PHILIPPINE REVIEW OF ECONOMICS
Vol. XXXVIII. No. 2. December 2001

Two Decades of Vector Autoregression (VAR) Modelling

Renato E. Reside, Jr. *

Abstract

A vector autoregression (VAR) is defined as a vector of endogenous variables regressed against its own lags. VARs therefore are considered part of a general class of simultaneous equations models. By construction, VAR analysis allows us to examine over time the dynamic impacts of innovations to variables on others. The following is a survey of the literature on vector autoregressions (VARs) in the last twenty years since it was first used for policy analysis by Christopher Sims [1980]. Initially, imposing a recursive structure on VAR disturbances had led to criticism that VAR modeling is atheoretical. In the last decade, however, a number of authors have attempted to remedy the problem by introducing new structural identification techniques. This has enhanced the ability of VARs to model dynamic economic relationships. VAR studies have been used primarily to identify the impacts of aggregate demand and supply shocks on aggregate output, as well as to identify the channels and impacts of monetary policy. The frontiers of current VAR research focus on open economy extensions, as well as on improving lag selection and estimation.

JEL classification: C32, C51
Keywords: Vector autoregressions

1. Introduction

In 1980, when Christopher Sims' seminal article using vector autoregressions (VAR) to model the monetary transmission mechanism in the US economy was published, the bulk of macroeconometric forecasting and modeling for policy analysis was being implemented through the use of large scale simultaneous equations models. Economists at the time were becoming skeptical about these large macroeconomic models, since they were unable to predict the high rates of inflation and unemployment in the 70's. These views were confirmed in the early eighties when large models began *underpredicting* the strength of the US economy.

* Assistant Professor, University of the Philippines School of Economics.

The problem with many large models, Sims argued, was that they included many "incredible restrictions" in that they tended to omit lagged values for many monetary variables in order to achieve identification. These were in contrast to his new VAR modeling technique, which employed less variables and imposed few a priori restrictions, such as exclusion restrictions on reduced form matrices, while accomplishing the same objective with far less complexity. A compact VAR model therefore could have the advantage of allowing for more efficient estimation, without imposing too many restrictions on the behavior of the economy. Early VARs (including Sims' original VAR) were labeled atheoretical in that they used only the observed time series properties of the data to forecast economic variables. Thus, both Keynesians and Monetarists could use the same VAR to forecast various economic variables even though the two groups had different views about the true structure of the economy [Hakkio and Morris 1984]. The only restrictions placed on the early VARs were restrictions that made the underlying structural model recursive and therefore, just-identified (in contrast to most large macroeconometric models, which are usually overidentified). Later versions of VAR would allow other types of restrictions to be imposed on the covariance matrix of the variables involved, some of which are typically not possible with larger models. Moreover, VAR models had the added features of being usable for forecasting and exogeneity/causality testing.

VAR models differed from large-scale econometric models in one other crucial aspect: they focused on the effects of the innovations or shocks (i.e., the residuals of the estimated reduced form of an unrestricted VAR), unlike larger models, which modeled shocks as exogenous changes in the levels of policy variables. By focusing on the effect of a shock to a policy variable, rather than its forecastable component implied by the monetary authorities' reaction function, VAR models can more easily accommodate the possibility that the two components of the policy variables may have different impacts on prices and output [Levy and Halikias 1997].[1]

These advantages were quickly recognized by other economists, who seized the opportunity to use VARs to fill perceived shortcomings in small-scale macroeconometric modeling. Sims and his cohorts began introducing VAR-based modeling into mainstream studies in economics [Litterman and Weiss 1981, Sims 1982, Gordon and King 1982, Doan, Litterman and Sims 1984].

However, several economists raised concerns over the use of an atheoretical process to make inferences about the structure of the economy. VAR modeling suffered a blow in 1985, when Cooley and LeRoy published their well-known critique of the Sims technique. They labeled VAR-based studies of the economy as "atheoretical macroeconomics", and argued, among other things, that for structural inferences from a VAR to be valid, restrictions based on theory ought to be imposed on VARs, and a recursive structure could never be always assumed to hold. Only if theory suggested a recursive ordering of variables within a VAR (and theory rarely does) could inferences from a VAR ever be valid.

[1] As in Barro's anticipated/unanticipated money papers [1977, 1978].

In response to Cooley and LeRoy, Sims [1986], Bernanke [1986], and Blanchard and Quah [1989] constructed VAR models using theory-based indentifying restrictions to identify structural coefficients. VARs no longer needed to be recursive, and modeling became far more flexible. The Sims, Bernanke, and Blachard and Quah studies lay the basis for improved identification in succeeding VAR studies. The Sims and Bernanke papers imposed contemporaneous restrictions on reduced form VAR shocks to achieve identification, while the last paper imposed long-run restrictions. Blanchard and Quah's paper represented a major breakthrough, because: (a) imposing long-run restrictions allowed VAR models to freely determine the short-run dynamics of the variables; and (b) there is much more agreement among economists about the long-run dynamics of macroeconomic variables. Many well-known restrictions implied by theory, such as long-run money neutrality could now be imposed on VAR models to achieve identification. Faust and Leeper [1997], though, discuss potential drawbacks to imposing long-run restrictions. Greene [1997] has remarked that the literature on VARs has come full circle as VAR modeling has evolved into the analysis of conventional dynamic structural simultaneous equations models.

Other researchers have since weighed in with their own criticisms of the VAR technique, notably, Rudebusch [1997], who surveyed VAR monetary policy studies and suggested that VAR monetary policy shocks were much more volatile than the surprise according to financial data (i.e., the difference between the forward federal funds rate and the realized federal funds rate). In addition, Rudebusch found little correlation between the two. Sims [1996] responded to earlier versions of the Rudebusch paper by arguing that although VARs did not include exogenous variables, they do not necessarily lead to bad estimates, but they do change the fitted residual. Sims also emphasized that the federal funds rate reacts quickly to new information about the state of the economy. Thus, the federal funds rate contains both policy surprises (the shocks researchers are interested in) as well as reactions to innovations in the state of the economy.

Other authors have attempted to provide more in-depth econometric interpretations of VAR results, notably Cochrane [1998], who argues that standard VAR policy tools such as impulse response functions are only relevant insofar as unanticipated policy matters. This is because impulse response functions of VARs depict the response of real variables to monetary policy innovations, residuals in the reduced form reaction function of the policymaker (the unexpected or unanticipated component of monetary policy).

Apart from the theoretical criticisms of VARs, VAR studies have produced their own share of controversial results, notably the price puzzle and the liquidity puzzle in early VAR studies of monetary policy, both of which numerous studies have attempted to resolve over the years. The price puzzle is the finding that negative innovations to money supply raise the price level (!) [Sims 1992]. The liquidity puzzle is the absence of a liquidity effect (i.e., a reduction in interest rates) in response to positive innovations to money supply [Leeper and Gordon 1992]. These controversies have been attributed to two factors: (a) shortcomings of the recursive VAR modeling approach; and (b) omitted variables in the VAR.

Both of these problems may lead to incorrect identification of monetary policy and the monetary policy instrument itself. The policy instruments in a VAR must be truly exogenous indicators of the stance of monetary policy – otherwise, shocks to it would not be representative of policy alone and we would not be able to measure the true impact of policy. Addressing the price puzzle could not be done without using better identification methods. Eichenbaum [1992] experimented with the use of narrower definitions of the monetary aggregate, replacing M1 in Sims' study with non-borrowed reserves, and got mixed results in eliminating the price puzzle. To eliminate the price puzzle, other authors include commodity prices in VAR studies, because it is found to be a good proxy for future inflation [Strongin 1995]. The price and liquidity puzzles have since been resolved through the use of improved structural identification techniques based on a more in-depth examination of the behavior of the monetary authority and of the economy [Leeper, Sims and Zha 1996].[2]

Another curious result caused by faulty identification in earlier VAR models is the result that shocks to money supply aggregates explain little of the forecast error variance in real variables [Sims 1992]. The reason is that the innovations to broad measures of money are a combination of endogenous responses to real shocks [King and Plosser 1984] and shifts in money demand [Bernanke and Blinder 1992]. The resolution of this problem lay in two approaches, one of them involving VARs. Bernanke and Blinder [1992] argue that the federal funds rate is the actual monetary policy instrument used by the Fed, and justify this on the basis of examining closely how the Fed actually implements policy. Using VAR models in which money innovations are compared to innovations in the funds rate, they confirm that the funds rate is superior to other candidate instruments (including several money measures) in that the funds rate contains more information about the future movement of real variables. In subsequent research for western economies, many authors designate shocks to central bank-influenced interest rates as the policy variable [Bernanke and Mihov 1998].[3]

More recent developments in VAR modeling have included the refinement of modeling the monetary policymaker's reaction function; open economy extensions; using panel data in VARs; and experimenting with asymmetrical lag lengths across VAR equations to achieve greater efficiency in estimation, as in Keating [2000].

---

[2] In an excellent survey of studies debating the existence of the liquidity effect, Pagan and Robertson [1995] suggest that in general, systems estimation techniques in VARs show that the liquidity effect exists.

[3] The second approach to identifying purely exogenous innovations in monetary policy is the Romer and Romer study [1989], in which they actually read minutes of Federal Open Market Committee (FOMC) meetings and identified dates in which they believed policy became contractionary.

2. What is a VAR?[4]

2.1 Compact form

VAR is essentially a time series technique, but it is flexible enough to incorporate features of conventional simultaneous equations models, such as restrictions on its reduced form to identify the structural relationships in the model (more on this later). Consider an n x 1 vector of variables, y_t. From its name, a vector autoregression (VAR) is a vector of variables, y_t, regressed on time lags of itself: y_{t-1}, y_{t-2}, …, y_{t-p}, where p is the number of periods. In matrix notation, this is expressed as:

$$y_t = A(L)\,y_{t-1} + \varepsilon_t \tag{1}$$

where A(L) is an n x n matrix polynomial in L, the lag operator, L = 0, 1, 2,…,p. Equation (1) is usually called an *unrestricted VAR*.[5] Suppose we add a constant term, μ, then we have:

$$y_t = \mu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t \tag{2}$$

where

ε_t is white noise and $\mathrm{cov}(\varepsilon_t) = \Omega$

For example, a VAR(2) of a 2 x 1 vector. Let $y_t = [x_t\,,\, z_t]'$. It follows that

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} A_{1,11} & A_{1,12} \\ A_{1,21} & A_{1,22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} A_{2,11} & A_{2,12} \\ A_{2,21} & A_{2,22} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{3}$$

Note that the VAR form in equation (3) is really y_t regressed on a series of predetermined variables (realizations of itself that are determined before time t, hence the term predetermined). Therefore, the VAR form in (3) can be interpreted as a reduced form of a system with n (= 2) endogenous variables, x_t and z_t.

---

[4] The succeeding sections on VAR modeling are based on Soderlind [2000] and Keating [1992].

[5] Why we need to distinguish between unrestricted and restricted VARs will be clear later.

2.2 Canonical form of an unrestricted VAR

A VAR(p) may be rewritten as a VAR(1). For example, a VAR(2) can be written as

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

or

$$y_t^* = \mu^* + Ay_{t-1}^* + \varepsilon_t^* \tag{4}$$

Suppose we ignore the first term in the previous equation. We get

$$y_t^* = Ay_{t-1}^* + \varepsilon_t^* \tag{5}$$

Next solve recursively to get another representation of y_t^*, the vector moving average representation, VMA, or impulse response function of the canonical form

$$\begin{aligned} y_t^* &= A(Ay_{t-2}^* + \varepsilon_{t-1}^*) + \varepsilon_t^* \\ &= A^2 y_{t-2}^* + A\varepsilon_{t-1}^* + \varepsilon_t^* \\ &= A^3 y_{t-3}^* + A^2 y_{t-2}^* + A\varepsilon_{t-1}^* + \varepsilon_t^* \\ &\vdots \\ &= A^{K+1} y_{t-K-1}^* + \sum_{s=0}^{K} A^s \varepsilon_{t-s}^* \end{aligned} \tag{6}$$

For a VAR to be stable, it must be that the first term in the previous equation (6) equals zero in the limit. This will only be the case if the eigenvalues of A are less than one in modulus. To see this, note that the matrix A may be decomposed using the spectral decomposition to yield:

$$A = Z\Lambda Z^{-1} \tag{7}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \text{ and } Z = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the main diagonal are the individual eigenvalues, and $z_1, z_2, \ldots, z_n$ are the associated eigenvectors of A. We can use the eigenvalue decomposition to justify the requirement that the eigenvalues must all be less than one for the VAR to be stable. Note that

$$A^2 = AA = Z\Lambda Z^{-1}Z\Lambda Z^{-1} = Z\Lambda\Lambda Z^{-1} = Z\Lambda^2 Z^{-1} \Rightarrow A^q = Z\Lambda^q Z^{-1} \tag{8}$$

Therefore, it follows that in general,

$$A^{K+1} = Z\Lambda^{K+1}Z^{-1} \tag{9}$$

where Z is the matrix of eigenvectors and Λ a diagonal matrix with eigenvalues.

$$A^{K+1} = Z\Lambda^{K+1}Z^{-1} \tag{10}$$

so that

$$\lim_{k\to\infty} \quad A^{K+1}y^K_{t-K-1} = 0 \tag{11}$$

(11) can only be satisfied if the eigenvalues of A are all less than one in modulus (this will be the case in (11) since Λ is diagonal, and all of the elements in the main diagonal will be fractions, whose value will tend to go to zero as its exponent, K, rises, so the RHS of (10) also goes to zero in the limit).

If we have a stable VAR, the first term in (6) vanishes. In this case, (6) can be re-written as the (modified) VMA form:

$$\begin{aligned} y^*_t &= \sum_{s=0}^{K} A^s \varepsilon^*_{t-s} \\ &= \varepsilon^*_t + A\varepsilon^*_{t-1} + A^2\varepsilon^*_{t-2}A^3\varepsilon^*_{t-3} + \ldots \end{aligned} \tag{12}$$

Note that this is still the canonical form, with

$$\begin{bmatrix} x_t \\ z_t \\ x_{t-1} \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} A_{1,11} & A_{1,12} & A_{2,11} & A_{2,12} \\ A_{1,21} & A_{1,22} & A_{2,21} & A_{2,22} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} A^2{}_{1,11} & A^2{}_{1,12} & A^2{}_{2,11} & A^2{}_{2,12} \\ A^2{}_{1,21} & A^2{}_{1,22} & A^2{}_{2,21} & A^2{}_{2,22} \\ A_{1,11} & A_{1,12} & A_{2,11} & A_{2,12} \\ A_{1,21} & A_{1,22} & A_{2,21} & A_{2,22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-2} \\ \varepsilon_{2t-2} \\ 0 \\ 0 \end{bmatrix} + \ldots \tag{13}$$

Note that we have an n (=2) variable VAR(p) in our example. We can conveniently extract the first n (=2) equations from the canonical form and write them as:

$$y_t = \varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + C_3\varepsilon_{t-3} + \ldots \tag{14}$$

and we now refer to this as the vector moving average (VMA) form of the VAR.

Stability of the VAR in (1) requires that the roots of the polynomial for the autoregressive part of the model

$$\left|1 - A_1 z - A_2 z^2 - \ldots - A_p z^p\right| = 0 \tag{15}$$

lie outside the unit circle. This condition is equivalent to the condition above that the eigenvalues of the matrix polynomial should be less than one in modulus.

There is an issue of whether the vector y_t should be stationary. Keating [1992] argues that if its elements each have a unit root, y_t should be replaced by Δy_t. Estimation can proceed in the same manner. Other authors, such as Sims [1980], suggest using levels instead, even if the variables contain a unit root, since differencing variables would needlessly throw away potentially useful information about their behavior.

Table 1 in the Appendix is a systematic classification of a number of VAR studies in the last two decades. Note that VARs find their greatest use in macroeconomics and monetary economics. In the former, it is used primarily in analyzing sources of business cycle fluctuations (demand vs. supply side). In the latter, VARs are used primarily in the analysis of the monetary transmission mechanism. All but one of the studies confine themselves to one type of restriction (only Gali's [1992] study combines contemporaneous with long-run restrictions). In recent years, open economy VARs have been attempted.

2.3 Toolkit for VAR estimation

2.3.1 How is a VAR estimated?

Since the VAR in (1) and (2) is a system of reduced form equations, the maximum likelihood estimate of the VAR (provided the disturbances are normally distributed and serially uncorrelated) is the same as OLS on each equation in the reduced form done separately. We therefore get consistent estimates of the coefficients of the VAR by implementing OLS on the RF equation by equation.

2.3.2 Standard VAR tools

Standard VAR applications usually involve the analysis of three outcomes: (1) Granger causality tests; (2) the impulse response function; and (3) forecast error variance decompositions. All three have become indispensable tools of VAR analysis:

Block-exogeneity Granger causality tests

If a variable z cannot help forecast x, then z does not Granger-cause x. To test this out, suppose x_t is an $n_1 \times 1$ vector and z_t is an $n_2 \times 1$ vector. Thus, if the $n_1 \times n_2$ matrices $A_{1,12} = \mathbf{0}$ and $A_{2,12} = \mathbf{0}$, then z fails to Granger-cause x, so it is said that z is exogenous with respect to x. This means that lagged shocks to z have no impact on x. These restrictions may be tested with a standard F test ın VAR systems.

The use of VAR analysis in the study of monetary policy can be traced back to Sims' [1973] paper on the use of VAR's in the study of money-income causality. In that paper, Sims runs a bivariate VAR consisting of money and income. Sims' results suggest Granger-causality running from money to income, so that money is exogenous with respect to income.

Impulse-response function

The standard approach taken is to examine pattern of responses to the vector of reduced form or structural form shocks (more on these later) to discern the time path of the effect of shocks on the endogenous variables over time.

Note from equation (14) that

$$\frac{\partial y_1}{\partial \varepsilon^1{}_1} = C_s \text{ so that } C_0 = 1 \tag{16}$$

Each C_s represents the response of the vector y_t to the vector of reduced form shocks in time t, ε_t. Therefore, the impulse response function {I, C_1, C_2, …} may be interpreted as the series of movements in y_t brought about by a succession of shocks over time.

Note that if we partition the coefficient matrices in equation (13), we find that

$$\begin{bmatrix} x_t \\ z_t \\ x_{t-1} \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & \vdots & A_2 \\ \cdots & \cdots & \cdots \\ I & \vdots & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} A_1^2 & \vdots & A_2^2 \\ \cdots & \cdots & \cdots \\ A_1 & \vdots & A_2 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-2} \\ \varepsilon_{2t-2} \\ 0 \\ 0 \end{bmatrix} + \ldots \tag{17}$$

Therefore, it follows that $C_s = A^s$, where C_s is the upper left block matrix in the coefficient matrices. This shows how the canonical form corresponds with the VMA.

As another alternative to the above procedure, we could proceed the following way: Let y_t be a column vector of n variables in the model. The reduced form VAR(p) would then be:

$$y_1 = A(L)\, y_{t+1} + \varepsilon_1$$

where (18)

$$A(L) = A_0 + A_1 L + A_1 L^1 + \ldots + A_p L^p$$

(18) implies that

$$[I - A(L)L]y_t = \varepsilon_t$$
$$y_t = [I - A(L)L]^{-1} \varepsilon_t \tag{19}$$

where it is clear that the last equation above is the impulse-response function, or VMA form of the VAR, relating the vector of endogenous variables to the reduced-form shocks.

$$\begin{aligned} y_t &= C(L)\varepsilon_t \\ y_t &= \varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + C_3\varepsilon_{t-3} + \dots \end{aligned} \tag{20}$$

which is equation (14).

Forecast error variance decomposition

The forecast error variance decomposition tells us the proportion of the movements in a time series due to its own shocks versus shocks to other variables. Analysis of the variance decomposition consists of examining it at various forecast horizons, to determine one or more variables are important in explaining movements in others. In monetary policy VAR studies, for instance [Bernanke and Blinder 1992, Bernanke and Mihov 1998], examination of variance decompositions in VARs were used in identifying the variable which reflected the stance of monetary policy, since this variable was expected to contain much information about the future movement of real variables.

The error forecast of the s-periods ahead forecast is

$$y_{t+s} - E_t y_{t+s} = \varepsilon_{t+s} + C_1\varepsilon_{t+s-1} + C_2\varepsilon_{t+s-2} + \dots + C_{s-1}\varepsilon_{t+1} \tag{21}$$

Therefore, the covariance matrix of the (s-periods ahead) forecast errors is:

$$E(y_{t+s} - E_t y_{t+s})(y_{t+s} - E_t y_{t+s})' = \Omega + C_1\Omega C_1' + \dots + C_{s-1}\Omega C'_{s-1} \tag{22}$$

Suppose we index the shocks by j = 1,..., n (there are n shocks because we have n variables in the VAR). If i = 1, ..., n denotes an index for the number of variables in the VAR, then we are looking at i variances of forecast errors due to j shocks.

If the shocks are uncorrelated, it is possible to calculate the fraction of the variance of the ith forecast error for time t + s, var($y_{i,t+s} - E_t y_{i,t+s}$), due to the jth shock, and examine the forecast error variance decomposition. To get these fractions, simply divide each element in the RHS of (22) by the total forecast error variance for the entire horizon on the LHS of (22). This fraction tells us what proportion of the forecast error in one variable (or a vector of variables) in time t + s is caused by lagged shocks to one or all of the variables in the system.

Note that the elements in the RHS of (22) are really sums of quadratic forms, with the elements of the C matrices serving as weights of the covariance terms, Ω. So, the forecast error variance is simply a weighted average of the covariance terms. The analysis of impulse-response functions and forecast error variance decompositions is called *innovation accounting*.

3. VAR identification issues

3.1 Historical background

Note that (1) and (2) are both reduced form equations, because the RHS of both equations is comprised of predetermined variables. In their widely cited 1985 paper, Cooley and LeRoy cautioned against drawing inferences from standard VAR tools such as impulse-response functions and variance decompositions unless restrictions based on economic theory (in the traditional approach of the Cowles Commission) were imposed on the reduced form of the VAR model. At the time, the authors using VAR estimation in their papers had been using a recursive (Cholesky decomposition) approach to identification, and Cooley and LeRoy felt they had done so without providing sufficient justification for the restrictions implied by such an identification approach. Because of the absence of restrictions from theory in VAR studies, Cooley and LeRoy branded the work of Sims and his cohorts as "atheoretical macroeconomics". They argued that VAR modelers either had to justify the recursive structure imposed on their models, or else identify VARs using restrictions based on theory. Failing this, conclusions drawn from impulse response functions and other VAR tools were meaningless. They had exposed weaknesses in Sims' modeling approach, suggesting that it suffered from flaws that did not afflict conventional overidentified simultaneous equations models. As a response to these charges, Sims [1985] and Bernanke [1986] developed structural vector autoregression techniques (SVAR). These methods were later refined by Blanchard and Quah [1989].

3.2 Identifying parameters from reduced form VARs

Since (1) and (2) are reduced forms, it follows that the VMA is a series of responses of y_t to the reduced form shocks to x_t and z_t. However, economists are usually more interested in determining the VMA and variance decomposition with respect to structural shocks, which we assume to be a linear function of the reduced form shocks:

$$u_t = F\varepsilon_t \quad \text{, with } \operatorname{var}(u_t) = D \tag{23}$$

where F is an invertible n x n matrix containing the coefficients of ε_t. The variance of the structural shocks is D. Usually, D is assumed to be diagonal. Using (23), we can rewrite the VMA in (14) as

$$y_t = F^{-1}u_t + C_1F^{-1}u_{t-1} + C_2F^{-1}u_{t-2} + C_3F^{-1}u_{t-3} + \dots \tag{24}$$

and the forecast error variance may be rewritten as:

$$E(y_{t+s} - E_t y_{t+s})(y_{t+s} - E_t y_{t+s})' = F^{-1}D(F^{-1})' + C_1F^{-1}D(F^{-1})'C_1' + \dots + C_{s-1}F^{-1}D(F^{-1})'C'_{s-1} \tag{25}$$

Both (24) and (25) have more profound economic interpretations than (14) and (19), respectively. However for us to be able to implement (24) and (25), we need to be able to identify the structural parameters of the *structural form* of the model. Since D is usually taken to be equal to I, then this usually amounts to choosing an appropriate form for F that allows us to achieve exact- or over-identification.

Suppose economic theory suggests that the structural form of a model is:

$$Fy_t = \alpha + B_1y_{t-1} + \dots + B_py_{t-p} + u_t, \; u_t \text{ is white noise, } \operatorname{cov}(u_t) = D \tag{26}$$

then, since F is invertible, then we could rewrite (26) as

$$y_t = F^{-1}\alpha + F^{-1}B_1y_{t-1} + \dots + F^{-1}B_py_{t-p} + F^{-1}u_t \tag{27}$$

$$= \mu + A_1y_{t-1} + \dots + A_py_{t-p} + \varepsilon_t \tag{28}$$

where ε_t is white noise and $\operatorname{cov}(\varepsilon_t) = \Omega$

This implies that

$$\mu = F^{-1}\alpha, \; A_s = F^{-1}B_s, \text{ and } \varepsilon_t = F^{-1}u_t \text{ , so that } \Omega = F^{-1}D(F^{-1})' \tag{29}$$

The ε_t terms are called *orthogonalized* innovations, since when multiplied by the matrix F, they are orthogonal to the reduced form innovations u_t. Note that (27) and (28) are VAR forms, which are really reduced form models, since each variable is regressed against its own lags and lags of other variables, so therefore, all regressors are predetermined.

Since the reduced form (RF) in (27) and (28) may be estimated directly from time series data, it is possible to recover the structural parameters in the structural form (SF) in (26) provided we impose enough restrictions on the structural parameters F, B_s, α, and D. The table below shows that we need to impose at least n^2 additional restrictions on the RF to achieve identification of all of the structural parameters.

Table 2: Deriving the structural parameters from the estimated reduced form VAR

Structural Form (SF)	*Number of unique parameters (to be identified or estimated)*	*Reduced Form (RF)*	*Number of unique parameters (known from estimation)*
F	n^2		
B_s	pn^2	A_s	pn^2
α	n	μ	n
D	n(n+1)/2 (symmetric)	Ω	n(n+1)/2 (symmetric)
Total number of parameters	(1+p) n^2 + n (1+((n+1)/2))		pn^2 + n (1+((n+1)/2))

Alternatively, we could use equation (19) in order to derive a more compact notation for impulse-responses to the structural shocks:

$$y_t = [I - A(L)L]^{-1} F^{-1} u_t$$

where we let

$$\theta(L) = [I - A(L)L]^{-1} F^{-1} \quad \text{so that } y_t = \theta(L) u_t \tag{30}$$

Example (SF of the 2 x 1 case):

$$\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} B_{1,11} & B_{1,12} \\ B_{1,21} & B_{1,22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} B_{2,11} & B_{2,12} \\ B_{2,21} & B_{2,22} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{31}$$

This system has 3 x 4 + 3 unique parameters. The VAR RF in (3) has 2 x 4 + 3. Therefore, we need at least 4 restrictions on {F, B_s, α, and D} to identify them from { A_1, A_2, and Ω). Over the past two decades, a variety of identification methods have emerged for the identification of structural VAR's. The identification methods in the following brief survey hold the key to understanding structural vector autoregressions.

3.3 Possible restrictions and identification patterns

Almost all VAR studies impose enough restrictions on the RF to just-identify the model.[6] For structural VARs, the most common identifying restrictions are those that impose zero restrictions on F and normalize the elements in the D matrix. Assume F is lower triangular (n(n-1))/2 restrictions), and that D = I (n(n+1))/2). This identification scheme yields just enough restrictions (n^2) to exactly identify the model. Normalizing the variances of the SF shocks to unity means that the impulse response of y generated by setting u to unity is the effect of a structural shock on y of the size of one standard deviation.

Note that a lower triangular F implies that the first variable can react to its own lags and the first shock, the second variable to its own lags and the first two shocks, etc. This identification pattern means that this is a recursive simultaneous equations model, so we have to be careful in ordering the variables in this VAR. In order to actually implement this identification pattern in computer programs such as Time Series Processor (TSP) and Regression Analysis for Time Series (RATS), we do not actually need to impose the restrictions ourselves. Computer programs do this by carrying out a Cholesky decomposition of Ω. This technique for achieving identification was first used by Sims in his first VAR paper [1980].

What follows are brief descriptions of general identification patterns for VARs, and examples of each from past VAR studies in chronological order.

3.3.1 Sims-type identification (Cholesky decomposition of Ω)

Perform a Cholesky decomposition of the covariance matrix of RF shocks, Ω, which is known directly from the estimation. This will (automatically) yield a lower triangular F matrix and restrict D = I. The Cholesky decomposition is carried out as follows: since Ω is symmetric, it follows by definition that it can be decomposed into two lower triangular matrices, X and X' with 1's on their main diagonal.

If Ω is a symmetric positive definite matrix, then there exists a lower triangular matrix X such that

$$\Omega = XX' \tag{32a}$$

so that

$$chol(\Omega) = X \tag{32b}$$

[6] The possible exception is the VAR study by Bernanke and Mihov [1997]. They impose over-identifying restrictions on monetary policy, and conduct tests of these restrictions.

However,

$$\Omega = F^{-1}D(F^{-1})',$$

so to pattern this after (32b), let D = I to get

$$\Omega = F^{-1}(F^{-1})' \qquad (33a)$$

and so,

$$chol(\Omega) = F^{-1} \qquad (33b)$$

Since F^{-1} is lower triangular, it follows that its inverse, F, will also be lower triangular. F will have 1's on its main diagonal, zero's above the main diagonal, and unrestricted parameters elsewhere. Thus, it follows that applying the F derived from the Cholesky decomposition on the SF in (26) yields a fully recursive system of equations (which, as is well known, is exactly-identified). Any time the D = I restriction is imposed, structural shocks are assumed to be of the magnitude of one standard deviation (or have unit variance). Signs of the elements in F may be chosen freely.

Notwithstanding the fact that a recursive structure is imposed upon all Sims-type VAR models, Cooley and LeRoy referred to such models as *atheoretical VAR models*, since most studies using the Sims identification at the time failed to provide theoretical justification for it.

Examples

a) Sims [1980]

In 1980, the first of Sims papers with VAR applied for forecasting purposes and impulse-response functions computed for policy analysis was published. Sims used a Cholesky decomposition to identify the structural form of an unrestricted VAR with variables money, real GNP, unemployment, wages, prices, and import prices.

b) Bernanke and Blinder [1992]

Many of the early papers using VAR techniques utilized the Sims-type identification technique. These include the well-known paper of Bernanke and Blinder [1992] on the monetary transmission mechanism, in which unrestricted (recursive) VAR's are used to identify the responses of bank loans, securities and deposits to structural disturbances emanating from the federal funds rate. Variables included in VARs estimated for this purpose include the funds rate, unemployment rate, log of CPI, and the three bank balance sheet variables.

The federal funds rate was found to be a good predictor of the future behavior of many real variables because it reflects the stance of monetary policy (since it reflects shocks to the supply of bank reserves). Using another set of unrestricted VAR estimates (this time with the funds rate, other interest rates and monetary aggregates in the VAR), Bernanke and Blinder find that the federal funds rate is superior to other interest rates and monetary aggregates in predicting a variety of real output measures. They also found that bank deposits fall in reaction to federal funds rate innovations (tightening monetary policy), and banks respond by reducing securities and loans in the short-run, and then later realigning their portfolios in the medium- to long-run, away from loans and into securities. They interpret these findings as consistent with the credit and money views of the monetary transmission mechanism, as well as the fact that money has real effects on the economy. Being the policy indicator, a measure of the federal funds rate is ordered last in all of the VARs estimated.

3.3.2 Bernanke-type identification (contemporaneous structural model)

For this method, take (33a), but do not restrict F to be lower triangular. Thus, we can freely impose zero restrictions on the elements of F and still have a just-identified system. However, D = I, as usual. Note that the Bernanke [1986] identifcation method greatly enhances the ability of VARs to model structural forms. Therefore, as before, we need to impose n(n-1))/2 restrictions on F, but we are not restricting the system to be recursive. Given these restrictions, we can solve for the remaining elements of B with a numerical method for solving non-linear equations (This is possible, since $A_s = F^{-1} B_s$. We automatically know B_s if we identify F, since A_s is known from the estimation of the RF). Models that utilize the Bernanke identification method are typically referred to as contemporaneous structural VAR models because they relate structural shocks (u_t) to reduced form innovations (ε_t) via the matrix of coefficients F by the relation $u_t = F\varepsilon_t$. Note, however, that placing contemporaneous restrictions on the reduced form implied placing restrictions on the *short-run* behavior of variables (or contemporaneous relationships between them). This problem would subsequently be addressed by Blanchard and Quah [1989].

Examples

a) Keating [1992]

The model ($u_t = F\varepsilon_t$) is the following

$$\begin{bmatrix} u_t^{AS} \\ u_t^{IS} \\ u_t^{Ms} \\ u_t^{Md} \end{bmatrix} = \begin{bmatrix} F_{11} & 0 & 0 & 0 \\ F_{21} & F_{22} & F_{23} & F_{24} \\ 0 & 0 & F_{33} & F_{34} \\ F_{41} & F_{42} & F_{43} & F_{44} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{p} \\ \varepsilon_t^{y} \\ \varepsilon_t^{r} \\ \varepsilon_t^{m} \end{bmatrix} \tag{34}$$

and this corresponds to the following model of the economy

$$u_t^{AS} = F_{11}\varepsilon_t^p$$
$$u_t^{IS} = F_{21}\varepsilon_t^p + F_{23}\varepsilon_t^r + F_{24}\varepsilon_t^m + F_{22}\varepsilon_t^y$$
$$u_t^{Ms} = F_{34}\varepsilon_t^m + F_{33}\varepsilon_t^r$$
$$u_t^{Md} = F_{41}\varepsilon_t^p + F_{42}\varepsilon_t^y + F_{43}\varepsilon_t^r + F_{44}\varepsilon_t^m \tag{35}$$

Clearly from (34), the structural shocks are linear functions of the reduced form shocks. Keating's structural model assumes that producers respond immediately to predetermined aggregate supply shocks. The second equation is a reduced-form IS equation, so that structural IS shocks are functions of shocks to prices, interest rates, money and income. The third equation allows interest rates set by the Fed to respond to changes in money stock (the Fed is assumed not to respond to aggregate measures of output and price[7]). The final equation is thus a short-run money demand function that relates nominal money holdings to nominal GDP and the interest rate. This specification is motivated by buffer stock theory, where short-run money holdings rise in proportion to nominal income. In the final equation, Keating imposes the additional restriction that $F_{41} = F_{42}$, so that the first two elements on the RHS of the last equation is the log of nominal GDP shock.

Note from equation (34) that the zero restrictions imposed on matrix F, along with the additional restriction that $F_{41} = F_{42}$ impose exactly six restrictions on the reduced form VAR, leading to a just-identified model, because D = I is assumed. This means that the variances of the four structural shocks are constrained to unit variances, and the four shocks are uncorrelated.

b) Karras [1993]

Karras' paper assumes that the US economy is driven by six structural shocks: oil, non-oil, aggregate supply, fiscal, monetary, aggregate demand and exchange rate innovations. The reduced form VAR is:

$$\begin{bmatrix} X_t \\ O_t \end{bmatrix} = \sum_{i=1}^{k} \begin{bmatrix} I - A_0 - a_0 \\ 0' \quad 1 \end{bmatrix}^{-1} \begin{bmatrix} A_i - a_1 \\ 0' \quad 1 \end{bmatrix} \begin{bmatrix} X_{t-i} \\ 0_{t-i} \end{bmatrix} + \begin{bmatrix} I - A_0 - a_0 \\ 0' \quad 1 \end{bmatrix}^{-1} \begin{bmatrix} u_t^x \\ 0_t^0 \end{bmatrix}$$

$$Z_t = \sum_{i=1}^{k} C_i Z_{t-i} + z_t \tag{36}$$

[7] This specification may be quite objectionable. Note that the identification of any structural VAR model is subject to the same pitfalls as the identification of any simultaneous equations model.

The estimated reduced form residuals are z = (d, m, y, p, e, o)'. These stand for the real federal government deficit, money supply, real GNP, GNP deflator, exchange rate, and price of oil, respectively.

The structural model with contemporaneous restrictions is:

$$
\begin{aligned}
&\text{Fiscal policy rule: } d_t = a_1 y_t + u_t^F \\
&\text{Monetary policy rule: } m_t = b_1 d_t + b_2 y_t + b_3 p_t + u_t^M \\
&\text{Aggregate demand: } y_t = c_1 o_t + c_2 d_t + c_3 m_t + c_4 p_t + u_t^D \\
&\text{Aggregate supply: } y_t = d_1 o_t + d_2 p_t + u_t^S \\
&\text{Nominal exchange rate: } e_t = f_1 o_t + f_2 d_t + f_3 m_t + f_4 y_t + f_5 p_t + u_t^E \\
&\text{Oil prices: } o_t = u_t^O
\end{aligned}
\tag{37}
$$

Note that the structural shocks are denoted by u^F, u^M, u^D, u^S, and u^E.

Karras finds that supply-side shocks are important for output fluctuations both in the short- and long-run, and that these affect output permanently. Demand side shocks have only transitory effects on output. All innovations affect inflation, but only money growth has a permanent effect on inflation. Faster money growth leads to a depreciation of the dollar, but so do higher budget deficits.

c) Haslag and Hein [1995]

Using contemporaneous structural VARs, the authors attempt to differentiate the impact of monetary policy implemented via changes in reserve requirements versus policy implemented via open market operations.

Given the following:

RAM = reserve adjustment magnitude (a measure of dollars freed (impounded) as a result of increases (decreases) in reserve requirement ratios.

H = monetary base

mm2 = growth rate of the M2 multiplier

INF = inflation

GDP = real GDP growth

The contemporaneous VAR model ($u_t = F\varepsilon_t$) is the following

$$\begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \\ u_t^4 \\ u_t^5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -F_{21} & 1 & -F_{23} & -F_{24} & 0 \\ 0 & 0 & 1 & -F_{34} & -F_{35} \\ -F_{41} & -F_{42} & 0 & 1 & 0 \\ -F_{51} & -F_{52} & -F_{53} & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^{RAM} \\ \varepsilon_t^{H} \\ \varepsilon_t^{mm2} \\ \varepsilon_t^{INF} \\ \varepsilon_t^{GDP} \end{bmatrix} \tag{38}$$

The first equation in this system postulates that RAM is a structural disturbance. The second equation specifies movements in high-powered money respond contemporaneously to RAM. The third equation is a money demand equation. It is postulated that both inflation and income affect the demand for M2 assets.

The authors find evidence that the Fed at least partially offsets reserve requirement changes (= movements in RAM) with open market operations (= changes in high-powered money). In addition, the authors find that innovations in the monetary base have the same contemporaneous effect on macroeconomic activity, regardless of how policy is implemented. However, the dynamic macroeconomic responses are quite different.

d) Leeper, Sims and Zha [1996]

This paper surveys the literature on identified VAR modeling for monetary policy analysis and attempts to determine the effects of monetary policy on key economic variables. The authors summarize current results, identify puzzles prevalent in results (e.g., the liquidity puzzle and the price puzzle). The liquidity effect hypothesizes that policy-induced increased liquidity brought about by a monetary expansion should lower interest rates. The price puzzle is the finding in some VAR studies that a monetary contraction produces inflation.

The authors argue that larger VAR models that capture the behavior of the Fed and the banking sector better reflect observed economic behavior. For example, one way the authors model the economy is by separating it into three components: (a) private sector variables that do not respond quickly to financial signals; (b) information – this component of nonpolicy behavior responds quickly to new information; and (c) the Federal Reserve's behavior. They suppose that the restrictions on the VAR could come in the form of:

Sector	*Variable*	*CPI*	*Y*	*RF*	*M1*
P	CPI	X	X		
P	Y	X	X		
I	RF	X	X	X	X
F	M1			X	

where the Xs indicate coefficients on the F matrix that are unrestricted, and the blanks indicate coefficients that are postulated to be zero. The first row gives the names of the variables, the first column gives the names of the sectors in which the disturbances originate. The authors extend their model further to explicitly account for the behavior of the banking sector.

e) Bernanke and Mihov [1998]

This paper aims to determine how the Bundesbank reacts to various macroeconomic variables such as inflation and monetary aggregates in order to ascertain the goals of monetary policy. They find that the Bundesbank is more appropriately described as an inflation targeter rather than a monetary targeter, contrary to its pronouncements.

Prior to ascertaining how the Bundesbank responds to changes in forecasts of inflation, money growth and other variables, they identify an appropriate indicator of the stance of monetary policy. Bernanke and Mihov make use of the following approach in identifying such an indicator.

The economy is assumed to be described by the following linear structural model:

$$Y_t = \sum_{i=0}^{k} B_i Y_{t-i} + \sum_{i=0}^{k} C_i P_{t-i} + A^y \upsilon_t^y$$

$$P_t = \sum_{i=0}^{k} D_i Y_{t-i} + \sum_{i=0}^{k} G_i P_{t-i} + A^p \upsilon_t^p \qquad (39)$$

where Y and P denote vectors of policy (P) and non-policy (Y) variables, respectively. The set of policy variables includes variables that are potentially useful as indicators of the stance of monetary policy, such as short-term interest rates and reserves measures. Non-policy variables include output and inflation, whose responses to monetary policy shocks they want to identify. The ν's are mutually uncorrelated structural disturbances, and these are pre-multiplied by the matrix A, allowing any disturbance in the Y block to enter into any equation in the same block (the same holds for P).

Note that the structural form in (39) may be written in the following matrix form:

$$\begin{bmatrix} I-B_0 & 0 \\ -D_0 & I-G_0 \end{bmatrix}\begin{bmatrix} Y_t \\ P_t \end{bmatrix}=\sum_{i=0}^{k}\begin{bmatrix} B_s & C_s \\ D_s & G_s \end{bmatrix}\begin{bmatrix} Y_{t-s} \\ P_{t-s} \end{bmatrix}+\begin{bmatrix} A^y & 0 \\ 0 & A^p \end{bmatrix}\begin{bmatrix} \upsilon_t^y \\ \upsilon_t^p \end{bmatrix} \tag{40}$$

To get the reduced form version, pre-multiply (40) by

$$\begin{bmatrix} I-B_0 & 0 \\ -D_0 & I-G_0 \end{bmatrix}^{-1} \tag{41}$$

which, by the rules of partitioned matrices, is equivalent to:

$$\begin{bmatrix} Y_t \\ P_t \end{bmatrix}=\sum_{i=0}^{k}A_s\begin{bmatrix} Y_{t-s} \\ P_{t-s} \end{bmatrix}+\begin{bmatrix} u_t^y \\ u_t^p \end{bmatrix} \tag{42}$$

The reduced form in compact notation is

$$\begin{bmatrix} Y_t \\ P_t \end{bmatrix}=\sum_{i=0}^{k}A_s\begin{bmatrix} Y_{t-s} \\ P_{t-s} \end{bmatrix}+\begin{bmatrix} u_t^y \\ u_t^p \end{bmatrix} \tag{43}$$

It is thus apparent that

$$u_t^p=(I-G_0)^{-1}A^p\upsilon_t^p$$

or

$$(I-G_0)^{-1}u_t^p=A^p\upsilon_t^p \tag{44}$$

Bernanke and Mihov use the following model to describe the workings of the reserves market:

$$\text{Total reserves demand:} \quad u_{TR} = -\alpha u_{CR} + \upsilon_t^d$$

$$\text{Lombard loans demand:} \quad u_{LL} = \beta(u_{CR} - u_{LR}) + \upsilon_t^b$$

$$\text{Nonborrowed reserves supply:} \quad u_{NBR} = \phi^d \upsilon^d + \phi^b \upsilon^d + \phi^s \upsilon^s + \upsilon^n$$

$$\text{Lombard rate:} \quad u_{LR} = \gamma^d \upsilon^d + \gamma^n \upsilon^n + \gamma^b \upsilon^b + \upsilon^s \tag{45}$$

This model is equivalent to equation (44) in matrix form:

$$\begin{bmatrix} 1 & 0 & -\alpha & 0 \\ 1 & -1 & -\beta & \beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{TR} \\ u_{NBR} \\ u_{CR} \\ u_{LR} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \varphi^d & \varphi^b & 1 & \varphi^s \\ \gamma^d & \gamma^b & \gamma^n & 1 \end{bmatrix} \begin{bmatrix} \upsilon^d \\ \upsilon^b \\ \upsilon^n \\ \upsilon^s \end{bmatrix} \tag{46}$$

In other words,

$$\begin{bmatrix} u_{TR} \\ u_{NBR} \\ u_{CR} \\ u_{LR} \end{bmatrix} = \begin{bmatrix} 1 - \dfrac{\alpha(1-\varphi^d+\beta\gamma^d)}{\alpha+\beta} & \dfrac{\alpha(1+\varphi^b+\beta\gamma^b)}{\alpha+\beta} & \dfrac{\alpha(1-\beta\gamma^n)}{\alpha+\beta} & \dfrac{\alpha(\varphi^s-\beta)}{\alpha+\beta} \\ \varphi^d & \varphi^b & 1 & \varphi^s \\ \dfrac{1-\varphi^d+\beta\gamma^d}{\alpha+\beta} & -\dfrac{1+\varphi^b-\beta\gamma^b}{\alpha+\beta} & -\dfrac{1-\beta\gamma^n}{\alpha+\beta} & -\dfrac{\varphi^s-\beta}{\alpha+\beta} \\ \gamma^d & \gamma^b & \gamma^n & 1 \end{bmatrix} \begin{bmatrix} \upsilon^d \\ \upsilon^b \\ \upsilon^n \\ \upsilon^s \end{bmatrix} \tag{47}$$

Note from (47) that this structural form has 12 unknown parameters (the eight unknown elements in the two 4 x 4 matrices in (47), as well as the four shock variances). All of these are to be estimated from ten reduced form residual variances and covariances (= 4(4+1) / 2 estimated parameters since the reduced form covariance matrix is symmetric). That is,

$$\text{cov}(u_t^P) = (I - G_0)^{-1} A^P E(\upsilon_t^P \upsilon_t^{P\,\prime}) A^{P\,\prime} (I - G_0)^{-1\,\prime}$$

$$\Omega = (I - G_0)^{-1} A^P D A^{P\,\prime} (I - G_0)^{-1} \tag{48}$$

so that $(I - G_0)^{-1} A^P$ is equivalent to the F^{-1} matrix in (26), (27), (29). Note that two additional restrictions need to be applied to identify the model. Bernanke and Mihov supply these by imposing additional restrictions implied by alternative operating procedures:

(a) Call rate smoothing – the call rate rate does not respond to (i.e., the Bundesbank offsets) reserves demand and Lombard borrowing shocks. This implies that in (47),

$$1 - \varphi^d + \beta\gamma^d = 0,\ 1 + \varphi^b - \beta\gamma^b = 0\ ,\ \varphi^s = 0$$

(b) Non-borrowed reserves targeting – nonborrowed reserves depend only on their own autonomous shocks and are not systematically adjusted in response to contemporaneous shocks in the bank reserves market. This implies that in (47),

$$\varphi^d = 0,\ \varphi^b = 0\ ,\ \varphi^s = 0$$

(c) Smoothing of the Lombard rate – the Lombard rate is independent of contemporaneous reserves market shocks. This implies that in (47),

$$\gamma^d = 0,\ \gamma^b = 0\ ,\ \gamma^n = 0$$

Each of these alternatives imposes three additional restrictions, leading to overidentification. A statistical comparison of models suggests that the Lombard rate model is the best model, so that the Lombard rate is superior to the call rate and nonborrowed reserves as the best indicator of the stance of monetary policy.

Bernanke and Mihov proceed by estimating VARs of the Lombard rate, as well as inflation forecasts and money growth forecasts. From the variance decompositions, they find that in short horizons, the forecast error for the Lombard rate is dominated by own shocks, but in longer horizons, it is dominated by shocks to the inflation forecast, leading them to conclude that the Bundesbank is more of an inflation targeter than a money growth targeter.

f) Soderlind [2000]

Soderlind explains how VARs are applied in the analysis of monetary policy. Partition the vector y_t into a policy instrument used by the central bank, s_t, variables which come before it, x_{1t}, and variables which come after it, x_{2t}:

$$y_t = \begin{bmatrix} x_{1t} \\ s_t \\ x_{2t} \end{bmatrix} \tag{49}$$

Now, rewrite (49) as

$$\begin{bmatrix} B_0^{11} & 0 & 0 \\ B_0^{21} & 1 & 0 \\ B_0^{31} & B_0^{32} & B_0^{33} \end{bmatrix} \begin{bmatrix} x_{1t} \\ s_t \\ x_{2t} \end{bmatrix} = \begin{bmatrix} B_1^{11} & B_1^{12} & B_1^{13} \\ B_1^{21} & B_1^{22} & B_1^{23} \\ B_1^{31} & B_1^{32} & B_1^{33} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ s_{t-1} \\ x_{2t-1} \end{bmatrix} + \ldots + \begin{bmatrix} B_2^{11} & B_2^{12} & B_2^{13} \\ B_2^{21} & B_2^{22} & B_2^{23} \\ B_2^{31} & B_2^{32} & B_2^{33} \end{bmatrix} \begin{bmatrix} x_{1t-p} \\ s_{t-p} \\ x_{2t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{st} \\ u_{2t} \end{bmatrix} \tag{50}$$

$$B_0 \quad \mathrm{y}_t \quad = B_1 \quad \mathrm{y}_{t-1} \; + \ldots + \quad B_2 \quad \mathrm{y}_{t-p} \quad + u_t$$

where B_0^{11} and B_0^{33} are lower triangular matrices with ones in the diagonal. The equation for the policy instrument s_t in (50) is

$$s_t = -B_0^{21} x_{1t} + \begin{bmatrix} B_1^{21} & B_1^{22} & B_1^{23} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ s_{t-1} \\ x_{2t-1} \end{bmatrix} + \ldots + \begin{bmatrix} B_p^{21} & B_p^{22} & B_p^{23} \end{bmatrix} \begin{bmatrix} x_{1t-p} \\ s_{t-p} \\ x_{2t-p} \end{bmatrix} + u_{st} \tag{51}$$

Note that this is a *reaction function* (since it depicts how the central bank's policy instrument responds to the contemporaneous values of the variables which precede it (x_{1t}) and the lagged values of all variables, plus a monetary policy shock, u_{st}. In other words, policy in time t is determined by a rule which depends upon the contemporaneous x_{1t} and all lagged variables, plus a monetary policy shock, u_{st}. Suppose s_t is the j^{th} element in y_t. Then the impulse-response with respect to the monetary policy shock is located in the j^{th} columns of the matrices in (50): B_0^{-1}, $B_1 B_0^{-1}$, $B_2 B_0^{-1}$, …

Suppose we want to analyze monetary policy shocks and the effects they have on other variables in the VAR system (for instance, output and prices). Identification of the model relies on the ordering of the variables in the VAR. That is, on the structure of the contemporaneous correlations as captured by B_0. Thus, it is important to understand how the results of the monetary policy shock could be changed if the variables are reordered.

It is sometimes found in VAR studies that policy surprises explain only a small fraction of the variance in y_t (this is the typical result for US studies after

1982). There are two explanations for this: (a) anticipated (systematic) monetary policy reduces the variance of output considerably; and (b) variance decompositions do not tell us about the potential effects of monetary policy surprises (but the impulse-response function does).

Note that contemporaneous structural VAR models constrain the dynamic behavior of the variables in the short-run. This concern motivated Blanchard and Quah to establish a new identification modality: one that imposed restrictions on the long-run impulse-responses.

3.3.3 Blanchard and Quah-type identification (identification via imposition of long-run restrictions)

The Blanchard and Quah [1989] identification technique further enhanced the ability of VAR models to identify structural forms, because restrictions were no longer imposed on the short-run (contemporaneous) behavior of variables. The main idea behind this method is to triangularize (i.e., apply the Cholesky decomposition) the matrix of long-run multipliers (which, also leads to an exactly-identified system).

Note from (20) that the VMA or impulse response function for the structural shocks u, is:

$$y_t = F^{-1}u_t + C_1F^{-1}u_{t-1} + C_2F^{-1}u_{t-2} + C_3F^{-1}u_{t-3} + \dots$$

$$y_t = C(L)F^{-1}u_t \tag{52}$$

From (52), the long-run response of y to u is captured by the sum of the coefficients of u:

$$\begin{aligned} \lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial u_t'} &= \lim_{s\to\infty} C_s F^{-1} \\ &= C(1)F^{-1} \end{aligned} \tag{53}$$

where $C(1) = \Sigma^{\infty}_{j=0} C_s$.

There must be (n(n-1)/2) restrictions imposed on these long-run responses (to get the needed n^2 restrictions for exact-identfication of all elements in F). In order to do this, we note that since

$$\varepsilon_t = F^{-1}u_t$$

It follows that

$$EC(1)F^{-1}u_t u_t'(F^{-1})'C(1)' = EC(1)\varepsilon_t \varepsilon_t' C(1)' \tag{54}$$

$$C(1)F^{-1}D(F^{-1})C(1)' = C(1)\Omega C(1)$$

$$C(1)F^{-1}(F^{-1})C(1)' = C(1)\Omega C(1) \tag{55}$$

We can thus apply the Cholesky decomposition to the RHS of (54) (the RHS of (55) is known from the estimation of the VAR RF) to obtain the lower triangular matrix Λ, the matrix containing the long-run responses of y to the structural shocks:

$$\Lambda = C(1)F^{-1} \tag{56}$$

We can solve for F^{-1} and thus, F, by using (56) (since C(1) is known from estimation of the RF)).

Example: (Imposing long-run restrictions in the i = 2 case)

Recall that

$$y_t = C(L)F^{-1}u_t$$

$$= (I + C_1 L + C_2 L^2 + C_3 L^3 + \ldots)F^{-1}u_t \tag{57}$$

Setting L = 1 in (57) yields the long-run coefficients of the structural shocks, C(1) F^{-1}

$$(I + C_1 + C_2 + C_3 + \ldots)F^{-1}u_t \tag{58}$$

Therefore, relating (58) to (56), we see that

$$\Lambda = C(1)F^{-1}$$

$$\begin{bmatrix} \text{unrestricted} & 0 \\ \text{unrestricted} & \text{unrestricted} \end{bmatrix} = (I + C_1 + C_2 + C_3 + \ldots)F^{-1} \tag{59}$$

In the i = 2 case, the LHS of (59) equals

$$\Lambda = (I + A^1 + A^2 + A^3 + ...)\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1}$$

$$= (I - A)^{-1}\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \text{unrestricted} & 0 \\ \text{unrestricted} & \text{unrestricted} \end{bmatrix} = \begin{bmatrix} 1 - A_{11} & -A_{12} \\ -A_{21} & 1 - A_{22} \end{bmatrix}^{-1}\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}^{-1} \qquad (60)$$

Note that the long-run restriction that the upper right hand element of Λ equals zero imposes one restriction on the elements of F (i.e., that the sum of the terms in the upper right hand element of the RHS of (60) equals zero). Remember that this restriction is just one equation that will help pin down one of the elements of F. The other three restrictions (equations) are provided by the fact that $F^{-1}(F^{-1})' = \Omega$ (since Ω is symmetric). This yields three equations, so that the system to be solved has four equations in four unknowns (again, exactly identified).

It is now clear that identification and estimation of the elements in the F matrix holds the key to the SVAR exercise. When routines written in RATS and TSP for estimating F recover this matrix, they then proceed to input F in (24) and (25) to produce the "structural" form of the VMA/impulse-response function and the variance decomposition. To recap, these are:

$$y_t = F^{-1}u_t + C_1F^{-1}u_{t-1} + C_2F^{-1}u_{t-2} + C_3F^{-1}u_{t-3} + ... \qquad (24)$$

$$E(y_{t+s} - E_t y_{t+s})(y_{t+s} - E_t y_{t+s})' = F^{-1}D(F^{-1})' + C_1F^{-1}D(F^{-1})'C_1' + ... + C_{s-1}F^{-1}D(F^{-1})'C'_{s-1} \qquad (25)$$

Since there are necessarily differences between the coefficients of the SF shocks and RF shocks, it follows that the impulse-response patterns generated by (24) will differ significantly from that of (14). The same will be true for the SF and RF variance decompositions.

The imposition of long-run identifying restrictions on a VAR implies that L should be set equal to 1 in equation (30). In more compact notation, this means that:

$$y_t = [I - A(1)]^{-1}F^{-1}u_t$$

where we let

$$\theta(1) = [I - A(1)]^{-1}F^{-1} \quad \text{so that } y_t = \theta(1)u_t \qquad (61)$$

From (61), it follows that

$$[I - A(1)]^{-1} E(F^{-1}u_t u_t' F^{-1'})[I - A(1)]^{-1'} = [I - A(1)]^{-1} F^{-1} D F^{-1'} [I - A(1)]^{-1'}$$

$$= \theta(1) u_t u_t' \theta(1)' \tag{62}$$

so that since $\varepsilon_t = F^{-1}u_t$ and $\mathrm{cov}(\varepsilon_t) = \Omega$, we get the link between the RF and the SF

$$[I - A(1)]^{-1} \Omega [I - A(1)]^{-1'} = \theta(1) D \theta(1)' \tag{63}$$

Where the LHS of (63) comes from the RF estimates and the RHS of (63) contain the SF long-run responses of the elements of y to the elements of u, $\theta(1)$. $\theta(1)$ may be extracted from the LHS of (63) using the Cholesky decomposition. Imposing the restriction D = I means that the Cholesky decomposition may be applied to the LHS of (63) to obtain the lower triangular matrix of long-run responses to the structural shocks, $\theta(1)$, such that the LHS of (63) = $\theta(1)\theta(1)'$. In this case, since $\theta(1)$ has a recursive structure, with the first variable affected only by lags of itself and not affected in the long-run by the other variables, the second variable affected in the long-run by its own lags and lags of the first, but not the others, and so on and so forth. Alternatively, not imposing D = I means that $\theta(1)$ need not be lower triangular. Note that both alternatives do not require F to be lower triangular.

Note that the long-run structural VAR's allow the data to determine short-run dynamics, because only long-run restrictions are imposed.

Examples

a) Blanchard and Quah [1989]

Blanchard and Quah estimate a two variable structural VAR for the log of GDP and the log of the first difference of the unemployment rate. The idea was to determine whether shocks to the economy came primarily from the demand side or the supply side. The identification scheme used implied that demand shocks have a temporary impact on the economy, while supply side shocks have a temporary and permanent impact on the economy.

The output shocks are associated with structural demand side shocks, while the unemployment shocks are associated with structural supply-side shocks. Demand side shocks have a hump-shaped impact on output, while supply shocks have a reverse hump-shaped impact on output.

b) Gali [1992]

Gali's model reflects a standard IS-LM framework, augmented with a Phillips curve. Shocks to Δy, the change in log of real GDP, are denoted ε_s, aggregate supply shocks. Shocks to Δi, the change in the yield on three-month T-bills, are denoted ε_{ms}, money supply shocks. Shocks to $i - \Delta p$, are denoted ε_{md}, money demand shocks, while shocks to real money demand, $\Delta m - \Delta p$ are denoted ε_{is}, IS shocks. The structural form of the VMA is:

$$Z_t = D(L)\varepsilon_t \tag{64}$$

The counterpart reduced form VMA is:

$$Z_t = E(L)\upsilon_t \tag{65}$$

E(L) is assumed to be invertible, so that the reduced form VAR representation is:

$$E(L)^{-1}Z_t = \upsilon_t \tag{66}$$

Gali assumes that the reduced form shocks, υ_t can be expressed as linear combinations of the structural shocks, ε_t:

$$\upsilon_t = S\varepsilon_t \quad \text{so that}$$

$$S^{-1}\upsilon_t = \varepsilon_t \tag{67}$$

and so $S^{-1}\ E(L)^{-1}\ Z_t = \varepsilon_t$. Gali imposes the following identifying restrictions on his model:

a) The variances of the structural disturbances are normalized to unity: D= I;
b) The structural disturbances are mutually orthogonal;
c) $\varepsilon_{ms,t}$, $\varepsilon_{md,t}$, and $\varepsilon_{is,t}$ have no permanent effects on y;
d) $\varepsilon_{ms,t}$ and $\varepsilon_{md,t}$, have no contemporaneous effects on y; and
e) The demand for real balances is unaffected by contemporaneous price changes, given the nominal rate and output (homogeneity in money demand).

Given these just-identifying restrictions, the elements of S can be uniquely solved for using a nonlinear simplex search technique.

Gali concludes that the dynamic response of the US economy to various disturbances matches closely most of the qualitative predictions of the Phillips curve-augmented IS-LM framework. Supply factors contribute a large part of short-run GNP fluctuations.

c) Ahmed, Ickes, Wang and Yoo [1993]

This study has two objectives: (a) to extend the identified VAR literature to an open-economy setting; and (b) to study the role of exchange rate regimes by examining the relative transmission properties of fixed and floating exchange rates. Identification of fundamental disturbances using a structural VAR model allows the authors to examine whether there is increased volatility in flexible exchange rate periods and to determined whether they are due to changes in the manner in which the economic system responds to shocks or due to a change in the volatility of the underlying fundamental disturbances themselves.

The authors employ the long-run modeling technique of Blanchard and Quah. They argue that in order to obtain a model that nests both fixed- and floating exchange rate periods, it is desirable to employ only long-run identifying restrictions, since most economists consider differences, if any, in the transmission across exchange rate regimes to be important predominantly in the short-run. Thus, their analysis makes it possible to discriminate between competing theories that share their long-run identification assumptions but differ widely in terms of short-run predictions and the role of exchange rate regimes.

The authors construct a real business cycle model of the economy to justify the imposition of long-run identifying restrictions. The resulting impulse responses have the following structural form:

$$\begin{bmatrix} \Delta n_{ht} \\ \Delta y_{ht} \\ \Delta y_{ft} \\ \Delta y^P_{ht} - \Delta y^P_{ft} \\ \Delta q_{ft} \\ \Delta m_{ht} - \Delta m_{ft} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & a_{32}(1) & 1 & 0 & 0 & 0 \\ 1 & 1-a_{32}(1) & -1 & 1 & 0 & 0 \\ a_{51}(1) & a_{51}(1)(1-a_{32}(1)) & -a_{51}(1) & a_{51}(1) & a_{51}(1) & 0 \\ a_{61}(1) & a_{62}(1) & a_{63}(1) & a_{64}(1) & a_{65}(1) & 1 \end{bmatrix} \begin{bmatrix} \tau_{ht} \\ \tau_t \\ \tau_{ft} \\ (1-\phi_h)\eta_{ft} - (1-\phi_f)\eta_{ht} \\ \varepsilon_{ft} - \varepsilon_{ht} \\ (1-\pi_f)\upsilon_{ht} - (1-\pi_h)\upsilon_{ft} \end{bmatrix}$$

(68)

where the first equation in the system models the domestic labor input (being driven by the shock τ_{ht}). The second equation models the change in output in the domestic economy (driven by two shocks). The third equation models output in the foreign country as being driven by the domestic economy's shock, as well as an own shock. The fourth equation depicts growth rate of the ratio of private domestic output to private foreign output (driven by a linear combination of shocks). The fifth equation implies that the change in the log of the relative price of the foreign good in units of the domestic consumption good is driven by all shocks except a relative money shock. The last equation describes the workings of the money market. The ratio of money supplies in the two countries reacts permanently to the three supply shocks (τ, τ_h and τ_f) as well as the relative fiscal shock ($[1 - \phi_h]\ \eta_{ft} - [1 - \phi_f]\ \eta_{ht}$), relative demand shock ($\varepsilon_h - \varepsilon_f$) and a relative money shock ($[1 - \pi_f]\ \upsilon_{ht} - [1 - \pi_h]\ \upsilon_{ft}$). Thus, the restrictions embodied in this model are:

a) changes in labor input are exogenous in the long-run;
b) the long-term behavior of output is supply-determined (it is determined by exogenous shocks to technology and the labor input);
c) the ratio of government purchases to output is exogenously chosen by fiscal authorities in the long-run; and
d) long-run neutrality of money holds.

The authors use data from the US and the UK to estimate the model. They show that country-specific supply shocks are very important in generating international business cycles. They also show that although the post-1973 flexible exchange rate period has been inherently more volatile, there are no differences in transmission of properties of economic disturbances across exchange rate regimes for the endogenous variables that are the subject of the study.

d) Ahmed and Park [1994]

Ahmed and Park examine sources of economic fluctuations in small open economies. The long-run multipliers in the structural model are given by:

$$\begin{bmatrix} \Delta y_{ft} \\ \Delta y_t \\ B_t \\ \Delta P_t \end{bmatrix} = \begin{bmatrix} a_{11}(1) & 0 & 0 & 0 \\ a_{21}(1) & a_{22}(1) & 0 & 0 \\ a_{31}(1) & a_{32}(1) & a_{33}(1) & 0 \\ a_{41}(1) & a_{42}(1) & a_{43}(1) & a_{44}(1) \end{bmatrix} \begin{bmatrix} \eta_t \\ \varepsilon_t \\ \upsilon_t \\ u_t \end{bmatrix} \tag{69}$$

Equation (69) implies that the rest of the world output growth (Δy_{ft}) is exogenously given (i.e., driven by own shocks, η_t). Also, the change in domestic output growth (Δy_t) is driven by innovations in: world output growth, domestic supply shocks, ε_t, *but not* by changes in domestic absorption shocks, υ_t, and changes in domestic price level shocks (nominal shocks), u_t. Furthermore, the trade balance (B), is determined by innovations in η_t, ε_t, and υ_t, but not by u_t. Finally, the change in the price level (ΔP) is determined by innovations in η_t, ε_t, ν_t, and u_t. The restrictions on the long-run multiplier imply that:

a) The rest of the world output growth is exogenously given to the domestic country in the long-run;
b) The long-run neutrality of aggregate demand holds (since nominal shocks do not affect real ouptut growth); and
c) Nominal disturbances have no cumulative effect on the trade balance.

e) Whitt [1995]

Whitt employs a slightly-modified Blanchard and Quah model to examine the historical pattern of aggregate demand and supply shocks in number of EMS countries to assess the desirability of monetary union. Countries with similar patterns of shocks are presumably better candidates for monetary union than those affected by widely disparate shocks. Again, a structural VAR is estimated with the restriction that demand shocks have no permanent effect on real output.

The structural form of the VMA impulse-response functions may be written in matrix form as:

$$\begin{bmatrix} \Delta Y_t \\ \Delta P_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) \\ C_{21}(L) & C_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{st} \\ \varepsilon_{dt} \end{bmatrix} \tag{70}$$

where ΔY_t is the growth rate of real output (percentage change in real industrial production), and ΔP_t is the inflation rate (percentage change in producer prices). Whitt imposes the following restrictions on the C(1) matrix of long-run multipliers:

1) $C_{12}(1) = 0$: aggregate demand disturbances (ε_{dt}) have no long-run impacts on the change un output (Y_t)

Thus, C(1) is restricted to be lower triangular, and again, its elements may be recovered via a Cholesky decomposition of the covariance matrix of reduced form VAR residuals along the same lines as Blanchard and Quah.

In most cases, supply shocks are positively correlated with those of Germany, but the negative correlation of demand shocks suggests that monetary union may not be desirable. Britain is a poor candidate for union, since both its supply and demand shocks are negatively correlated with Germany's.

f) Thomas [1997]

Thomas uses a structural VAR representation of the Mundell-Fleming model to analyze movements in Sweden's real exchange rate. He shows that demand shocks account for a higher fraction of real shocks in Sweden than in several EMU-member countries. If real demand shocks are due to controllable macroeconomic policies, then the cost of relinquishing the exchange rate is no higher (maybe lower?) for Sweden than for most core EMU countries.

The long-run responses of the structural shocks to the three variables are shown by the following relation:

$$\begin{bmatrix} \Delta Y_t \\ \Delta RER_t \\ \Delta CPI_t \end{bmatrix} = \begin{bmatrix} C_{11}(1) & C_{12}(1) & C_{13}(1) \\ C_{21}(1) & C_{22}(1) & C_{23}(1) \\ C_{31}(1) & C_{32}(1) & C_{33}(1) \end{bmatrix} \begin{bmatrix} \varepsilon_s \\ \varepsilon_d \\ \varepsilon_f \end{bmatrix} \tag{71}$$

Thomas imposes the following restrictions:

2) C_{12} (1) = 0: aggregate demand disturbances (ε_d) have no long-run impacts on output (Y_t)
3) C_{13} (1) = 0: aggregate nominal disturbances (ε_f) have no long-run impacts on output
4) C_{23} (1) = 0: aggregate nominal disturbances (ε_f) have no long-run impacts on the real exchange rate (RER).

CPI_t is the consumer price index. Thus, C(1) is upper triangular.

g) Hoffmaister and Roldos [1997]

The authors compare business cycles in Asia and Latin America using SVAR analysis with panel data. The evidence for countries in these regions suggests that: (a) the main source of output fluctuations is supply shocks, even in the short run; (b) the real exchange rate is driven mostly by fiscal shocks; and (c) terms of trade shocks are important for trade balance fluctuations but not for output or real exchange rate fluctuations. However, in Latin America, as opposed to Asia, output is affected more by external and domestic demand shocks.

The impulse responses are of the form:

$$\begin{bmatrix} \Delta r^{*} \\ \Delta p^{*}_{mt} \\ \Delta y_t \\ \Delta q_t \\ \Delta p_t \end{bmatrix} = \begin{bmatrix} a_{11}(1) & 0 & 0 & 0 & 0 \\ a_{21}(1) & a_{22}(1) & 0 & 0 & 0 \\ a_{31}(1) & a_{32}(1) & a_{33}(1) & 0 & 0 \\ a_{41}(1) & a_{42}(1) & a_{43}(1) & a_{44}(1) & 0 \\ a_{51}(1) & a_{52}(1) & a_{53}(1) & a_{54}(1) & a_{55}(1) \end{bmatrix} \begin{bmatrix} \varepsilon^{i*}_t \\ \varepsilon^{tt}_{mt} \\ \varepsilon^{s}_t \\ \varepsilon^{f}_t \\ \varepsilon^{n}_t \end{bmatrix} \tag{72}$$

Note that the long-run dynamic multipliers are contained in the matrix A(1). We interpret this in the following manner:

a) The change in the world real interest rate (Δr^*) is only affected in the long-run by its own shocks, (ε_t^{i*}), and not by any other variable;
b) The change in the world price of intermediate inputs (Δr^*) is only affected in the long-run by the world real interest rate, (ε_t^{i*}), as well as the country's terms of trade, (ε_{mt}^{tt*});
c) The change in real output (Δy) is affected in the long-run by the world real interest rate, (ε_t^{i*}), as well as the country's terms of trade, (ε_{mt}^{tt*}), as well as supply shocks, (ε_t^{s});

d) The change in the real exchange rate (Δq) is affected in the long-run by the world real interest rate, (ε_t^{i*}), as well as the country's terms of trade, (ε_{mt}^{tt*}), as well as supply shocks, (ε_t^s); and
e) The change in the domestic price level (Δp) is affected in the long-run by all of the shocks.

Nominal shocks (ε_t^n) are shocks to either money or the nominal exchange rate. These are assumed to be neutral (with respect to the first four variables in the SVAR) in the long-run. The model captures the effects of nominal shocks via an unspecified equation for the evolution of the price level. The authors reason that, owing to the different exchange rate regimes in the sample of countries in the study, it is difficult to establish whether the evolution of the price level is determined by money supply, the nominal exchange rate, or both. The restrictions implied by the model as a whole are:

a) Domestic innovations do not effect external disturbances (the small open economy assumption);
b) Long-run economic restrictions: Nominal shocks have no long-run impact on the level of domestic output and the real exchange rate, fiscal shocks can have long-run effects on the real exchange rate, but not on output; and
c) The structural innovations are mutually orthogonal.

h) Cushman and Zha [1997]

VAR modeling and identification has produced its share of controversies in the past ten years. Recursive modeling of economies has been criticized for generating unusual results. On the one hand, VAR modeling of the US economy has shown that output and exchange rates respond to innovations in interest rates and monetary aggregates in ways consistent with traditional monetary analyses. However, this is not the case for small, open economies.

Sims [1992] analyzed five major industrial countries using VARs, using domestic interest rate innovations to represent monetary policy shocks. For some countries, positive domestic interest rate innovations are associated with persistent increases in domestic price levels (a "price puzzle") and depreciation in the domestic currency (an "exchange rate puzzle"). The same results were obtained in a similar paper for G-7 countries by Grilli and Roubini [1995]. Cushman and Zha [1997] suggest that what is needed to eliminate such puzzles is an identification scheme appropriate for small, open economies. Applied to Canada, both puzzles are absent, and the dynamic responses to the identified monetary policy shock are consistent with standard theory and highlight the exchange rate as a transmission mechanism.

Open economy VAR modeling appears to be at the frontier of VAR modeling. Recent examples of open economy VAR studies include those by Ahmed, Ickes, Wang and Yoo [1993], Ahmed and Park [1994], Hoffmaister and Roldos [1997], Cushman and Zha [1997], and Bardsen and Klovland [2000].

i) Keating and Nye [1999]

The authors estimate Blanchard and Quah's model using post-World War II and pre-World War I data on output and unemployment from G7 countries. They find that demand shocks typically explain more output variance in the post-World War II period than in the pre-World War I period. This result is consistent with the view that price adjustment was slower in the latter half of the twentieth century. A second finding is that supply shocks cause the unemployment rate to rise for some countries and fall for others. This result implies that at least two types of supply shocks are at work.

4. Other VAR issues

4.1 Choosing the appropriate lag length[8]

Appropriate lag length is commonly determined by using the Akaike information criterion (AIC) [1973]:

$$AIC(q) = \ln\frac{e'e}{T} + \frac{2q}{T} \qquad (73)$$

If some maximum Q is known, $Q \geq q$ can be chosen to minimize AIC(q). Other studies suggest other methods. Other methods for choosing lag length can be found in Sims [1980] and the Enders [1996].

4.2 How to construct confidence intervals for impulse-response functions[9]

Standard errors for the parameters, the impulse responses, and the variance decompositions are calculated using the Monte Carlo approach of Runkle [1987], which simulates the VAR model to generate distributions for these results. The actual residuals are sampled, and the sampled residuals are used as shocks to the estimated VAR. After the artificial series are generated, they are used to perform the same structural VAR analysis. After 200 replications of the model, standard errors are calculated for the parameter estimates, the impulse responses and the variance decompositions.

[8] Based on Greene [1997].
[9] Based on Keating [1992].

5. Conclusion

In spite of the lively debate on the merits of VARs, the field of macroeconomics has since 1980, been inundated with literature using the original version or modifications of Sims' time series modeling technique. The use of VARs made it possible for conventional time series techniques being utilized in the business cycle literature to be used in conjunction with traditional simultaneous equations modeling. This has allowed economists to discriminate and identify supply and demand shocks causing business cycles, determine the persistence of these shocks, identify the impacts of monetary policy interventions by the central bank, ascertain the monetary transmission mechanism, and determine whether US data supports the IS-LM model, among other applications. Recently, open economy extensions of VAR modeling have become prevalent in the literature, initiated by Sims [1992].

Because of the ability of VARs to produce an accounting of innovations that is useful for policy analysis and forecasting, as well as its later capacity to accommodate structural modeling, many researchers adopted VAR methods in their study *in spite of caveats from other economists.* Impulse-response functions and variance decompositions have all since become standard tools in macroeconomic analysis and standard lessons in graduate econometrics classes, necessitating the creation of VAR programming commands for many econometric software applications such as RATS, TSP Eviews and others. It appears that the most profound contribution of the VAR literature in the last twenty years is its contribution to the recognition that systems methods are superior to single equations methods in identifying the effects of monetary policy [Pagan and Robertson 1995]. In this regard, proper identification of the dynamic impacts of monetary policy entails carefully modeling the behavior of the monetary authorities and how they use their instruments to implement policy. Non-recursive VAR modeling has done much to refine the modeling of the monetary transmission mechanism in this regard.

Since VAR modeling appears here to stay for the meantime, it appears that economists should devote more time to refining the VAR technique to address remaining criticisms of the technique. The frontiers of current VAR analysis already see econometricians in this direction. There also appears to be a need to refine open economy VAR modeling. Pagan and Robertson [1995] suggest that an open economy VAR system requires the addition of a foreign interest rate (to allow for the possibility of uncovered interest parity). In a recursive system, the foreign interest rate would need to appear as the first variable, and the exchange rate will appear after the domestic interest rate (see Eichenbaum and Evans [1995]).

References

Ahmed, S., B. Ickes, P. Wang, and B. Sam Yoo [1993] "International business cycles." *American Economic Review* 83: 335-59.

_____, and J. Ha Park [1994] "Sources of macroeconomic fluctuations in small open economies." *Journal of Macroeconomics* 16:1-36.

Bardsen, G. and J. Klovland [2000] "Shaken or stirred? Financial deregulation and the monetary transmission mechanism in Norway." *Scandinavian Journal of Economics* 102: 563-83.

Barro, R. [1977] "Unanticipated money, growth, and unemployment in the United States." *American Economic Review* 67: 101-15.

Bayoumi, T. and B. Eichengreen. "Macroeconomic adjustment under Bretton Woods and the Post-Bretton-Woods Float: an impulse-response analysis." *The Economic Journal* 104: 813-27.

Bernanke, B. [1986] "Alternative explanations of the money-income correlation." *Carnegie-Rochester Conference Series on Public Policy* 25:49-100.

____, and A. Blinder [1992] "The federal funds rate and the channels of monetary transmission." *American Economic Review* 82: 901-21.

____, and I. Mihov [1997] "What does the Bundesbank target?" *European Economic Review* 41: 1025-53.

____, and ____ [1998] "Measuring monetary policy." *Quarterly Journal of Economics* 113:869-902.

Blanchard, O. and D. Quah [1989] "The dynamic effects of aggregate demand and aggregate supply disturbances." *American Economic Review* 79:655-73.

Cecchetti, S. [1995] "Distinguishing theories of the monetary transmission mechanism." *Federal Reserve Bank of St. Loius Review* 77:83-97.

Clarida, R. and J. Gali [1994] "Sources of real exchange rate fluctuations: how important are nominal shocks?" *Carnegie-Rochester Conference Series on Public Policies* 4:1-56.

Cochrane, J. [1998] "What do VARs mean? Measuring the output effects of monetary policy." *Journal of Monetary Economics* 41:277-300.

Cooley, T., S. LeRoy [1985] "Atheoretical macroeconomics: a critique." *Journal of Monetary Economics* 16:283-308.

Copelman, M., and A. Werner [1995] "The monetary transmission mechanism in Mexico." *Board of Governors of the Federal Reserve System International Finance Discussion Paper* 521.

Doan, T., R. Litterman, and C. Sims [1984] "Forecasting and conditional projection using realistic prior distributions." *Econometric Review* 3:1-144.

Eichenbaum, M. [1992] "Comments on "interpreting the macroeconomic time series facts: the effects of monetary policy." by C. Sims. *European Economic Review* 36:1001-11.

____, and C. Evans [1995] "Some empirical evidence on the effects of shocks to monetary policy on exchange rates." *Quarterly Journal of Economics*.

Faust, J., and E. Leeper [1994] "When do long-run identifying restrictions give reliable results?" *Federal Reserve Bank of Atlanta Working Paper* 94-2.

Gali, J. [1992] "How well does the IS-LM model fit postwar U.S. data?" *Quarterly Journal of Economics* 57:709-38.

Gamber, E., and F. Joutz [1993] "An application of estimating structural vector autoregression models with long-run restrictions." *Journal of Macroeconomics* 15:723-745.

Gordon, D. and E. Leeper [1993] "The dynamic impacts of monetary policy: an exercise in tentative identification." *Journal of Political Economy* 102:1228-47.

Greene, W. [1997] *Econometric Analysis.* 3rd Edition. Prentice Hall, Upper Saddle River, New Jersey.

Hakkio, C., and C. Morris [1984] "Vector autoregressions: a user's guide." *Federal Reserve Bank of Kansas City Working Paper* 84-10.

Hamilton, J. [1994] *Time series analysis*. Princeton University Press, Princeton, New Jersey.

Haslag, J., and S. Hein [1992] "Macroeconomic activity and monetary policy actions: some preliminary evidence." *Journal of Money, Credit and Banking* 24: 431-46.

____, and ____ [1995] "Does it matter how monetary policy is implemented?" *Journal of Monetary Economics* 35: 359-86.

Hoffmaister, A., and J. Roldos [1997] "Are business cycles different in Asia and Latin America?" *International Monetary Fund Working Paper* 97/9.

Horvath, J., M. Kandil, and S. Sharma [1996] "On the European monetary system: the spillover effects of German shocks and disinflation." *Manuscript, Southern Illinois University and University of Wisconsin-Milwaukee.*

Karras, G. [1993] "Sources of U.S. macroeconomic fluctuations: 1973-1989." *Journal of Macroeconomics* 15:47-68.

Keating, J. 1992. "Structural approaches to vector autoregressions." *Federal Reserve Bank of St. Loius Review* 74:37-57.

Keating, J. and J. Nye [1998] "The dynamic effects of aggregate demand and supply disturbances in the G7 countries." *Journal of Macroeconomics* 21:263-78.

____ [2000] "Macroeconomic modeling with asymmetric vector autoregressions." *Journal of Macroeconomics* 22:1-28.

Lastrapes, W. and G. Selgin [1995] "The liquidity effect: identifying short-run interest rate dynamics using long-run restrictions." *Journal of Macroeconomics* 17:387-404.

Leeper, E., and D. Gordon [1992] "In search of the liquidity effect." *Journal of Monetary Economics* 29:341-69.

Leeper, E., C. Sims, and Tao Zha [1996] "What does monetary policy do?" *Brookings Papers on Economic Activity* 2:1-63.

Levy, J. and I. Halikias [1997] "Aspects of the monetary transmission mechanism under exchange rate targeting: the case of France." *International Monetary Fund Working Paper* 97/44.

Litterman, R. and L. Weiss [1985] "Money, real interest rates, and output: a reinterpretation of postwar U.S. data." *Econometrica* 53:129-56.

McMillin, W [2001] "The effects of monetary policy shocks: comparing contemporaneous versus long-run identifying restrictions." *Southern Economic Journal* 67:618-636.

Pagan, A. and J. Robertson [1995] "Resolving the liquidity effect." *Federal Reserve Bank of St. Loius Review* 77:33-54.

Rudebusch, G. [1997] "Do measures of monetary policy in a VAR make sense?" Manuscript, Federal Reserve Bank of San Francisco.

Romer, C., and D. Romer [1989] "Does monetary policy matter? A new test in the spirit of Friedman and Schwatz." in O. Blanchard and S. Fischer, eds, *NBER Macroeconomics Annual 1989* 121-70. Cambridge, Mass.: MIT Press.

Runkle, D. [1987] "Vector autoregressions and reality." *Journal of Business and Economic Statistics* 5: 437-42.

Sarte, P. [1997 "On the identification of structural vector autoregressions." *Federal Reserve Bank of Richmond Economic Quarterly* 83: 45-67

Shapiro, M., and M. Watson [1988] "Sources of business cycle fluctuations." in Ss Fischer, ed., *NBER Macroeconomics Annual 1988.* Cambridge, Mass.: MIT Press.

Sims, C. [1972] "Money, income and causality." *American Economic Review* 65:540-52.

____. [1980] "Macroeconomics and reality." *Econometrica* 48:1-48.

____. [1986] "Are forecasting models usable for policy analysis?" *Quarterly Review of the Federal Reserve Bank of Minneapolis* 10 *(Winter)*: 2-16.

____. [1992] "Interpreting the macroeconomic time series facts: the effects of monetary policy." *European Economic Review* 36: 975-1000.

____. [1996] "Comment on Glenn Rudebusch's "Do measures of monetary policy in a VAR make sense?" Manuscript, Yale University.

Soderlind, P. [2000] *Lecture Notes in Econometrics.* Stockholm School of Economics.

Thomas, A. [1997] "Is the exchange rate a shock absorber? the case of Sweden." *International Monetary Fund Working Paper* 97/176.

Whitt, J., Jr. [1995] "European Monetary Union: evidence from structural VARs." *Federal Reserve Bank of Atlanta Working Paper* 95-1