

VECTOR AUTOREGRESSIONS: A USER'S GUIDE

Craig S. Hakkio and Charles S. Morris

**November 1984
RWP 84-10**

Research Division
Federal Reserve Bank of Kansas City

The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Kansas City or of the Federal Reserve System.

Both authors are economists with the Federal Reserve Bank of Kansas City.

Table of Contents

	<u>page</u>
1. VARS: A TECHNICAL DISCUSSION	2
a. Univariate time series models	2
b. Multivariate time series models	7
2. USES OF VARS	15
a. Forecasting with a VAR: unconditional and conditional	15
b. The moving average representation of a VAR	19
c. Exogeneity	20
d. Variance decomposition and exogeneity	24
3. A SIMPLE TWO-VARIABLE VAR EXAMPLE	26
4. ESTIMATION OF A SIMPLE VAR	33
a. Choice of lag length	33
b. Estimation of the VAR	35
c. Variance decomposition of the VAR	38
d. Forecasts using the VAR	39
5. THE RELATIONSHIP BETWEEN STANDARD STRUCTURAL ECONOMETRIC MODELS AND VAR MODELS	40
6. AN EXAMPLE OF VARS VERSUS STRUCTURAL MODELS	44
a. The data generating model	44
b. The structural model and a VAR	46
c. Estimation	48
d. Conclusions	51
7. SUMMARY AND CONCLUSIONS	53
TABLES	55
FIGURES	60
FOOTNOTES	68
REFERENCES	71
APPENDIX A: METHOD USED TO CHOOSE THE DGM PARAMETERS	76
APPENDIX B: COMPUTER PROGRAMS	78

Widespread concern has surfaced in recent years over the inability of large scale macroeconomic models to produce accurate forecasts. The poor predictive performance of the large macroeconomic models first appeared in the 1970s as most models predicted inflation and unemployment rates well below actual rates. More recently, most large models did not predict the strength with which the U.S. economy would rebound in 1983 from the recession that ended in 1982. For example, in February 1983 the Council of Economic Advisers predicted that real gross national product would grow 3.1 percent from the fourth quarter of 1982 to the fourth quarter of 1983 and that the unemployment rate including armed forces stationed in the United States would be 10.4 percent in the fourth quarter of 1983.^{1/} Actual fourth quarter over fourth quarter real GNP growth was 6.2 percent, while the unemployment rate fell to 8.4 percent by the fourth quarter of 1983.

Because of the large models' poor forecasting performance, the use of vector time series models for forecasting purposes has been proposed by some economists as an alternative to structural econometric models. Vector time series, or vector autoregressive, models are atheoretical models that use only the observed time series properties of the data to forecast economic variables. Because they are atheoretical, vector time series models can be estimated without imposing the prior restrictions that are necessary to identify and estimate structural econometric models. Thus, both Keynesians and Monetarists could use the same vector autoregression (VAR) to forecast various economic variables even though the two groups have a different view of the true structure of the economy. Furthermore, although a VAR cannot be used to make inferences about the structure of an economy, they can be used to estimate parameters of interest to policymakers, such as the interest elasticity of real income.

This article provides a theoretical and empirical introduction to vector autoregressive modeling. The first section provides a technical discussion of time series models in general, with special attention devoted to VARs. In section two, the uses of a VAR are discussed. A simple two variable VAR is presented in section three. In section four a VAR is estimated using U.S. data on money, income, and prices. Section five shows how the VAR is related to a standard structural econometric model. A set of Monte Carlo experiments is used in section six to evaluate the forecasting performance of the VAR relative to a simple structural model.

1. VARS: A TECHNICAL DISCUSSION

A vector autoregressive model is a fairly general multivariate time series model. Thus, it may be easiest to understand what a VAR is by first discussing a specific univariate time series model. The multivariate model will then be discussed as a generalization of the univariate model.

a. Univariate time series models

In a univariate time series model the past history of a single variable is used to model the behavior of that same series. In general, a variable at time t , y_t , is modeled as a function of past values of that variable plus current and past random error terms. The difference between univariate and multivariate time series models, therefore, is that in a multivariate model y_t is modeled as a function of current and past values of other variables and their random error terms as well.

The simplest univariate time series model is the first-order autoregressive, AR(1), model,

$$(1.1) \quad y_t = ay_{t-1} + u_t \quad |a| < 1$$

$$(1.2) \quad \begin{aligned} E(u_t) &= 0 \\ E(u_t^2) &= \sigma^2 \\ E(u_t u_s) &= 0, \quad t \neq s \\ E(y_t u_s) &= 0, \quad t < s. \end{aligned}$$

The AR(1) model expresses the current value of y as the sum of a fraction of the previous value of y and a random white noise disturbance term. The conditions for u to be a white noise process are given by equations (1.2). Because (1.1) does not include a constant term, y can be viewed as a variable with a zero mean or as the deviation of a variable from its non-zero mean.

The primary use of a univariate time series model such as equation (1.1) is to forecast future values of the variable y . In order to forecast future values of y , it is necessary to first specify the information set that is known at the time the forecast is made. If the information set at time $t-1$ contains y_{t-1} , then the conditional expectation of y_t at $t-1$ given y_{t-1} is

$$(1.3) \quad \begin{aligned} E(y_t | y_{t-1}) &= E(ay_{t-1} | y_{t-1}) + E(u_t | y_{t-1}) \\ &= ay_{t-1}. \end{aligned}$$

Given the information set at time $t-1$, the conditional expectation of y_t at time $t-1$, is an optimal, in a mean square error sense, linear forecast of y_t . If the information set at time $t-1$ does not contain any past values of y , then the unconditional expectation of y_t is

$$(1.4) \quad \begin{aligned} E(y_t) &= E(ay_{t-1}) + E(u_t) \\ &= 0. \end{aligned}$$

This result is most easily seen after writing y_t in terms of past error terms alone. Defining the lag operator L as $L^i x_t = x_{t-i}$, and letting Σ run from $i=0$ to ∞ , (1.1) may be rewritten as

$$\begin{aligned}y_t &= aLy_t + u_t \\(1-aL)y_t &= u_t \\y_t &= \left(\frac{1}{1-aL}\right)u_t \\&= (\Sigma a^i L^i)u_t \\&= \Sigma a^i u_{t-i}.\end{aligned}$$

Thus,

$$E(y_t) = \Sigma a^i E(u_{t-i}) = 0.$$

Because past values of the variable are usually known to the forecaster, the conditional expectation is generally used to forecast future values of y .

The above transformation of (1.1) to

$$(1.5) \quad y_t = \sum_{i=0}^{\infty} a^i u_{t-i}$$

is the equivalent moving average (MA) representation of the first-order autoregressive representation (1.1). The MA representation of (1.1) is defined if and only if y is a stationary process. Stationarity is imposed by including the condition that $|a| < 1$ in (1.1).^{2/} That is, the term $(1-aL)^{-1}$ equals $\sum_{i=0}^{\infty} a^i L^i$ if and only if $|a| < 1$. More generally, y is stationary if

and only if the root of the characteristic equation from (1.1),

$$(1.6) \quad 1-aL = 0,$$

is greater than one.^{3/}

Economic time series data are usually not stationary. Stationarity usually can be achieved by taking first or second differences of the raw data or of the natural logarithm of the raw data. Stationarity implies that the effect of a non-zero current error term on future values of y declines over time in the sense that y will converge toward its unconditional expected value.^{4/}

Univariate time series models usually take a more general form than the AR(1) model. The Wold decomposition theorem states that any stationary stochastic process may be written as the sum of a deterministic component and a stochastic moving average, possibly of infinite order, component. Assuming that there is no deterministic component, the stationary stochastic process, y , may be written as

$$(1.6) \quad y_t = b^*(L)u_t$$
$$b^*(L) = 1-b_1L-b_2L^2-\dots$$
$$E(u_t) = 0$$
$$E(u_t^2) = \sigma^2$$
$$E(u_t u_s) = 0, \quad t \neq s.$$

If the roots of the characteristic equation $b^*(L) = 1-b_1L-b_2L^2-\dots = 0$ lie outside the unit circle, the process (1.6) is said to be invertible and may be written as a pure AR model, possibly of infinite order,

$$(1.7) \quad a^*(L)y_t = u_t, \quad a^*(L) = b^*(L)^{-1}$$
$$a^*(L) = 1-a_1L-a_2L^2-\dots,$$

or $y_t = a(L)y_{t-1} + u_t$

$$a(L) = (1 - a^*(L))L^{-1} = a_1 + a_2L + a_3L^2 + \dots \underline{5/}$$

Thus, if y follows an AR(1) process, $a^*(L) = 1 - a_1L$ and in (1.6) $b_i = a_1^i$.

It is often the case that $b^*(L)$ may be written as the ratio of two relatively low order polynomials in the lag operator,

$$(1.8) \quad y_t = \frac{d^*(L)u_t}{c^*(L)}$$

$$c^*(L) = 1 - c_1L - c_2L^2 - \dots - c_pL^p$$

$$d^*(L) = 1 - d_1L - d_2L^2 - \dots - d_qL^q.$$

If y_t can be written as (1.8), then $b^*(L)$ is said to be a rational polynomial and y is said to follow an ARMA(p,q) process.6/ Equation (1.8) can be rewritten as

$$(1.9) \quad y_t = c(L)y_{t-1} + d(L)u_{t-1} + u_t$$

$$c(L) = (1 - c^*(L))L^{-1} = c_1 + c_2L + c_3L^2 + \dots + c_pL^{p-1}$$

$$d(L) = (d^*(L) - 1)L^{-1} = -d_1 - d_2L - d_3L^2 - \dots - d_qL^{q-1}$$

The optimal linear forecast of y_t given all past information about y is the conditional expectation of y given all past values of y ,

$$(1.10) \quad E_{t-1}y_t = E(y_t | y_{t-1}, y_{t-2}, \dots) = c(L)y_{t-1} + d(L)u_{t-1}.$$

The methods used to identify the numerical values of p and q and to estimate $c(L)$ and $d(L)$ may be found in Charles Nelson (1973).

Although (1.9) is a general univariate model, it is still quite limited. As shown in equation (1.10), only past information about y is used to forecast future values of y . For example, if y represents real GNP, forecasts of future real GNP are formed using past values of real GNP alone. The problem with this procedure is that other variables, say, interest rates and money, may be useful in predicting future values of real GNP. This problem is solved by using multivariate time series models.

b. Multivariate time series models

In a multivariate time series model the interaction between several variables is used to forecast each individual variable. Thus, in a multivariate time series model the forecast at time t of a variable y is a function of past values of itself and current and past values of all other variables in the system.

As with univariate time series models, the simplest multivariate time series model is a first-order autoregressive, VAR(1), model,

$$(1.11) \quad y_t = Ay_{t-1} + u_t$$

$$(1.12) \quad E(u_t) = 0$$

$$E(u_t u_t') = \Sigma$$

$$E(u_t u_s') = 0 \quad t \neq s$$

$$E(y_t u_s') = 0 \quad t < s,$$

where y is an $n \times 1$ vector of variables, A is an $n \times n$ matrix of coefficients, and u is an $n \times 1$ vector of white noise disturbance terms. As with the univariate model, if the information available at time $t-1$ is y_{t-1} , the conditional expectation of y_t given y_{t-1} is

$$(1.13) \quad E(y_t | y_{t-1}) = Ay_{t-1}.$$

Given the information set at time $t-1$, the conditional expectation of y_t at time $t-1$ is an optimal linear forecast of y_t .

If y is a stationary series, an equivalent vector moving average (VMA) representation of equation system (1.11) may be derived. The vector stochastic process y is stationary if and only if the roots of the characteristic equation

$$\det(I-AL) = 0$$

lie outside the unit circle.^{7/} The equivalent VMA representation of (1.11) is

$$(1.14) \quad y_t = \sum_{i=0}^{\infty} A^i u_{t-i}.$$
^{8/}

The Wold decomposition theorem applies to any stationary vector stochastic process, say, y . Assuming there is no deterministic component, y may be written as a stochastic VMA process, possibly of infinite order,

$$(1.15) \quad y_t = B^*(L)u_t$$

$$B^*(L) = I - B_1^*L - B_2^*L^2 - \dots$$

$$E(u_t) = 0$$

$$E(u_t u_t') = \Sigma$$

$$E(u_t u_s') = 0 \quad t \neq s.$$

If the roots of the characteristic equation

$$\det(B^*(L)) = 0$$

lie outside the unit circle, y_t is said to be an invertible process and may be written as a pure VAR model, possibly of infinite order,

$$(1.16) \quad A^*(L)y_t = u_t, \quad A^*(L) = B^*(L)^{-1}$$

$$A^*(L) = I - A_1L - A_2L^2 - \dots,$$

or

$$(1.17) \quad y_t = A(L)y_{t-1} + u_t$$

$$A(L) = (I - A^*(L))L^{-1} = A_1 + A_2L + A_3L^2 + \dots \quad \underline{9/}$$

Thus, if y follows a VAR(1) process, $A^*(L) = I - A_1L$ and in (1.15) $B_i = A_1^i$.

Finally, if $B^*(L) = D^*(L)/C^*(L)$,

$$(1.18) \quad y_t = \frac{D^*(L)u_t}{C^*(L)}$$

$$C^*(L) = I - C_1L - C_2L^2 - \dots - C_pL^p$$

$$D^*(L) = I - D_1L - D_2L^2 - \dots - D_qL^q,$$

then $B^*(L)$ is said to be a rational matrix polynomial and y is said to follow a vector ARMA(p,q), or VARMA(p,q), process. The optimal linear forecast of y_t given all past information about y is the conditional expectation of y given all past values of y ,

$$(1.19) \quad E_{t-1}y_t = E(y_t | y_{t-1}, y_{t-2}, \dots) = C(L)y_{t-1} + D(L)u_{t-1}$$

$$C(L) = (I - C^*(L))L^{-1} = C_1 + C_2L + C_3L^2 + \dots + C_pL^{p-1}$$

$$D(L) = (D^*(L) - I)L^{-1} = -D_1 - D_2L - D_3L^2 - \dots - D_qL^{q-1}$$

Although the theoretical development of multivariate time series models is quite similar to the theoretical development of univariate time series models, identifying the numerical values of p and q in multivariate models is much more difficult. Well documented procedures exist for identifying p and q in univariate models. However, not only is it difficult to identify p and q in VARMA models, but the robustness of the identification and estimation procedures known to date is unclear.^{10/}

To get around this problem, multivariate time series models are usually estimated as pure VAR(p) models. Recall that the VAR representation of a stochastic vector of variables exists only if the process is invertible. Invertibility implies that the elements of the coefficient matrices A_i in equation system (1.16) approach zero as i increases. Although $A^*(L)$ may be an infinite order polynomial in the lag operator L , a VAR(p) model will be a fairly accurate representation of the true model if p is chosen such that the elements of A_i , $i > p$, are close to zero.

Once p is chosen, the p -th order VAR may be estimated by ordinary least squares. Because the right-hand side variables in (1.17) are all past values of y , they are all predetermined. Thus, the system (1.17) can be consistently estimated using ordinary least squares without being concerned about the existence of simultaneous equations bias. Furthermore, estimating each equation separately using ordinary least squares produces asymptotically efficient estimates because the right-hand side variables are the same in every equation. Thus, VARs are easy to estimate because efficient and consistent estimates can be produced without using system estimation procedures. Although the relationship between VARs and standard econometric models will be discussed in section three, it is useful to note that in terms of the standard econometric literature, the system of equations (1.17) is the reduced form of some underlying structural system of equations.

Another useful transformation of (1.17) is to write the system of equations as a classical recursive system. From (1.15), the contemporaneous covariance matrix of u is

$$E(u_t u_t') = \Sigma.$$

Using the Cholesky decomposition of Σ , the covariance matrix may be written as

$$\Sigma = H^{-1} H^{-1'}$$

or

$$H \Sigma H' = I,$$

where H^{-1} is lower triangular. Because H is also lower triangular, premultiplying (1.17) by H results in the recursive system

$$H y_t = H A(L) y_{t-1} + H u_t$$

or

$$(1.20) \quad H y_t = A^{**}(L) y_{t-1} + e_t$$

$$A^{**}(L) = H A(L) = H A_1 + H A_2 L + H A_3 L^2 + \dots$$

$$e_t = H u_t$$

$$E(e_t) = E(H u_t) = 0$$

$$E(e_t e_t') = E(H u_t u_t' H') = H \Sigma H' = I$$

$$E(e_t e_s') = E(H u_t u_s' H') = H O H' = 0 \quad t \neq s. \underline{11/}$$

Thus, the equation for y_{it} contains contemporaneous values of y_j , $j < i$, and the vector error term is a vector white noise process, with the variance of each component normalized to equal one and the contemporaneous covariance between all individual error terms in e equal to zero. That is,

$$\begin{aligned}
 h_{11}y_{1t} &= \sum_{s=1}^{\infty} (a_{11,s}y_{1t-s} + a_{12,s}y_{2t-s} + \dots + a_{1n,s}y_{nt-s}) + e_{1t} \\
 h_{22}y_{2t} &= -h_{21}y_{1t} + \sum_{s=1}^{\infty} (a_{21,s}y_{1t-s} + a_{22,s}y_{2t-s} + \dots + a_{2n,s}y_{nt-s}) + e_{2t} \\
 &\vdots \\
 h_{ii}y_{it} &= -\sum_{j=1}^{i-1} h_{ij}y_{jt} + \sum_{s=1}^{\infty} (a_{i1,s}y_{1t-s} + a_{i2,s}y_{2t-s} + \dots + a_{in,s}y_{nt-s}) + e_{it} \\
 &\vdots \\
 h_{nn}y_{nt} &= -\sum_{j=1}^{n-1} h_{nj}y_{jt} + \sum_{s=1}^{\infty} (a_{n1,s}y_{1t-s} + a_{n2,s}y_{2t-s} + \dots + a_{nn,s}y_{nt-s}) + e_{nt}
 \end{aligned}$$

where

$$\begin{aligned}
 e_{1t} &= h_{11}u_{1t} \\
 e_{2t} &= h_{21}u_{1t} + h_{22}u_{2t} \\
 &\vdots \\
 e_{it} &= h_{i1}u_{1t} + h_{i2}u_{2t} + \dots + h_{ii}u_{it} \\
 &\vdots \\
 e_{nt} &= h_{n1}u_{1t} + h_{n2}u_{2t} + \dots + h_{nn}u_{nt}
 \end{aligned}$$

Although both u_{it} and e_{it} are referred to as the error or innovation in y_i , clearly they are not equal to each other unless $H=I$ --which is true only if $\Sigma = I$ to start with. The difference between e and u is that the information set used to calculate u in (1.17) is different from that used to calculate e in (1.20). From (1.17)

$$u_{1t} = y_{1t} - E(y_{1t} | y_{t-1}, y_{t-2}, \dots)$$

$$u_{2t} = y_{2t} - E(y_{2t} | y_{t-1}, y_{t-2}, \dots)$$

$$u_{nt} = y_{nt} - E(y_{nt} | y_{t-1}, y_{t-2}, \dots),$$

whereas from (1.20)

$$e_{1t} = y_{1t} - E(y_{1t} | y_{t-1}, y_{t-2}, \dots)$$

$$e_{2t} = y_{2t} - E(y_{2t} | y_{1t}, y_{t-1}, y_{t-2}, \dots)$$

$$e_{nt} = y_{nt} - E(y_{nt} | y_{1t}, y_{2t}, \dots, y_{n-1t}, y_{t-1}, y_{t-2}, \dots).$$

That is, u_{it} is the error in predicting y_{it} when the information set used to predict y_{it} includes all past values of the vector y , whereas e_{it} is the error in predicting y_{it} when the information set used to predict y_{it} includes all past values of the vector y and current values of y_j for all $j < i$.

Although u is invariant to the ordering of the variables in y in (1.17) for a given sample of data, e is not invariant to the ordering of the variables in y in (1.20) because different orderings lead to different information sets being used to predict y_{it} in the triangularized system. Technically, different orderings lead to different H matrices and to different linear combinations of u_{jt} , $j=1,2,\dots,i-1$, in e_{it} .

The corresponding VMA representation of the recursive VAR in (1.20) is

$$Hy_t = HA(L)y_{t-1} + Hu_t$$

$$H(I-A(L)L)y_t = Hu_t$$

$$y_t = (I-A(L)L)^{-1}H^{-1}Hu_t$$

or

$$(1.21) \quad y_t = B(L)e_t$$

$$B(L) = (I - A(L)L)^{-1}H^{-1} = B^*(L)H^{-1} = H^{-1} - B_1L - B_2L^2 - \dots$$

$$e_t = H^{-1}u_t$$

The VMA representation (1.21) is recursive with respect to the current error term, where the error term is a white noise process in which all contemporaneous covariance terms are zero and the variance of each element of the noise vector is normalized to equal one. That is,

$$(1.22) \quad y_{1t} = h_{11}^* e_{1t} - \sum (b_{11,se1t-s} + b_{12,se2t-s} + \dots + b_{1n,se nt-s})$$

$$y_{2t} = h_{21}^* e_{1t} + h_{22}^* e_{2t} - \sum (b_{21,se1t-s} + b_{22,se2t-s} + \dots + b_{2n,se nt-s})$$

$$\vdots$$

$$y_{it} = h_{i1}^* e_{1t} + h_{i2}^* e_{2t} + \dots + h_{ii}^* e_{it} - \sum (b_{i1,se1t-s} + b_{i2,se2t-s} + \dots + b_{in,se nt-s})$$

$$\vdots$$

$$y_{nt} = h_{n1}^* e_{1t} + h_{n2}^* e_{2t} + \dots + h_{nn}^* e_{nt} - \sum (b_{n1,se1t-s} + b_{n2,se2t-s} + \dots + b_{nn,se nt-s}),$$

where \sum runs from $s=1$ to ∞ and h_{ij}^* is the (i,j) element of H^{-1} . Note that the equation for y_{it} contains its own contemporaneous error term as well as the error term from the y_{jt} equation, $j < i$.

2. USES OF VARS

Four uses of VARS are discussed in this section. First, forecasting with a VAR is discussed. Second, the moving average representation of a VAR is discussed. Third, conditions for the exogeneity of a subset of variables are derived. Fourth, the variance decomposition of a VAR is discussed.

a. Forecasting with a VAR: unconditional and conditional

VARs are often used as an alternative to structural econometric models for forecasting purposes. Once the lag length of the vector autoregressive representation of a vector of variables, y , is chosen, the VAR(p) model may be written as

$$(2.1) \quad A^*(L)y_t = u_t$$
$$A^*(L) = I - A_1L - A_2L^2 - \dots - A_pL^p$$
$$E(u_t) = 0$$
$$E(u_t u_t') = \Sigma$$
$$E(u_t u_s') = 0 \quad t \neq s$$
$$E(y_t u_s') = 0 \quad t < s.$$

Assume the relevant information set available at time t , I_t , consists of all current and past values of the vector y ,

$$I_t = y_t, y_{t-1}, \dots .$$

The optimal linear forecast of y_{t+1} , given the information available at time t , I_t , is the conditional expectation of y_{t+1} given I_t . This is just the linear least squares projection of y_{t+1} on I_t ,

$$(2.2) \quad E(y_{t+1}|I_t) = E_t y_{t+1} = A(L)y_t$$
$$A(L) = (I - A^*(L))L^{-1} = A_1 + A_2L + A_3L^2 + \dots + A_pL^{p-1},$$

where E_t is the conditional expectation operator, conditional on all information available at time t . Equation system (2.1) can be updated and rewritten as

$$(2.3) \quad y_{t+1} = A(L)y_t + u_{t+1}.$$

Substituting (2.2) into (2.3), results in

$$(2.4) \quad y_{t+1} = E_t y_{t+1} + u_{t+1}.$$

Thus, using (2.2), the system of equations (2.3) can be orthogonally decomposed into anticipated and unanticipated components, where $E_t y_{t+1}$ is the anticipated or forecasted component of y_{t+1} and u_{t+1} is the unanticipated component.

The chain rule of forecasting can be used to forecast any future value of y . Updating equation (2.3) results in

$$y_{t+k} = A(L)y_{t+k-1} + u_{t+k}.$$

Taking the expectation of y_{t+k} conditional on information available at time t , and noting that $E_t u_{t+k} = 0$, yields the recursive expectation formula,

$$(2.5) \quad E_t y_{t+k} = A(L)E_t y_{t+k-1}.$$

The k-step ahead forecast of y can be equivalently written as

$$(2.6) \quad E_t y_{t+k} = A(L)^k y_t.$$

This is derived from writing y_{t+k} as

$$\begin{aligned} y_{t+k} &= A(L)y_{t+k-1} + u_{t+k} \\ &= A(L)(A(L)y_{t+k-2} + u_{t+k-1}) + u_{t+k} \\ &= A(L)^k y_t + A(L)^{k-1} u_{t+1} + A(L)^{k-2} u_{t+2} \\ &\quad + \dots + A(L) u_{t+k-1} + u_{t+k} \end{aligned}$$

and taking the conditional expectation of y. Similarly,

$$(2.7) \quad E_t y_{t+k-1} = A(L)^{k-1} y_t.$$

Equation (2.6) can be rewritten as

$$(2.8) \quad E_t y_{t+k} = A(L) A(L)^{k-1} y_t.$$

Substituting (2.7) into (2.8) results in the recursive expectation formula,

$$(2.5) \quad E_t y_{t+k} = A(L) E_t y_{t+k-1}.$$

Thus, for a given information set, (2.5) is a simple recursive rule for determining the optimal linear forecast of the vector y_{t+k} , $k > 0$. That is, once a VAR is estimated it can be used to provide optimal linear forecasts of the variables of interest.

Although it is an unfortunate terminology, the conditional expectation of y_{t+k} at time t conditional on all information available at time t is referred to as the unconditional forecast of y_{t+k} . That is, the term conditional takes on a different meaning when it modifies "forecast" than when it modifies "expectation."

Strictly speaking, VARs can only be used to make unconditional forecasts. The problem with this is that if a subset of the vector y , say y^w , are policy variables, policymakers are interested in the forecast of y_{t+k} given all past information about y at time t and given alternative future paths of the variables $y^w = y_{t+1}^w, y_{t+2}^w, \dots, y_{t+k}^w$. According to the Lucas critique, if any path other than the unconditional forecast is assigned to future values of y^w , the coefficients in $A(L)$ will generally change. Because the coefficients in $A(L)$ change, conditional forecasts--forecasts in which specific future paths are chosen for y^w --cannot be made with a VAR.^{12/} Thus, strictly speaking, VARs cannot be used for policy analysis because policy analysis utilizes conditional forecasts to choose among a set of policy alternatives.

There are conditions, however, in which the Lucas critique is not very damaging in the sense that conditional forecasts are reasonably accurate. If proposed future values of y^w are similar to realized past values of y^w , the coefficients in $A(L)$ will not change much so that the conditional forecast will be fairly accurate. Furthermore, if economic agents perceive policy changes with a lag, then in the short run $A(L)$ will not change. In that case, conditional forecasts will be quite accurate in the short run. Because changes in policy rarely deviate by large amounts from previous policies, VARs should be a relatively accurate forecasting tool.^{13/}

b. The moving average representation of a VAR

The equivalent moving average representation of a VAR can be used to forecast the response of the system of variables y to an unanticipated change in any of the component variables of y . Equation system (1.21), which is a compact notation for (1.22), is rewritten below as (2.9) for convenience,

$$(2.9) \quad \begin{aligned} y_t &= B(L)e_t \\ B(L) &= H^{-1} + B_1L + B_2L^2 + \dots \\ E(e_t) &= 0 \\ E(e_t e_t') &= I \\ E(e_t e_s') &= 0 \quad t \neq s, \end{aligned}$$

where H^{-1} is a lower triangular matrix. The i th element of e_t , e_{it} , represents the unanticipated component of the i th element of y , y_{it} , at time t --that is, e_{it} is the part of y_{it} that can not be predicted when the information set contains current values of y_j , $j < i$, and all past values of the entire vector y . The coefficient matrix $B(L)$ represents the response of the system to a one standard error shock, or innovation, in e_t . The typical element of $B(L)$, $b_{ij}(L)$, is the response of all future values of y_i to a one standard error one-time current innovation in y_j . Since $b_{ij}(L)$ is the impulse response function of y_i with respect to a shock in y_j , the j -th column of $B(L)$, $b_{\cdot j}(L)$, is the impulse response function of the entire vector y with respect to a shock in y_j . In other words, $b_{\cdot j}(L)$ describes the typical response (and the optimal linear forecast) of the vector y to a normalized (one standard error) one-time innovation in the j -th component of y , e_j . Because the errors, say, e_{jt} and e_{it} , $j \neq i$, are not correlated with each other by construction, the j -th column of $B(L)$ is the response of the vector y to the shock in y_j alone. Thus, there is no need to be concerned about feedback between errors across equations.

c. Exogeneity

The exogeneity of a set of variables with respect to another set can be defined in terms of zero-restrictions on the coefficient matrices $C^*(L)$ and $D^*(L)$ in the VARMA (p,q) model (1.18). The VARMA (p,q) model (1.18) can be partitioned as

$$(2.10) \quad \begin{bmatrix} * & * \\ C_{11}(L) & C_{12}(L) \\ * & * \\ C_{21}(L) & C_{22}(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} * & * \\ D_{11}(L) & D_{12}(L) \\ * & * \\ D_{21}(L) & D_{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix},$$

where $E(u_t) = 0$
 $E(u_t u_t') = \Sigma$
 $E(u_t u_s') = 0 \quad t \neq s,$

and y_1 is $n_1 \times 1$, y_2 is $n_2 \times 1$, C_{11}^* is $n_1 \times n_1$, C_{22}^* is $n_2 \times n_2$, C_{21}^* is $n_2 \times n_1$, C_{12}^* is $n_1 \times n_2$, and D_{ij}^* is the same dimension as C_{ij}^* , $i, j = 1, 2$. The set of variables y_1 is said to be exogenous with respect to the set of variables y_2 if and only if $C_{12}^*(L) = 0$, $D_{12}^*(L) = 0$, and $D_{21}^*(L) = 0$. Intuitively, y_1 is exogenous with respect to y_2 because y_{1t} is not affected directly or indirectly by lagged values of y_2 . The condition $C_{12}^*(L) = 0$ implies that lagged values of y_2 do not directly enter the y_{1t} equation. The conditions $D_{12}^*(L) = 0$ and $D_{21}^*(L) = 0$ imply that lagged values of y_2 do not indirectly enter the y_{1t} equation through the error term.

The exogeneity conditions in terms of the restrictions on $C^*(L)$ and $D^*(L)$ in (2.10) have implications for the coefficient matrices in both the VAR and VMA representations of (2.10). If the exogeneity conditions are imposed on (2.10), the VAR representation of (2.10) is,

$$(2.11) \quad \begin{bmatrix} D_{11}^*(L)^{-1}C_{11}^*(L) & 0 \\ D_{22}^*(L)^{-1}C_{21}^*(L) & D_{22}^*(L)^{-1}C_{22}^*(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} .$$

Partitioning the VAR representation (1.16) conformably and imposing the restrictions implied by (2.11) on $A^*(L)$ results in

$$\begin{bmatrix} A_{11}^*(L) & 0 \\ A_{21}^*(L) & A_{22}^*(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} .$$

Thus, in terms of the VAR representation, y_1 is exogenous with respect to y_2 if and only if $A_{12}^*(L) = 0$. The VMA representation of (2.10), after imposing the exogeneity conditions, is

$$(2.12) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} C_{11}^{1*}(L)D_{11}^*(L) & 0 \\ C_{21}^{2*}(L)D_{11}^*(L) & C_{22}^{2*}(L)D_{22}^*(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} ,$$

where $C_{ij}^*(L)$ is the (i,j) block in $C^*(L)^{-1}$. Partitioning the VMA representation (1.15) conformably and imposing the restrictions implied by (2.12) on $B^*(L)$ results in

$$(2.13) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} B_{11}^*(L) & 0 \\ B_{21}^*(L) & B_{22}^*(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} .$$

Thus, in terms of the VMA representation, y_1 is exogenous with respect to y_2 if and only if $B_{12}^*(L) = 0$.

Although y_1 is often defined to be exogenous with respect to y_2 if and only if $C_{12}^*(L) = D_{12}^*(L) = D_{21}^*(L) = 0$ and $E(u_{1t}u_{2t}) \equiv \Sigma_{12} = 0$, the condition

$\Sigma_{12} = 0$ is not necessary. The reason the condition $\Sigma_{12} = 0$ is not necessary is that the time series representation for y_t can always be normalized such that there is no contemporaneous correlation between any of the error terms. This is most easily shown using the VMA representation (2.13). For a given covariance matrix Σ in (2.13), there exists a lower triangular matrix P such that $P\Sigma P' = I$. Equation (2.13) may be rewritten as

$$(2.14) \quad \begin{aligned} y_t &= [B^*(L)P^{-1}][Pu_t] \\ &\equiv \tilde{B}^*(L)\tilde{u}_t \\ E(\tilde{u}_t\tilde{u}_t') &= E(Pu_tu_t'P') = P\Sigma P' = I. \end{aligned}$$

Because P^{-1} is lower triangular, $\tilde{B}^*(L)$ is lower triangular if and only if $B^*(L)$ is lower triangular. That is, if $E(u_{1t}u_{2t}') \neq 0$, there exists a normalization such that $\tilde{B}_{12}^*(L) = 0$ and $E(\tilde{u}_{1t}\tilde{u}_{2t}') = 0$ if and only if $B_{12}^*(L) = 0$. The condition $E(u_{1t}u_{2t}') = 0$ need not be imposed because it can always be met through an appropriate normalization that will result in a lower triangular matrix $\tilde{B}^*(L)$ if and only if $B^*(L)$ is lower triangular. Therefore, a necessary and sufficient condition for y_1 to be exogenous with respect to y_2 is that the matrix of coefficients on u_{2t} in the y_{1t} equation of (2.13), $B_{12}^*(L)$, is zero or that the matrix of coefficients on \tilde{u}_{2t} in the y_{1t} equation of (2.14), $\tilde{B}_{12}^*(L)$, is zero.

Although $\tilde{B}_{12}^*(L) = 0$ is necessary and sufficient for y_1 to be exogenous with respect to y_2 , $\tilde{B}_{21}^*(L) = 0$ is sufficient but not necessary for y_2 to be exogenous with respect to y_1 . In terms of (2.10), y_2 is exogenous with respect to y_1 if and only if $C_{21}^*(L) = D_{12}^*(L) = D_{21}^*(L) = 0$, which implies $\tilde{B}_{21}^*(L) = 0$ in (2.13). That is, y_2 is exogenous with respect to y_1 if and only if $B^*(L)$ is upper triangular. Now assume that y_2 is exogenous with respect to

y_1 so that $B^*(L)$ is upper triangular. Since P^{-1} is lower triangular, $\tilde{B}^*(L)$ will not be upper triangular unless P^{-1} is also a diagonal matrix. Thus, $\tilde{B}^*(L)$ is upper triangular if and only if $B^*(L)$ is upper triangle and P is diagonal. Therefore, if $\tilde{B}^*(L)$ is upper triangular, then $B^*(L)$ is upper triangular and y_2 is exogenous with respect to y_1 . But if $\tilde{B}^*(L)$ is not upper triangular, y_2 may be exogenous with respect to y_1 because $B^*(L)$ may be upper triangular. That is, $\tilde{B}_{21}^*(L) = 0$ is sufficient but not necessary for y_2 to be exogenous with respect to y_1 .

Finally, the VMA representation (2.9) can be used to determine the exogeneity of a set of variables with respect to another set. The VMA representation (2.9) is

$$y_t = B(L)e_t$$

where $B(L) \equiv B^*(L)H^{-1}$, $e_t \equiv Hu_t$, and H is lower triangular. This is simply (2.14) with $P=H$, $\tilde{B}^*(L) = B(L)$, and $\tilde{u}_t = e_t$. Partitioning (2.9) as

$$(2.15) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix},$$

y_1 is exogenous with respect to y_2 if and only if $B_{12}(L) = 0$. That is, y_1 is exogenous with respect to y_2 if and only if the typical element of $B(L)$, $b_{ij}(L)$, equals zero for $i=1,2, \dots, n_1$ and $j = n_1+1, n_1+2, \dots, n$; in other words, if and only if the impulse response functions that relate y_1 to e_2 are identically zero. It also follows that the condition $B_{21}(L) = 0$, that is, $b_{ij}(L) = 0$ for $i = n_1+1, n_1+2, \dots, n$ and $j = 1, 2, \dots, n_1$, is sufficient but not necessary for y_2 to be exogenous with respect to y_1 .

d. Variance decomposition and exogeneity

The degree to which a set of variables is considered exogenous with respect to another set of variables can be determined using the VMA representation of a VAR model by computing the percentage of the expected k-period-ahead squared prediction error of a variable produced by an innovation in another variable. Using representation (2.9), the *i*th element of *y* at time *t+k* is

$$(2.16) \quad y_{i,t+k} = \sum_{j=1}^n \sum_{s=0}^{\infty} b_{ij,s} e_{j,t+k-s}.$$

The conditional expectation of $y_{i,t+k}$, conditional on all information available at time *t* is

$$(2.17) \quad E_t y_{i,t+k} = \sum_{j=1}^n \sum_{s=k}^{\infty} b_{ij,s} e_{j,t+k-s}.$$

Subtracting (2.17) from (2.16) results in the k-period-ahead conditional prediction error of $y_{i,t+k}$, conditional on all information available at time *t*,

$$y_{i,t+k} - E_t y_{i,t+k} = \sum_{j=1}^n \sum_{s=0}^{k-1} b_{ij,s} e_{j,t+k-s}.$$

Because e_{it} is uncorrelated with e_{jt} , $i \neq j$, the expected k-period ahead squared prediction error of $y_{i,t+k}$ conditional on information available at time *t* is,

$$E(y_{i,t+k} - E_t y_{i,t+k})^2 = \sum_{j=1}^n \sum_{s=0}^{k-1} b_{ij,s}^2 E e_{j,t+k-s}^2$$

or

$$(2.18) \quad E(y_{i,t+k} - E_t y_{i,t+k})^2 = \sum_{j=1}^n \sum_{s=0}^{k-1} b_{ij,s}^2$$

because $E_t^2 e_{j,t+k-s} = 1$ by construction. Note that (2.18) is also the k-period-ahead forecast error variance of $y_{i,t}$. From (2.18), the part of the expected k-period-ahead squared prediction error of y_i produced by the innovation in $y_{j,t}$, $e_{j,t}$, is

$$(2.19) \quad \sum_{s=0}^{k-1} b_{ij,s}^2$$

Therefore, the percentage of the expected k-period-ahead squared prediction error of $y_{i,t}$, $E(y_{i,t+k} - E_t y_{i,t+k})^2$, produced by an innovation in $y_{j,t}$, written as $PCNT(k, y_i; y_j)$, is

$$(2.20) \quad PCNT(k, y_i; y_j) = \frac{\sum_{s=0}^{k-1} b_{ij,s}^2}{\sum_{j=1}^n \sum_{s=0}^{k-1} b_{ij,s}^2} .$$

The value of $PCNT(k, y_i; y_j)$ can be used to determine whether y_i is exogenous with respect to y_j . To match the definition of exogeneity given by the condition $B_{12}(L) = 0$ in the VMA representation (2.15), let y_i be the first variable of the vector y --that is, let $y_i = y_1$. The necessary and sufficient condition for y_1 to be exogenous with respect to the remaining variables in the vector y , y_2 , is $B_{12}(L) = 0$ in (2.15), which is equivalent to the condition that $PCNT(k, y_1; y_j) = 0$ for all k greater than zero and all j greater than one. Therefore, y_1 is exogenous with respect to y_2 if and only if $PCNT(k, y_1; y_j) = 0$ for all $k > 0$ and all $j > 1$.^{14/} It also follows that y_j , $j > 1$, is exogenous with respect to y_1 if $PCNT(k, y_j; y_1) = 0$ for all k

greater than zero, while y_j might be exogenous with respect to y_1 if $PCNT(k, y_j; y_1) \neq 0$ for all k greater than zero. That is, $PCNT(k, y_j; y_1) = 0$ is sufficient but not necessary for y_j to be exogenous with respect to y_1 .

Up to this point, the discussion of VARs has been quite general. To help understand this section and to see how VARs are used, a simple two variable VAR is used to discuss the concepts in sections one and two.

3. A SIMPLE TWO-VARIABLE VAR EXAMPLE

This section discusses the results of sections one and two in the context of a simple two-variable VAR, with $p=1$. The VAR can be written as

$$(3.1) \quad x_t = ax_{t-1} + by_{t-1} + u_t$$

$$(3.2) \quad y_t = cx_{t-1} + dy_{t-1} + v_t$$

$$(3.3) \quad E \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u_t & v_t \end{bmatrix} = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} .$$

The moving average representation can be obtained by recursive substitution, to obtain

$$(3.4) \quad x_t = u_t + au_{t-1} + (a^2 + bc)u_{t-2} + \dots \\ + bv_{t-1} + (ab + bd)v_{t-2} + \dots$$

$$(3.5) \quad y_t = cu_{t-1} + (ac + cd)u_{t-2} + \dots \\ + v_t + dv_{t-1} + (bc + d^2)v_{t-2} + \dots .$$

Consider the effect of an x innovation, an increase in u_t . According to (3.4), if u_t increases by 1 unit then x_t increases by 1 unit. In addition, since u_t and v_t are correlated, a change in u_t will be associated with a change in v_t on average, which causes y_t to change (according to equation (3.5)). For simplicity, assume the variance of u_t and v_t equal 1 and the covariance (correlation) equals r . Then, if u_t increases by 1, v_t will, on average, increase by r and y_t will increase by r . Therefore, because u and v are contemporaneously correlated, changes in u and v occur simultaneously so that a change in v cannot be attributed to a pure y innovation.

How does one decompose this contemporaneous correlation? Write (3.1) and (3.2) as a vector equation,

$$(3.6) \quad \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

$$E \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u_t & v_t \end{bmatrix} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} = \Sigma .$$

Consider the Cholesky decomposition of $\Sigma : H^{-1}H^{-1'} = \Sigma$, where H^{-1} is lower triangular. Then

$$H^{-1} = \begin{bmatrix} 1 & 0 \\ r & \sqrt{1-r^2} \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 \\ -r & 1 \\ \hline \sqrt{1-r^2} & \sqrt{1-r^2} \end{bmatrix} .$$

Next, multiply both sides of (3.6) by H to obtain

$$(3.8) \quad x_t = ax_{t-1} + by_{t-1} + e1_t$$

$$(3.9) \quad y_t = rx_t + (c-ra)x_{t-1} + (d-rb)y_{t-1} + \sqrt{1-r^2} e2_t$$

$$(3.10) \quad e1_t = u_t$$

$$(3.11) \quad e2_t = (-ru_t + v_t)/\sqrt{1-r^2} \quad .$$

Notice that $\text{cov}(e1, e2) = 0$. Equations (3.10)-(3.11) correspond to the matrix equation (1.20), where

$$\begin{bmatrix} e1_t \\ e2_t \end{bmatrix} = H \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad .$$

Equations (3.8)-(3.11) constitute a recursive system of equations: x and y are jointly endogenous; there are no RHS endogenous variables in (3.8); x_t appears in the y_t equation; the correlation between the errors, $e1$ and $e2$, equals 0.

Consider the errors $e1$ and $e2$. Suppose u_t increases by 1 unit, so that $e1$ increases by 1. An increase in u of 1 unit will, on average, increase v by r units (since the correlation= r). According to (3.11), since u increases by 1, $-ru$ decreases by r and v increases by r . Thus, there is no change in $e2$. That is, $e2$ has been constructed so that $e1$ and $e2$ are uncorrelated, which was the purpose of multiplying the error vector in (3.6) by H . In addition, according to (3.10)-(3.11), the variance of $e1$ and $e2$ equals 1, regardless of the variance of u and v .

Consider next the MAR of equations (3.8)-(3.9). Recursive substitution yields:

$$(3.12) \quad x_t = e_{1t} + (a+br)e_{1t-1} + (a^2+bc+br(a+d))e_{1t-2} + \dots \\ + b\sqrt{(1-r^2)} e_{2t-1} + b(a+d)\sqrt{(1-r^2)} e_{2t-2} + \dots$$

$$(3.13) \quad y_t = re_{1t} + (c+dr)e_{1t-1} + [c(a+d)+r(bc+d^2)]e_{1t-2} + \dots \\ + \sqrt{(1-r^2)} e_{2t} + d\sqrt{(1-r^2)} e_{2t-1} + (bc+d^2)\sqrt{(1-r^2)} e_{2t-2} + \dots$$

As above, e_1 and e_2 have zero correlation. However, an increase in e_1 increases x by 1 unit and increases y by r units, since e_1 directly enters the y equation.

The errors e_1 and e_2 can be considered the innovations in the new (x,y) system given by (3.8)-(3.9). By inspection, e_1 and e_2 are defined as

$$(3.14) \quad e_{1t} = x_t - E(x_t | x_{t-1}, y_{t-1})$$

$$(3.15) \quad e_{2t} = (y_t - E(y_t | x_t, x_{t-1}, y_{t-1})) / \sqrt{(1-r^2)}.$$

Notice that the information sets in the two equations are different. These should be compared to the innovations in the (x,y) system given by (3.1)-(3.2),

$$(3.16) \quad u_t = x_t - E(x_t | x_{t-1}, y_{t-1})$$

$$(3.17) \quad v_t = y_t - E(y_t | x_{t-1}, y_{t-1}).$$

The triangularized systems--(3.8)-(3.11) and (3.12)-(3.13)--are, unfortunately, only one possible transformation. These systems arose by ordering the variables such that x was first. Consider, instead, an ordering where y is first,

$$(3.18) \quad \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} d & c \\ b & a \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ u_t \end{bmatrix}$$

$$(3.19) \quad E \begin{bmatrix} v_t \\ u_t \end{bmatrix} \begin{bmatrix} v_t & u_t \end{bmatrix} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} = \Omega .$$

In general, Ω will be different from Σ since the variances of u and v are generally different. However, in this case the Cholesky decomposition is the same as before. Therefore, multiplying both sides of (3.18) by H yields:

$$(3.20) \quad y_t = dy_{t-1} + cx_{t-1} + e_{3t}$$

$$(3.21) \quad x_t = ry_t + (b-rd)y_{t-1} + (a-rc)x_{t-1} + \sqrt{(1-r^2)} e_{4t}$$

$$(3.22) \quad e_{3t} = v_t$$

$$(3.23) \quad e_{4t} = (-rv_t + u_t) / \sqrt{(1-r^2)}.$$

The recursive system given by (3.8)-(3.11) is different from the one given by (3.20)-(3.23). Although the H matrix in (3.20)-(3.23) happens to be the same in this example as that in (3.8)-(3.11), $e_1 \neq e_4$ and $e_2 \neq e_3$ because different linear combinations of u and v enter the two sets of equations. The x innovation in the system given by (3.8)-(3.11) was defined in equation (3.14); for the system (3.20)-(3.23) it is defined as,

$$(3.24) \quad e_{4t} = (x_t - E(x_t | y_t, x_{t-1}, y_{t-1})) / \sqrt{(1-r^2)}.$$

In particular, the information sets are different in (3.14) and (3.24). Also, in (3.8)-(3.11) an x innovation influences x and y , whereas in (3.20)-(3.23),

an x innovation influences only x . As stated earlier, the reason is that the variables are ordered differently in (3.8)-(3.11) than in (3.20)-(3.23). Of course, if the correlation between u and v is zero ($r=0$), then the ordering does not matter.

The issue of exogeneity can be investigated using the system given by (3.8)-(3.11), and its MAR given by (3.12)-(3.13). According to the exogeneity conditions discussed in section 2, x is exogenous with respect to y if and only if

$$(3.25) \quad \begin{aligned} b\sqrt{(1-r^2)} &= 0 \\ b(a+d)\sqrt{(1-r^2)} &= 0 \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned}$$

That is, e_2 does not enter the x equation. This set of equations implies that $b=0$. (The alternative, $r=1$, is not allowed since it would imply that Σ is singular.) But according to equation (3.1) or (3.8), this implies that x can be written as

$$x_t = ax_{t-1} + u_t,$$

so that y_{t-1} does not enter the x equation; that is, x is exogenous.^{15/}

Equations (3.12)-(3.13) can also be used to forecast x and y . Due to the complexity of the MAR, consider the 2-step-ahead forecast of x_t ,

$$\begin{aligned} x_{t+2} &= e_{1t+2} + (a+br)e_{1t+1} + (a^2+bc+br(a+d))e_{1t} + \dots \\ &\quad + b\sqrt{(1-r^2)} e_{2t+1} + b(a+d)\sqrt{(1-r^2)} e_{2t} + \dots \\ E_t x_{t+2} &= (a^2+bc+br(a+d))e_{1t} + \dots \\ &\quad + b(a+d)\sqrt{(1-r^2)} e_{2t} + \dots \end{aligned}$$

Therefore, the 2-step-ahead forecast error is given by

$$x_{t+2} - E_t x_{t+2} = e_{1t+2} + (a+br)e_{1t+1} + b\sqrt{(1-r^2)} e_{2t+1} .$$

Recalling that the variance of e_1 and e_2 equals 1 and the covariance equals 0, the variance of the 2-step-ahead forecast error is given by

$$(3.26) \quad E(x_{t+2} - E_t x_{t+2})^2 = 1 + (a+br)^2 + b^2(1-r^2) .$$

This corresponds to the formula given by (2.18).

Equation (3.26) can be used to examine the variance of the 2-step-ahead prediction error. The percentage of the variance explained by the x innovation, e_1 , is given by

$$\frac{100 * (1 + (a+br)^2)}{1 + (a+br)^2 + b^2(1-r^2)} .$$

The percentage of the variance explained by the y innovation, e_2 , is given by

$$\frac{100 * b^2(1-r^2)}{1 + (a+br)^2 + b^2(1-r^2)} .$$

Suppose that x is exogenous with respect to y . According to (3.25) this implies that $b=0$. Substituting this into the above implies that 100 percent of the variance of the 2-step-ahead forecast error in x is explained by x innovations and 0 percent is explained by y innovations.

4. ESTIMATION OF A SIMPLE VAR

In this section a simple three variable VAR is estimated. The three variables are the growth in real income, GY, the growth in the money supply, GM, and the growth in prices (inflation), GP. The VAR is estimated using quarterly data, from 1953:3 to 1983:4. Subsection (a) discusses the choice of lag length. Subsection (b) discusses the actual estimation of the VAR. Included in this section are the VAR results and the associated moving average representation. Subsection (c) discusses the variance decomposition of the system. Finally, subsection (d) reports the results of an unconditional and conditional forecast.

To fix notation, using equation (1.17), the model estimated can be written as:

$$\begin{aligned}
 (4.1) \quad & GM_t = a_{10,0} + \sum a_{11,s} GM_{t-s} + \sum a_{12,s} GY_{t-s} + \sum a_{13,s} GP_{t-s} + u_{1t} \\
 & GY_t = a_{20,0} + \sum a_{21,s} GM_{t-s} + \sum a_{22,s} GY_{t-s} + \sum a_{23,s} GP_{t-s} + u_{2t} \\
 & GP_t = a_{30,0} + \sum a_{31,s} GM_{t-s} + \sum a_{32,s} GY_{t-s} + \sum a_{33,s} GP_{t-s} + u_{3t},
 \end{aligned}$$

where Σ runs from $s=1, N$ and $A_s = (a_{ij,s})$ in equation (1.17).

a. Choice of lag length

The choice of lag length, N , is accomplished via an asymptotic χ^2 test. First, a maximum possible lag length is chosen, call it $NMAX$. Second, equation (4.1) is estimated with $N=1, \dots, NMAX$. Third, for each value of N , the covariance matrix of the residuals is calculated, call it Σ_N . Fourth, one tests if lag length $N1$ is a restriction on a lag length of $N2$, where $N1$ and $N2$ are less than $NMAX$. Two χ^2 -tests are available-- $CHI1$ and $CHI2$,

$$(4.2) \quad \text{CHI1} = T * (\ln(\det \Sigma_{N1}) - \ln(\det \Sigma_{N2}))$$

$$(4.3) \quad \text{CHI2} = (T - c) * (\ln(\det \Sigma_{N1}) - \ln(\det \Sigma_{N2})),$$

where T = the number of observations and c = a correction factor. The correction factor, suggested by Sims (1980), is an attempt to include a degrees of freedom correction to a χ^2 -statistic, analogous to an F -statistic. The value of c equals the number of variables in each unrestricted equation of the VAR. For this example, $c = 1 + 3 * N2$. CHI1 and CHI2 are each asymptotically distributed χ^2 with $3 * 3 * (N2 - N1)$ degrees of freedom since there are $N2 - N1$ zero restrictions for each of the three variables in each of the three equations.

Table 1 reports the results of choosing a lag length for a (GY, GM, GP) VAR, with $N_{\text{MAX}} = 10$, using both CHI1 and CHI2 . The entries in Table 1 report the marginal significance level (msl) for the test statistic. (The marginal significance level equals the probability of obtaining a value of a χ^2 random variable greater than or equal to the value of the calculated statistic. That is, it is the area under a χ^2 distribution to the right of the calculated statistic. A value less than 0.05 implies a rejection at the 95% level.) The lower half of Table 1 reports the results for CHI1 . Looking at row 9, lags 1 to 7 are restrictions on a lag of 9, while a lag of 8 is not a restriction ($\text{msl} = 0.15$). In row 10, lags 1 to 7 are restrictions while lags 8 and 9 are not restrictions ($\text{msl} = 0.23$ and 0.48). Thus, CHI1 indicates a lag of 8 should be chosen. The top half of the Table reports the results for CHI2 . Looking at column 9, lags 1-4 and 6 are restrictions on a lag of 9, while lags of 5, 7, and 8 are not restrictions. Since 6 is a restriction on 9, a lag length of 5 is inappropriate. Also, looking at column 8, a lag of 7 is a restriction on a lag of 8 ($\text{msl} = 0.04$). Therefore, CHI2 also indicates a lag of 8 should be chosen.

b. Estimation of the VAR

The results of estimating a VAR for (GM,GY,GP) are given in Table 2. To conserve space, only the sums of the lag coefficients are reported. The individual coefficients ($a_{ij,s}$) do not mean much due to the multicollinearity in the variables.

As discussed in section two, a convenient way to summarize and interpret the results in Table 2 is to calculate the moving average representation (MAR). To calculate the MAR, the variables must be ordered. The ordering chosen is (GM,GY,GP). The interpretation is as follows. Current money innovations enter the GM, GY and GP equations, current income innovations enter only the GY and GP equations, and current inflation innovations are allowed to enter only the GP equation.

One can interpret the MAR as tracing the response of GM, GY and GP to a one standard deviation shock to each of the three variables. Using equation (1.21) and defining $B_s=(b_{ij,s})$, the MAR of (4.1) can be written as

$$\begin{aligned}
 (4.4) \quad GM_t &= b_{10,0} + \sum_{s=0}^{\infty} b_{11,s}e^{1t-s} + \sum_{s=1}^{\infty} b_{12,s}e^{2t-s} + \sum_{s=1}^{\infty} b_{13,s}e^{3t-s} \\
 GY_t &= b_{20,0} + \sum_{s=0}^{\infty} b_{21,s}e^{1t-s} + \sum_{s=0}^{\infty} b_{22,s}e^{2t-s} + \sum_{s=1}^{\infty} b_{23,s}e^{3t-s} \\
 GP_t &= b_{30,0} + \sum_{s=0}^{\infty} b_{31,s}e^{1t-s} + \sum_{s=0}^{\infty} b_{32,s}e^{2t-s} + \sum_{s=0}^{\infty} b_{33,s}e^{3t-s}
 \end{aligned}$$

Consider the meaning of the sequence $b_{11,0}, b_{11,1}, b_{11,2}, \dots$. Write the first equation in (4.4) for successive t , as

$$GM_t = b_{10,0} + b_{11,0}e^{1t} + b_{11,1}e^{1t-1} + b_{11,2}e^{1t-2} + \dots$$

$$GM_{t+1} = b_{10,0} + b_{11,0}e_{1t+1} + b_{11,1}e_{1t} + b_{11,2}e_{1t-1} + \dots$$

$$GM_{t+2} = b_{10,0} + b_{11,0}e_{1t+2} + b_{11,1}e_{1t+1} + b_{11,2}e_{1t} + \dots$$

Written in this way, it is easy to see that $b_{11,k}$ equals the effect of an increase in e_{1t} on GM_{t+k} . That is,

$$b_{11,0} = dGM_t/d e_{1t}$$

$$b_{11,1} = dGM_{t+1}/d e_{1t}$$

·
·
·

$$b_{11,k} = dGM_{t+k}/d e_{1t},$$

where $d e_{1t} = 1$, which equals a one standard deviation change in u_{1t} by construction.

Figure 1 plots the response of GM, GY, and GP to a one standard deviation shock in GM. That is, Figure 1 plots $b_{11,s}$, $b_{21,s}$, $b_{31,s}$ against s . As can be seen, the change in GM_{t+k} is generally positive, but converges to zero. That is, an unexpected innovation in GM at time t means that GM will be higher in the future but will eventually go back to zero. For example, $b_{11,0} = dGM_t/d e_{1t} = .00569$ but $b_{11,40} = dGM_{t+40}/d e_{1t} = .000569$. The change in GY_{t+k} is initially positive, for periods 1-4, but is generally negative from periods 5-40. That is, a temporary, unexpected increase in money growth leads to an increase in real income growth for four quarters but then leads to a decrease in real income growth. An unexpected money innovation leads to a gradual increase in GP that peaks 10 periods after the shock, then slowly falls back towards zero.

Figure 2 plots the response of GM, GY and GP to a one standard deviation shock in GY; that is, it plots $b_{12,s}$, $b_{22,s}$, $b_{32,s}$. (Notice that $b_{12,0}=0$ reflecting the restriction that a current innovation in GY has no effect on GM_t , due to the assumed ordering.) After an innovation in GY, GM is negative in periods 2-7, positive in periods 8-10, and then fluctuates between positive and negative values but converges to zero. An innovation in income growth leads to positive income growth in the short run (periods 1 and 2), negative income growth in the medium run (periods 3-7), and then fluctuates but converges to zero. Finally, a current innovation in GY leads to a slight fall in GP today, a rise in GP in period 2 and then GP fluctuates near zero.

Figure 3 plots the response of GM, GY and GP to a one standard deviation shock in GP; that is, it plots $b_{13,s}$, $b_{23,s}$, $b_{33,s}$ against s . An innovation in GP leads to negative money growth for periods 2-8. From period 9 on, money growth is generally positive. An innovation in GP leads to negative income growth in periods 2-9, positive income growth in periods 10-18, and generally negative income growth in periods 19-40. Finally, an innovation in GP leads to positive inflation today and in periods 2-10, a negative inflation in periods 11-17, positive inflation in periods 18-40, and then converges to zero.

A one standard deviation shock in GM, GY, or GP leads to a permanent increase in M, Y, or P. Consider, for example, a one-standard deviation innovation in GM, and assume money growth is expected to be 0. Then, a temporary, one period, increase in GM (with GM expected to return to zero), leads to M being permanently higher. Therefore, the effects on M, Y, and P following one standard deviation innovations in GM, GY, and GP, is found by cumulating the responses plotted in Figures 1-3. As was shown earlier, $b_{11,k} = dGM_{t+k}/d\epsilon_t$. Therefore the increase in M_{t+k} following a money growth innovation is given by

$$dM_{t+k}/delt = \sum_{s=0}^k b_{11,s}.$$

Figures 4-6 plot the response of M, Y, and P (in logarithms) to one standard deviation shocks in GM, GY, and GP--equivalently, permanent shocks in M, Y, and P. Since the figures are simply alternative representations of Figures 1-3, only Figure 4 will be discussed. Figure 4 indicates that a permanent shock to M (at time t+1) leads to ever higher values of M and P. However, the higher value of M leads eventually to a lower level of Y.

c. Variance decomposition of the VAR

The variance decomposition of (GM, GY, GP) is contained in Table 3. As indicated in section two, the forecast variance of each series can be decomposed into components caused by each of the three series. Table 3 reports the percentage of the k-step-ahead squared prediction error in (GM, GY, GP) due to innovations in (GM, GY, GP), for selected values of k. For example, 77.01 percent of the variance in the 16 quarter (4 year) ahead forecast error of GM is due to innovations in GM; 45.57 percent of the variance in the 10 year ahead (k=40) forecast error of GP is due to innovations in GP.

Table 3 can be used to determine the exogeneity of GM, GY, and GP. A series is exogenous if 100 percent of the k-step-ahead forecast error variance is due to innovations in the series itself. It appears that GM is exogenous, while GY and GP are not exogenous. GM is exogenous since most of the forecast error variance is attributable to GM: 92.51 percent at 1 year, 77.01 percent at 4 years and 78.15 percent at 10 years. Notice, however, that 33 percent of the 4 to 10 year-ahead forecast error variance in GY is due to innovations in GM, and 57 percent of the 4 to 10 year-ahead forecast error variance is due to innovations in GY. (If GY were exogenous, then larger numbers in the GY

column and smaller numbers in the GM column in the middle section of the Table are expected.)^{16/} Finally, innovations in GM explain 50 percent of the 8 to 10 year ahead forecast error variance in GP, while GP explains only 46 percent.

d. Forecasts using the VAR

Unconditional forecasts of GM, GY, and GP are readily calculated from the estimated model, according to equation (4.1). Forecasts are generated through the end of the century, 1999:4. Figure 7 plots the historical values and the predicted values of GM, GY, and GP. The figures refer to annual growth rates, quarter 1 over quarter 1. As can be seen, all variables converge: GM converges to 6.21 percent, GY converges to 2.69 percent and GP converges to 5.57 percent. Using the identity that $MV=PY$, where V equals velocity, these numbers imply velocity rising at a rate of 2.05 percent.

In addition to unconditional forecasts, one can also generate a set of conditional forecasts. Of interest to the Federal Reserve is a policy of reducing money growth to reduce inflation. Therefore, consider a forecast of GY and GP, conditional on a reduction in GM. In particular, suppose GM is gradually reduced to 1 percent by the year 1988 and then kept at 1 percent until 1999:4. The forecast of GY and GP is obtained from equations (4.1b) and (4.1c), after imposing the new money supply policy. As noted earlier, this procedure is subject to the Lucas critique. Since during the estimation period there was not a sustained period of low money growth, it is likely that switching to such a policy would lead to changes in the parameters of equation (4.1), especially since the unconditional forecast of GM tends to 6 percent. As can be seen in Figure 8, GY converges to 3.6 percent and GP converges to 1.3 percent, when money growth is reduced to 1 percent. These results indicate that reducing GM from 6.2 percent (the unconditional forecast) to 1 percent (the conditional forecast) leads to an increase in GY from 2.7 percent to 3.6

percent, a fall in GP from 5.6 percent to 1.3 percent, and a rise in velocity growth from 2.1 percent to 3.9 percent.

5. THE RELATIONSHIP BETWEEN STANDARD STRUCTURAL ECONOMETRIC MODELS AND VAR MODELS

It has been shown by Arnold Zellner and Franz Palm (1974) and Arnold Zellner (1979) that any structural econometric model may be viewed as a restricted vector time series model. This section will focus on a more narrow aspect and show that a VAR is an unrestricted reduced form of an unknown structural model.^{17/}

A typical structural econometric model may be written as (5.1),

$$(5.1) \quad J(L)y_t + M(L)x_t = N(L)u_t$$
$$J(L) = J_0 + J_1L + J_2L^2 + \dots + J_jL^j$$
$$M(L) = M_0 + M_1L + M_2L^2 + \dots + M_mL^m$$
$$N(L) = N_0 + N_1L + N_2L^2 + \dots + N_wL^w$$
$$E(u_t) = 0$$
$$E(u_t u_t') = I$$
$$E(u_t u_s') = 0 \quad t \neq s$$
$$E(x_t u_s') = 0 \quad \text{for all } t, s,$$

where y_t is an $n \times 1$ vector of endogenous variables, x_t is a $k \times 1$ vector of exogenous variables, u_t is an $n \times 1$ white noise vector, J_i is an $n \times n$ matrix of coefficients, M_i is an $n \times k$ matrix of coefficients, and N_i is an $n \times n$ matrix of coefficients. The vector of exogenous variables is assumed to be generated by the vector autoregressive-moving average process,

$$(5.2) \quad Q(L)x_t = R(L)v_t$$
$$Q(L) = Q_0 + Q_1L + Q_2L^2 + \dots + Q_hL^h$$

$$R(L) = R_0 + R_1L + R_2L^2 + \dots + R_rL^r$$

$$E(v_t) = 0$$

$$E(v_tv_t') = I$$

$$E(v_tv_s') = 0 \quad t \neq s$$

$$E(v_tv_s') = 0 \quad \text{for all } t, s,$$

where Q_i is a $k \times k$ matrix of coefficients and v_t is a $k \times 1$ white noise vector. In general, j , m , h , and r are small and w is usually zero. Equations (5.1) and (5.2) can be combined as

$$(5.3) \quad \begin{bmatrix} J(L) & M(L) \\ 0 & Q(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} N(L) & 0 \\ 0 & R(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix}.$$

If u and v are invertible, (5.3) may be written as a pure VAR model,

$$(5.4) \quad \begin{bmatrix} J^*(L) & M^*(L) \\ 0 & Q^*(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \end{bmatrix},$$

where $J^*(L) = N(L)^{-1}J(L)$, $M^*(L) = N(L)^{-1}M(L)$, and $Q^*(L) = R(L)^{-1}Q(L)$.

The reduced form of (5.1) is

$$(5.5) \quad y_t = J_0^{-1}(J_0L^{-1} - J(L)L^{-1})y_{t-1} - J_0^{-1}M(L)x_t + J_0^{-1}N(L)u_t.$$

The disturbance term in (5.5) is a vector moving average process, not white noise. A reduced form equation with a white noise disturbance term can be derived from the first equation of (5.4) as

$$(5.6) \quad y_t = J_0^{*-1}(J_0^*L^{-1}-J^*(L)L^{-1})y_{t-1} - J_0^{*-1}M^*(L)x_t + u_t^*$$

where $u_t^* = J_0^{*-1}u_t$.

Note that although $E(u_t u_t') = I$, $E(u_t^* u_t^{*'}) = J_0^{*-1} u_t u_t' J_0^{*-1} \neq I$ unless $J_0^* = I$. The exogeneity conditions imposed on x imply that (5.1), (5.5), and (5.6) are statistically independent of (5.2) so that (5.1), (5.5) and (5.6) can be estimated without taking account of the process generating x .

A more general reduced form can be derived from the system (5.3),

$$(5.7) \quad \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} J_0 & M_0 \\ 0 & Q_0 \end{bmatrix}^{-1} \begin{bmatrix} J_0 L^{-1} - J(L)L^{-1} & M_0 L^{-1} - M(L)L^{-1} \\ 0 & Q_0 L^{-1} - Q(L)L^{-1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} \\ + \begin{bmatrix} J_0 & M_0 \\ 0 & Q_0 \end{bmatrix}^{-1} \begin{bmatrix} N(L) & 0 \\ 0 & R(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix}.$$

Note that the disturbance term in the reduced form (5.7) is a vector moving average process. A general reduced form with a white noise disturbance term can be derived from (5.4) as

$$(5.8) \quad \begin{bmatrix} y_t \\ x_t \end{bmatrix} = - \begin{bmatrix} J_0^* & M_0^* \\ 0 & Q_0^* \end{bmatrix}^{-1} \begin{bmatrix} J_0^{*L^{-1}} - J^*(L)L^{-1} & M_0^{*L^{-1}} - M^*(L)L^{-1} \\ 0 & Q_0^{*L^{-1}} - Q^*(L)L^{-1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} \\ + \begin{bmatrix} u_t^* & 0 \\ 0 & v_t^* \end{bmatrix}$$

where

$$\begin{bmatrix} u_t^* & 0 \\ 0 & v_t^* \end{bmatrix} = \begin{bmatrix} J0^* & M0^* \\ 0 & Q0^* \end{bmatrix}^{-1} \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

and

$$E \begin{bmatrix} u_t \\ v_t \end{bmatrix} [u_t' \ v_t'] = \Sigma.$$

Note that in general $\Sigma \neq I$.

A VAR model is an unrestricted reduced form of some unknown structural system of equations. A VAR(p) model can be written as

$$(5.9) \quad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$A_{ij}(L) = A_{ij1} + A_{ij2}L + A_{ij3}L^2 + \dots + A_{ijp}L^{p-1}$$

$$E(e_{1t}) = E(e_{2t}) = 0$$

$$E(e_t e_t') = \Sigma, \quad e_t = [e_{1t}' \ e_{2t}']'$$

$$E(e_{1t} e_{1s}') = E(e_{2t} e_{2s}') = E(e_{1t} e_{2s}') = 0 \quad t \neq s.$$

Comparing (5.8) and (5.9), it is clear that the reduced form model is just a restricted VAR model, where the restriction on the VAR model is that $A_{21}(L) = 0$. Thus, if exogeneity restrictions are not imposed on a set of variables, a system of equations such as (5.9) can be estimated. If exogeneity restrictions are imposed on a set of variables in y , say, y_2 , then the VAR may be run as

$$(5.10) \quad y_{1t} = A_{11}^*(L)y_{1t-1} + A_{12}^*(L)y_{2t} + e_{1t}^*$$
$$A_{11}^*(L) = A_{11,1}^* + A_{11,2}^*L + A_{11,3}^*L^2 + \dots + A_{11,p}^*L^{p-1}$$
$$A_{12}^*(L) = A_{12,0}^* + A_{12,1}^*L + A_{12,2}^*L^2 + \dots + A_{12,p}^*L^p \quad ,$$

which is equivalent to the reduced form (5.6). Note that when a set of variables are exogenous, contemporaneous values of the exogenous variables are included as right-hand side variables in the VAR.

6. AN EXAMPLE OF VARS VERSUS STRUCTURAL MODELS

In this section the robustness of VAR estimates are compared with two stage least squares (2SLS) estimates of a structural model. To do this, a data generating model (DGM), a VAR, and a structural model (SM) are needed. The DGM is a set of equations that is used to generate the observations. This is discussed in subsection (a). Subsection (b) presents the structural model that is estimated and the VAR. Finally, subsection (c) presents several experiments using the VAR and the structural model.

a. The data generating model

This subsection describes the method by which the observations on an economy are generated. By necessity, the economy is simple. The endogenous variables are income (y), interest rates (r), and money (m). The DGM contains three equations describing the behavior of (y, r, m),

$$(6.1) \quad A*ENDO(t) = B*ENDO(t-1) + C*EXO(t) + ERROR(t)$$

where $ENDO(t) = (y_t, r_t, m_t)$

$$EXO(t) = (1, z_{1t}, z_{2t}, z_{3t}, z_{4t})$$

$$(6.3c) \quad z_{3t} = .85z_{3t-1} + w_{3t}$$

$$(6.3d) \quad z_{4t} = .90z_{4t-1} + w_{4t}$$

The covariance matrix and correlation matrix of w_1, \dots, w_4 is given by

$$(6.4a) \quad \text{cov} = \begin{bmatrix} 25.00 & & & \\ 1.25 & 25.00 & & \\ 1.25 & 2.50 & 25.00 & \\ 1.25 & 1.25 & 1.25 & 25.00 \end{bmatrix}$$

$$(6.4b) \quad \text{corr} = \begin{bmatrix} 1.00 & & & \\ .05 & 1.00 & & \\ .05 & .10 & 1.00 & \\ .05 & .05 & .05 & 1.00 \end{bmatrix} .$$

Given equations (6.1) to (6.4), it is possible to generate a set of observations on (y_t, r_t, m_t) and $(z_{1t}, z_{2t}, z_{3t}, z_{4t})$.

b. The structural model and a VAR

This subsection first describes the structural model of the economy that the economist believes is accurate. Then, a VAR is specified.

The structural model contains an IS curve, a money supply equation, and a money demand equation. It is assumed that several zero restrictions have been imposed on the model. The structural model is given by,

$$(6.5) \quad (\text{IS}) \quad y_t = c_{11} + a_{12}r_t + b_{12}r_{t-1} + c_{12}z_{1t}$$

$$(6.6) \quad (\text{MS}) \quad M_t = c_{21} + a_{22}r_t + c_{23}z_{2t}$$

$$(6.7) \quad (\text{MD}) \quad M_t = c_{31} + a_{31}y_t + a_{32}r_t + b_{33}M_{t-1} + c_{34}z_{3t} .$$

The system of equations given by (6.5) - (6.7) is fairly conventional.

Given a set of observations on (y, r, m) and (z_1, z_2, z_3, z_4) it is possible to estimate the structural model. The appropriate estimation procedure is 2SLS or 3SLS; in the experiments, 2SLS is used. The set of instruments is $(1, r_{t-1}, m_{t-1}, z_{1t}, z_{2t}, z_{3t})$.

The specification of a VAR imposes several assumptions. First, it is assumed that z_1-z_3 are known to be exogenous. VARs are often estimated with all variables assumed to be endogenous so that no contemporaneous variables are on the right-hand side of the VAR. However, as was shown in section five, if a set of variables are assumed to be exogenous, contemporaneous values of these variables may enter the right hand side of the VAR. Second, it is assumed that no lags of z_1-z_3 appear in the VAR. Third, it is assumed that only one lag of the endogenous variables appear in the VAR. With these assumptions, the VAR is given by

$$(6.8) \quad y_t = a_0 + a_1 y_{t-1} + a_2 r_{t-1} + a_3 M_{t-1} + a_4 z_{1t} + a_5 z_{2t} + a_6 z_{3t}$$

$$(6.9) \quad r_t = b_0 + b_1 y_{t-1} + b_2 r_{t-1} + b_3 M_{t-1} + b_4 z_{1t} + b_5 z_{2t} + b_6 z_{3t}$$

$$(6.10) \quad M_t = c_0 + c_1 y_{t-1} + c_2 r_{t-1} + c_3 M_{t-1} + c_4 z_{1t} + c_5 z_{2t} + c_6 z_{3t}.$$

Note that equations (6.8) - (6.10) is not the true reduced form of equations (6.5) - (6.7) because (6.8) - (6.10) includes y_{t-1} while the true reduced form of (6.5) - (6.7) does not. Thus, (6.8) - (6.10) can be thought of as an unconstrained reduced form of a number of underlying, but unknown, structural systems of equations. Equations (6.8) - (6.10) would be the reduced form of

equations (6.5) - (6.7) only if the coefficients a_1 , b_1 , and c_1 in (6.8) - (6.10) were found to equal zero. Equations (6.8) - (6.10) can each be estimated by OLS.

c. Estimation

In this subsection, observations are generated according to the DGM, and the structural model (SM) and the VAR are estimated. The structural model and the VAR are then compared on two criteria. To examine the robustness of the SM versus the VAR, observations are generated according to several different DGMs. Each set of observations generated by a DGM is referred to as an experiment.

The basic DGM model (with parameters) is given by

$$(6.11) \quad (\text{IS}) \quad y_t = 1300 - .2r_t - .5r_{t-1} + z1_t + u_t$$

$$(6.12) \quad (\text{MS}) \quad M_t = 147 + r_t + z2_t + e_t$$

$$(6.13) \quad (\text{MD}) \quad M_t = -850 + .7y_t - .1r_t + z3_t + .3M_{t-1} + v_t \quad .$$

The method used to choose these parameters is given in Appendix A. Notice, comparing the basic DGM, equations (6.11) - (6.13), to the SM, equations (6.5) - (6.7), reveals that the same functional form is used in both. The stochastic structure of the structural errors, equation (6.2), and the exogenous variables, equations (6.3) and (6.4), remain the same in all experiments.

The set of experiments considered amounts to slight variations in the basic DGM. In all experiments, the same SM and VAR are estimated. In experiment 1, observations are generated according to the basic DGM. Since the form of the basic DGM and the SM are the same, estimating the SM corresponds to estimating a "true" model. By altering the basic DGM, the

robustness of the SM and the VAR to various model misspecifications can be determined. In experiment 2, $.1y_{t-1}$ is added to the IS curve in the DGM. In experiment 3, a larger value of lagged income, $.5y_{t-1}$, is added to the IS curve in the DGM. In experiment 4, $.5y_{t-1}$ is added to the MD equation in the DGM. In experiment 5, $.5z_4t$ is added to the IS curve in the DGM, while in experiment 6 it is added to the MD equation in the DGM. Experiments 2 - 4 are intended to reflect a misspecification of the SM. Notice, the VAR is not "misspecified" in these experiments. However, experiments 5 and 6 reflect a misspecification (a left-out variable) of both the SM and the VAR.

The robustness of the SM and the VAR to misspecification are compared on two grounds. First, the value of dy_t/dr_t is calculated. According to the SM, dy/dr represents the slope of the IS curve. According to the VAR, dy/dr represents the response of y , relative to the response of r , to a change in an exogenous variable. The values of dy_t/dr_t for the DGM, the SM, and the VAR are given by

$$dy_t/dr_t = -0.2 \quad (\text{DGM})$$

$$dy_t/dr_t = a_{12} \quad (\text{SM})$$

$$dy_t/dr_t = a_5/b_5 \quad (\text{VAR}).$$

While the value for the DGM and the SM are clear, the value for the VAR needs explanation. In the VAR, it is assumed that there is some variable controllable by policymakers that influences the money market, but not the goods market. In terms of a structural model, there is an exogenous variable entering the LM schedule but not the IS schedule. That is, to obtain $dy/dr = a_5/b_5$, it is assumed that z_2 is exogenous and that it is controlled by the Federal Reserve, but z_2 has no direct effect on the goods market. A change in

z_2 changes r by $dr/dz_2 = b_5$ and changes y by $dy/dz_2 = a_5$. Therefore, a change in z_2 causes a change in y relative to the change in r ,

$$\frac{dy_t}{dr_t} = \frac{dy_t}{dz_{2t}} \bigg/ \frac{dr_t}{dz_{2t}} = \frac{a_5}{b_5} .$$

In addition to calculating dy/dr , the standard error of the income equation is calculated. Call this value STD. For the DGM and the SM, this is the standard error of the reduced form income equation. For the VAR, it is the standard error of equation (6.8). According to equation (6.1), the covariance matrix of the reduced form is $A^{-1} \Sigma A^{-1}$, where Σ is the covariance matrix of ERROR.

For each DGM experiment, 100 sets of 260 observations are generated. To eliminate the effect of initial conditions, the first ten observations are dropped. For each set of observations, dy/dr and STD is calculated. Then, on the basis of the 100 sets, the average value and the root mean square error (RMSE) for dy/dr and STD are calculated. The RMSE is calculated about the DGM value and about the estimated value.

The results of the six experiments are given in Tables 4 and 5. Table 4 presents the estimates of dy/dr and its RMSE. Table 5 presents the estimates of STD and its RMSE.

In all experiments, the VAR estimate of dy_t/dr_t is closer to the DGM value than is the SM estimate. However, in experiments 1, 2, 4, 5, and 6 the RMSE of the SM estimate is less than the RMSE of the VAR estimate. This means that although the "average" value of dy/dr from the VAR is closer to the "true" value, in general the SM gives more efficient estimates of dy/dr (since the RMSE is less). However, the VAR does "better" than the SM in a RMSE sense in experiment 3, a case where the IS curve is badly misspecified (a "large"

value of y_{t-1} in the DGM). At least on the basis of these experiments, it may be concluded that point estimates from a VAR are more robust to misspecification than point estimates from a structural model. This is important since in general one does not know the "true" model.

Table 5 reports the estimates of the standard error (STD) of the reduced form income equation. The estimate of STD from the SM is closer to the true DGM value of STD in experiments 1, 2, and 4, while the estimate of STD from the VAR is closer to the true DGM value in experiments 3, 5, and 6. The point estimates of STD from the VAR are closer to the true DGM values of STD in those experiments where the IS curve is badly misspecified. Also, the RMSE is smaller for the SM estimate than for the VAR estimate in experiments 2, 4, and 6. Although not as clear as before, the VAR appears more robust to misspecification than the SM.

d. Conclusions

In this section the robustness of a SM with respect to model misspecification was compared with the robustness of a VAR. The data were generated by several different models, but the same SM and VAR were estimated in each experiment. In some experiments only the SM was misspecified, while in other experiments both the SM and VAR were misspecified.

The results indicate that point estimates from a VAR are more robust to model misspecification than point estimates from a structural model. On the other hand, point estimates from a structural model are more efficient, except when the structural model is badly misspecified.

Although the experiments performed in this section cannot be used to explain why point estimates from the VAR are more robust, it is likely that the VAR was more robust because it is a less restrictive system of equations.

The SM was formulated by imposing several restrictions on the coefficient matrices A, B, and C in equation (6.1). Given this structural model, as represented in equations (6.5) - (6.7), there is a unique reduced form model. The VAR, however, can be interpreted as an unconstrained reduced form that is consistent with many unknown structural models. This is the sense in which the VAR is atheoretical--the underlying structural model is admitted to be unknown. If a variable in the DGM is omitted from the estimated SM, say, a lagged endogenous variable as in experiments 2-4, then the estimated structural coefficients will be biased. In a VAR, however, lagged variables enter each equation so that the estimated VAR does not suffer from the error made in estimating the SM. Furthermore, in experiments 5 and 6, where a variable in the DGM was omitted from both the VAR and the SM, the VAR was more robust. Thus, it appears that the severity of the bias increases with the number of restrictions imposed on the estimated model.

The results reported in this section are very preliminary and they should be viewed with caution. Indeed, there are several directions in which the experiments performed could be improved upon. First, a more extensive DGM should be used to generate the data. In experiments 2-6, the SM was misspecified to a greater extent than was the VAR. But, in general, VARs are usually misspecified in the sense that potentially relevant variables that are included in a SM are excluded from a VAR. Whereas a SM has many variables and equations with little dynamic interaction, a VAR has fewer variables with richer dynamic interaction. To reflect these differences, a more extensive DGM should be used to generate the data. Then the SM could be estimated with most of the variables in the DGM included in the SM but with some incorrect restrictions, while the VAR would be estimated with more variables excluded from the VAR than are excluded from the SM. Second, the assumptions made

about the estimated form of the VAR are usually not known a priori and, therefore, should be relaxed before the VAR is estimated. Before the VAR was estimated, it was assumed that it was known that z_1 - z_3 were exogenous, that no lags of z_1 - z_3 were to appear in the VAR, and that only one lag of the endogenous variables was to appear in the VAR. Instead, the VAR could be estimated with all variables assumed to be endogenous and with several lags of all variables. Various statistical tests would then be used to determine the appropriate lag length and whether any of the variables are exogenous. Experiments designed along these lines would produce much stronger and more conclusive results.

7. SUMMARY AND CONCLUSIONS

This paper has provided a theoretical and empirical introduction to vector autoregression models and some preliminary results on the robustness of VAR models to model misspecification relative to standard structural econometric models. In section one a theoretical discussion of time series models in general, and of VAR models in particular, was presented. Section two presented the ways in which VARs are used. This was followed by a simple two-variable VAR example in section three. A VAR model using U.S. data on money, income, and prices was estimated in section four. Examples of the way VAR models are used were also provided. In section five the relationship between standard structural models and VAR models was discussed. There it was shown that a VAR can be viewed as an unconstrained reduced form of some unknown underlying structural model. Finally, the robustness of VAR models to model misspecification errors, relative to standard structural models, was tested in section six.

The results in section six, although preliminary, indicated that point estimates from VAR models are more robust to model misspecification than point estimates from structural models. It was suggested that stronger and more conclusive results were likely to be reached through further research that uses a more extensive data generating model and that imposes fewer a priori assumptions on the form of the estimated VAR.

TABLE 1

VAR(GM,GY,GP): 1955:3 - 1983:4

$$CHI1 = T*(\ln(\det N2) - \ln(\det N1))$$

$$CHI2 = (T-c)*(\ln(\det N1) - \ln(\det N2))$$

<u>N1</u>	<u>N2</u>									
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>
1	*	.03	.01	.03	.00	.01	.00	.00	.00	.02
2	.02	*	.07	.12	.00	.02	.01	.00	.01	.03
3	.00	.04	*	.37	.01	.03	.01	.01	.01	.04
4	.01	.05	.27	*	.00	.02	.01	.00	.01	.03
5	.00	.00	.00	.00	*	.56	.16	.05	.07	.18
6	.00	.00	.00	.00	.42	*	.06	.02	.03	.11
7	.00	.00	.01	.00	.04	.02	*	.04	.08	.23
8	.00	.00	.00	.00	.00	.00	.01	*	.35	.60
9	.00	.00	.00	.00	.00	.01	.01	.15	*	.72
10	.00	.00	.00	.00	.00	.00	.02	.23	.48	*

Notes: The lower half of the Table reports the marginal significance for CHI1, for whether N2 is a restriction on N1. The upper half of the Table reports the marginal significance level for CHI2, for whether N1 is a restriction on N2.

TABLE 2

1955:3 to 1983:4

<u>Equation</u>	<u>Constant</u>	<u>Sum of Coefficients on:</u>		
		<u>GM</u>	<u>GY</u>	<u>GP</u>
GM _t	.003 (.002)	.844 (.240)	-.122 (.225)	-.011 (.247)
GY _t	.011 (.003)	.315 (.333)	-.065 (.313)	-.608 (.343)
GP _t	-.002 (.001)	.159 (.134)	.234 (.126)	.857 (.138)

Note: Standard errors for the sum of the coefficients are reported in parentheses.

TABLE 3

Variance Decomposition of (GM, GY, GP)

Percentage of the expected k-step-ahead Squared Prediction Error in (GM, GY, GP) Produced by Innovations in (GM, GY, GP)

Percentage of the expected k-step-ahead Squared Prediction Error in:	k	Innovation in:		
		GM	GY	GP
GM	1	100.00	0.0	0.0
	4	92.51	2.67	4.82
	8	81.02	10.57	8.42
	16	77.01	13.24	9.75
	24	77.44	12.79	9.77
	32	77.90	12.46	9.64
	40	78.15	12.22	9.62
GY	1	12.46	87.53	0.0
	4	31.48	65.84	2.69
	8	31.50	61.28	7.22
	16	33.52	57.33	9.14
	24	33.43	57.31	9.25
	32	33.61	57.04	9.36
	40	33.71	56.95	9.34
GP	1	0.50	0.00	99.49
	4	7.66	0.94	91.40
	8	26.26	2.47	71.27
	16	45.95	3.40	50.65
	24	47.83	3.31	48.87
	32	50.09	3.17	46.74
	40	51.34	3.09	45.57

TABLE 4

Estimates of dy_t/dr_t^*

Experiment	dy_t/dr_t		Precision of dy_t/dr_t			
	VAR	SM	RMSE about DGM		RMSE about Estimate	
			VAR	SM	VAR	SM
1	-.199	-.189	.189	.090	.190	.090
2	-.199	-.288	.190	.136	.191	.104
3	-.199	-.752	.190	.597	.191	.226
4	-.200	-.185	.190	.119	.191	.119
5	-.238	-.276	.243	.170	.242	.152
6	-.205	-.192	.208	.100	.209	.100

Notes: Experiment 1: DGM = Structural Model (SM)
 Experiment 2: DGM = SM plus $.1y_{t-1}$ in IS curve
 Experiment 3: DGM = SM plus $.5y_{t-1}$ in IS curve
 Experiment 4: DGM = SM plus $.5y_{t-1}$ in MD equation
 Experiment 5: DGM = SM plus $.5z^4_t$ in IS curve
 Experiment 6: DGM = SM plus $.5z^4_t$ in MD equation

*In the DGM, $dy_t/dr_t = -0.2$.

TABLE 5

Estimates of the Standard Error of the Income Equation*

<u>Experiment</u>	<u>Standard Error</u>		<u>Precision of Standard Error</u>			
			<u>RMSE about</u>		<u>RMSE about</u>	
			<u>DGM</u>		<u>Estimate</u>	
	<u>VAR</u>	<u>SM</u>	<u>VAR</u>	<u>SM</u>	<u>VAR</u>	<u>SM</u>
1	9.816	9.821	.423	.446	.455	.448
2	9.816	9.846	.453	.445	.455	.447
3	9.815	10.428	.452	.755	.454	.476
4	9.815	9.853	.454	.453	.455	.455
5	10.816	10.825	1.095	1.101	.497	.494
6	9.840	9.848	.448	.446	.450	.448

Notes: See Table 4.

* The standard error in the DGM is 9.839374.

FIGURE 1

Response of GM(+), GY(□) and GP(#)
to a One Standard Deviation Innovation in GM

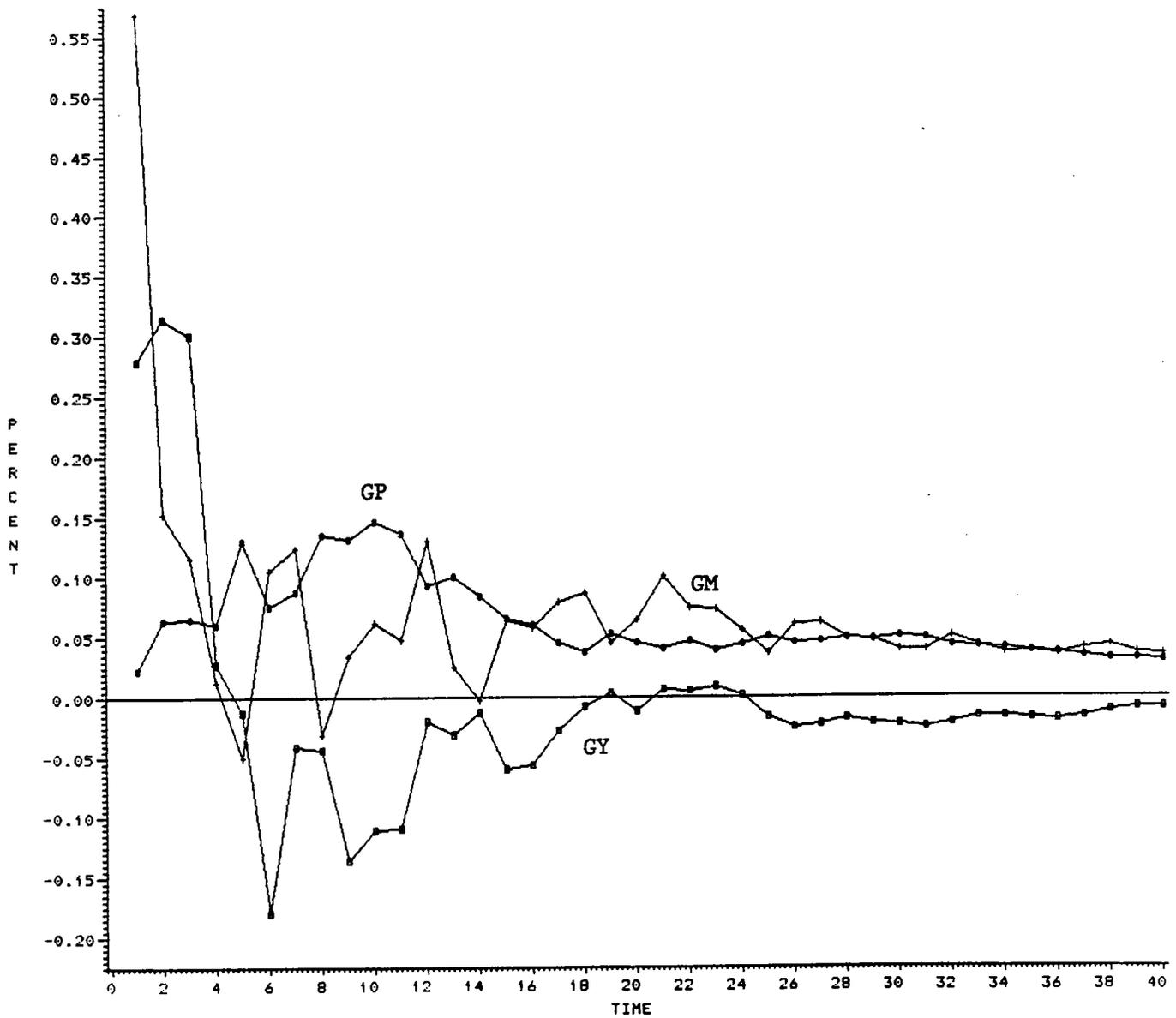


FIGURE 2

Response of GM(+), GY(□) and GP(#)

to a One Standard Deviation Innovation in GY

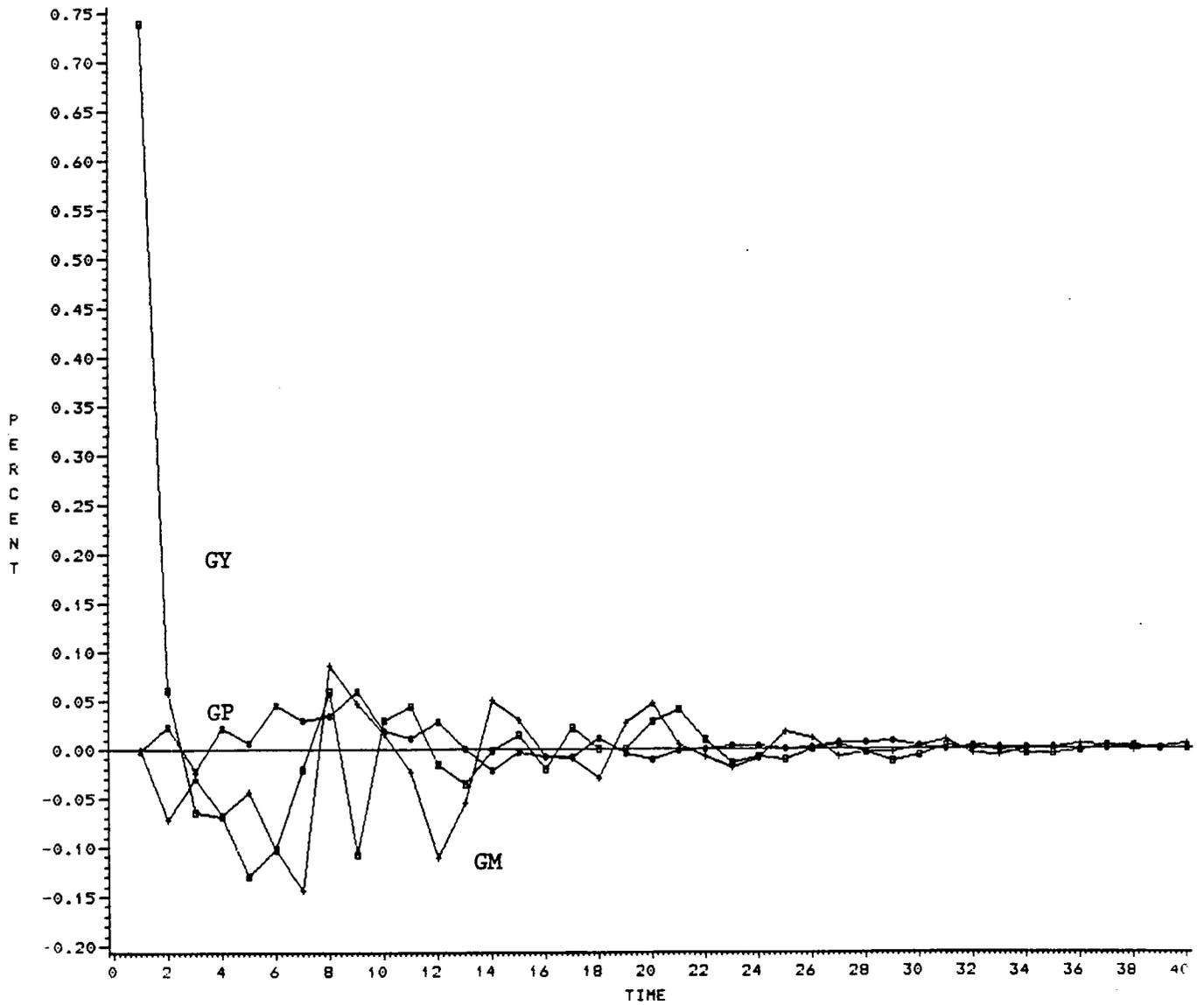


FIGURE 3

Response of GM(+), GY(□) and GP(#)

to a One Standard Deviation Innovation in GP

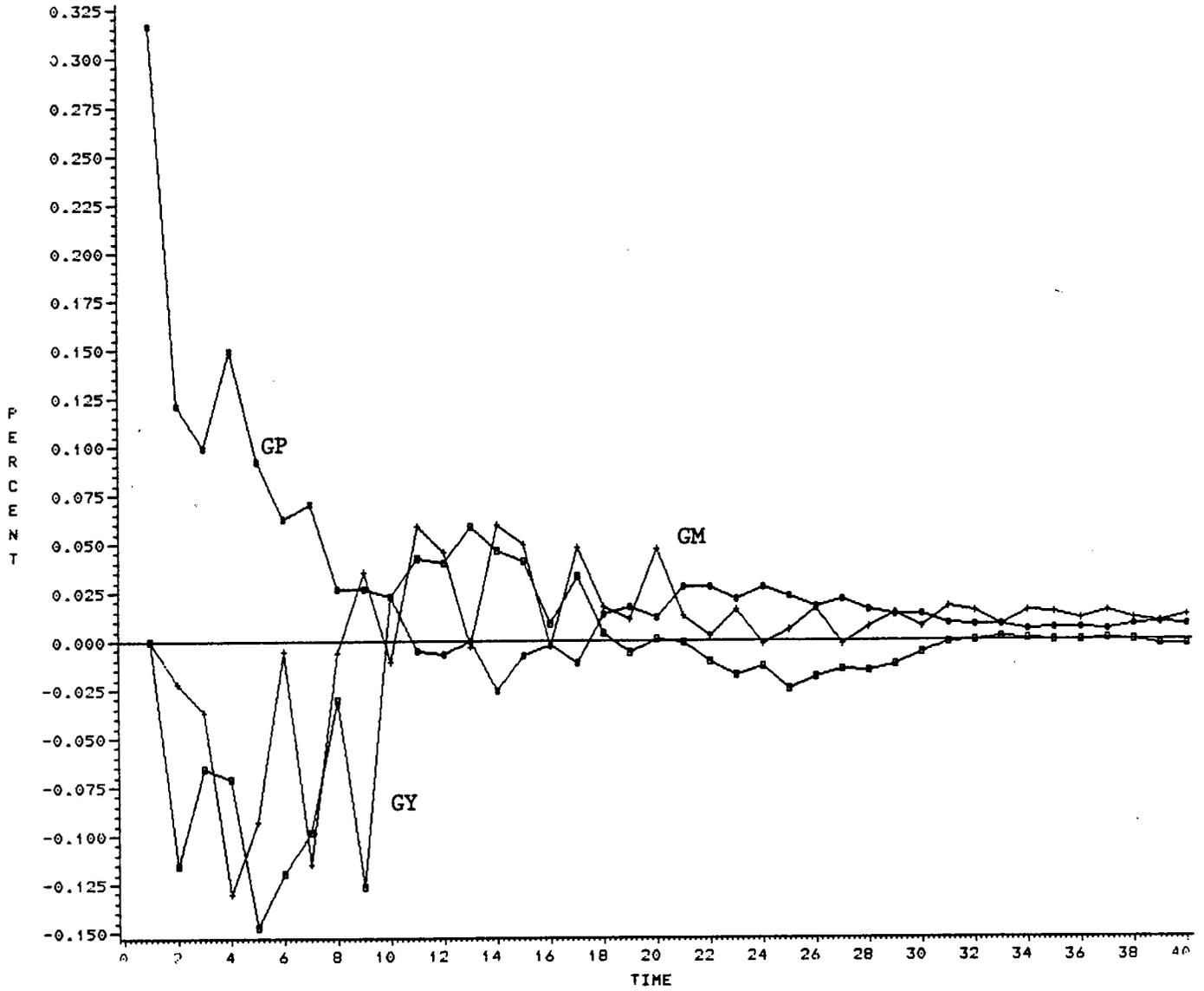


FIGURE 4

Cumulative Response of GM(+), GY(□) and GP(#) to a One Standard Deviation Innovation in GM

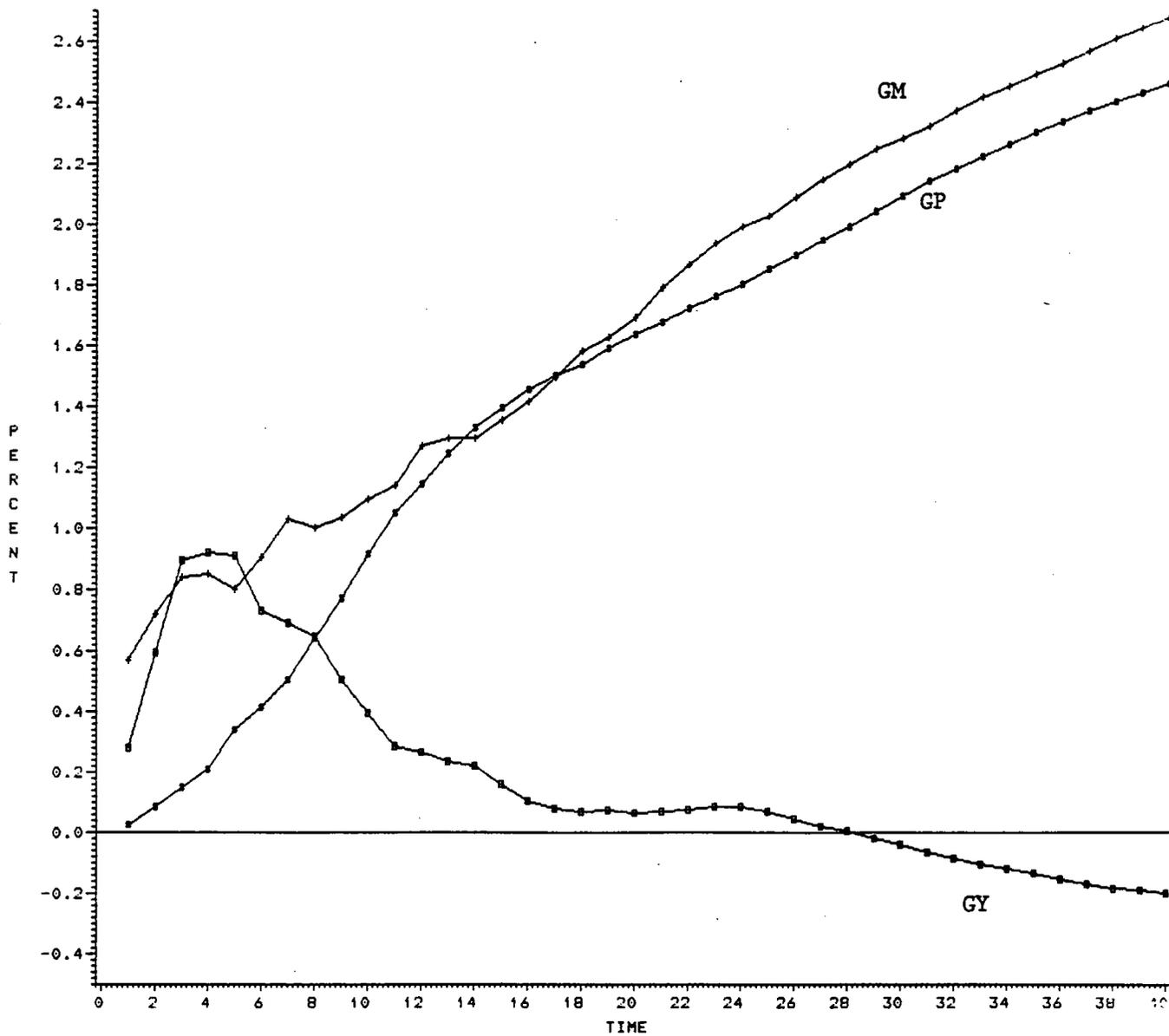


FIGURE 5

Cumulative Response of GM(+), GY(\square) and GP($\#$) to a One Standard Deviation Innovation in GY

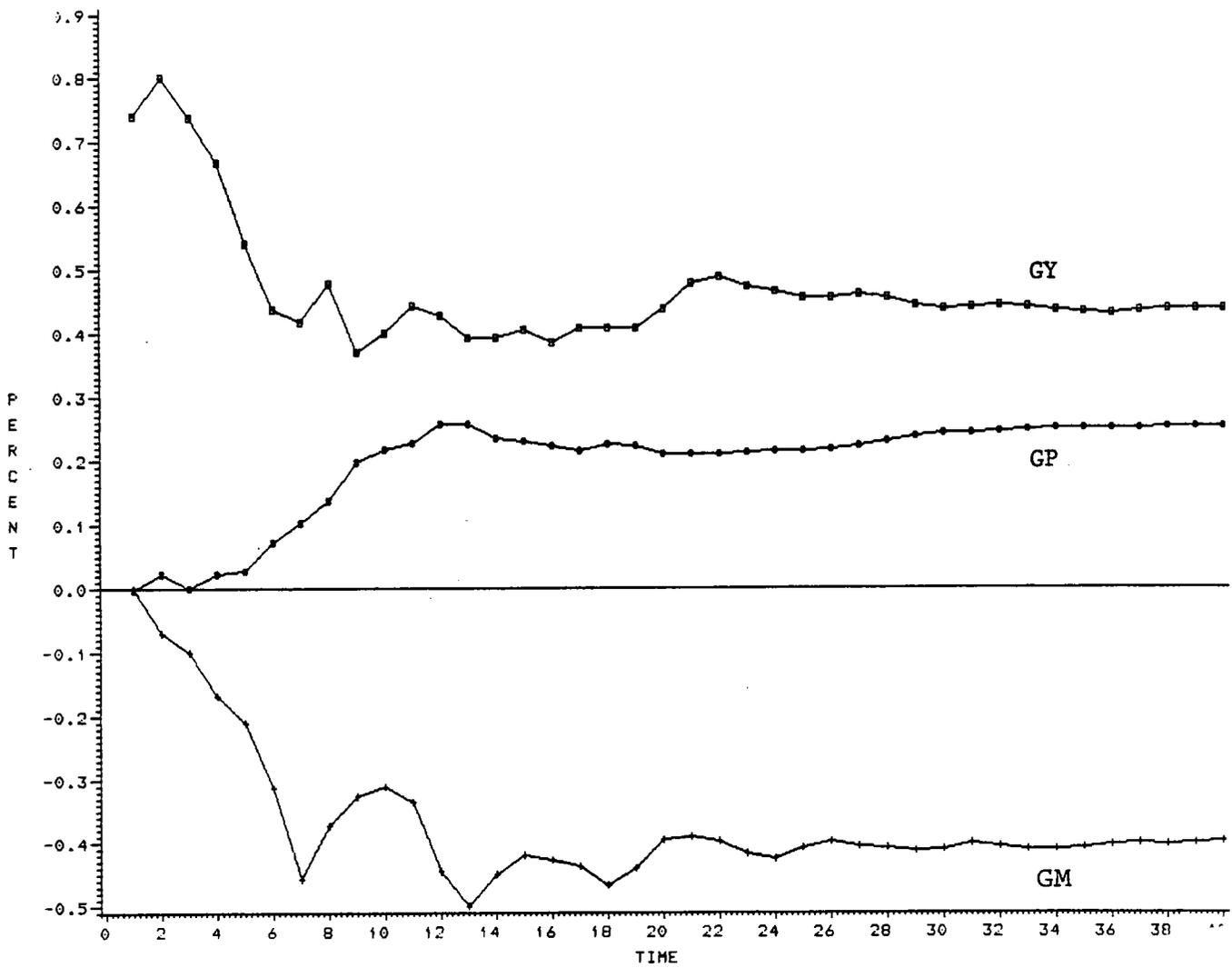


FIGURE 6

Cumulative Response of GM(+), GY(\square) and GP($\#$) to a One Standard Deviation Innovation in GP

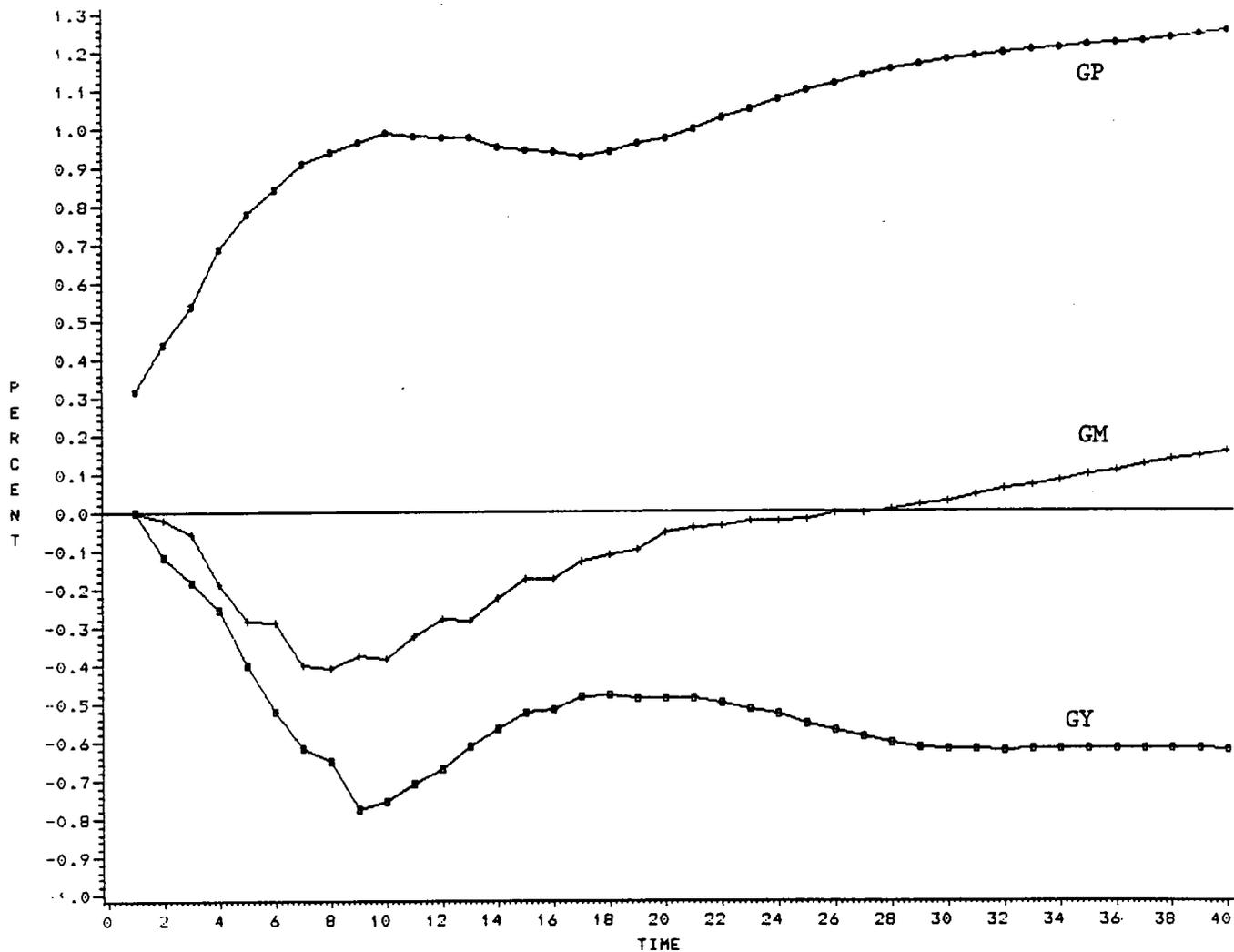


FIGURE 7

Unconditional Forecast of GM(+), GY(□) and GP(#)

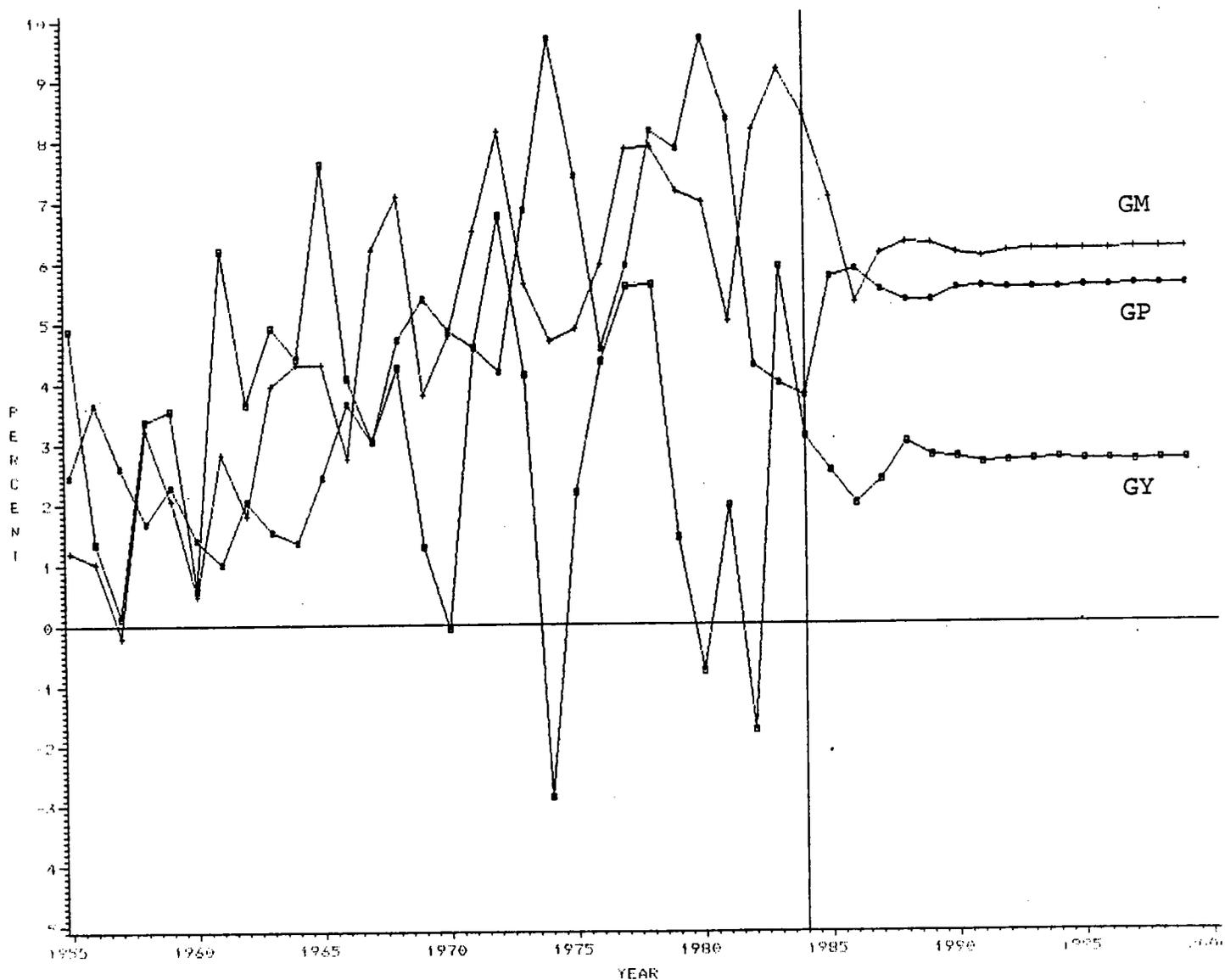
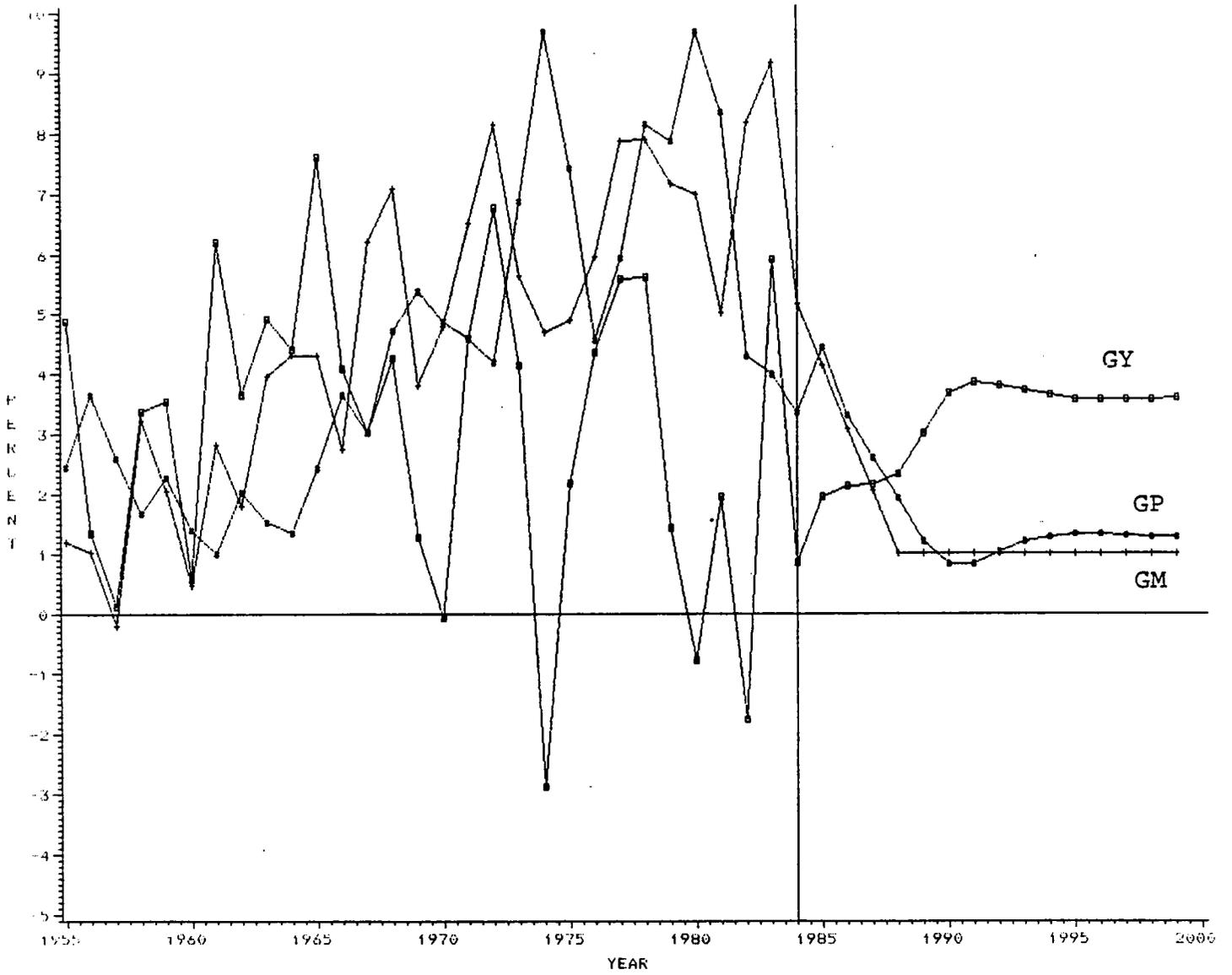


FIGURE 8

Conditional Forecast of GM(+), GY(□) and GP(#), Given GM



Footnotes

We would like to thank William Keeton and other members of the staff of the Economic Research Division at the Federal Reserve Bank of Kansas City for their comments and suggestions on an earlier draft. Lyle Matsunaga provided able research assistance. Any remaining errors are our own. The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Kansas City or of the Federal Reserve System.

1. Council of Economic Advisers, Economic Report of the President, U.S. Government Printing Office, Washington, D.C., 1983, pp. 143.
2. Actually, y only has to be a wide-sense stationary process. The time series process y is wide-sense stationary if $E(y_t)$ is constant and $E(y_t y_{t-s})$ depends only on the difference $t-s$. Wide-sense stationarity is also referred to as covariance or second-order stationarity.
3. In more general models the characteristic equation will have several roots, some of which may be complex roots. Thus, the more general condition for stationarity is that the roots of the characteristic equation lie outside the unit circle.
4. If (1.1) included a constant term, the constant term would be the unconditional expected value.
5. In (1.7) y is known to be stationary because (1.7) was derived from (1.6). If it was not known that (1.7) came from (1.6), it would be necessary to show that the roots of the characteristic equation $(1-a(L))L = a^*(L) = 0$ lie outside the unit circle.
6. Equation (1.8) is often called an ARIMA $(p,0,q)$ process. If y itself was not stationary, but could be transformed into a stationary series by taking d differences of y , the model is called an ARIMA (p,d,q) time series model.

7. For a first-order vector autoregressive process, an equivalent stationarity condition is that the eigenvalues of the coefficient matrix A in (1.11) lie inside the unit circle. Multiplying the equation $\det(I-AL) = 0$ by $1/L$ results in $\det(I\lambda-A) = 0$ where $\lambda = 1/L$. The roots of this equation are the eigenvalues of the matrix A . Thus, if the roots of $\det(I-AL) = 0$ lie outside the unit circle, the inverses of the roots--which are the eigenvalues--must lie inside the unit circle.

8. Stationary implies that A^i approaches zero as i goes to infinity. The matrix A can be written as $F\Lambda F'$ where F is an orthogonal matrix composed of the eigenvectors of A and Λ is a diagonal matrix with the eigenvalues of A along the diagonal. Using $A=F\Lambda F'$, $A^2 = F\Lambda F'F\Lambda F' = F\Lambda^2 F'$ because the orthogonality property of F implies that $F'F = I$. Thus, $A^i = F\Lambda^i F'$. If y is stationary then the elements of Λ all lie inside the unit circle which implies that Λ^i approaches the zero matrix as i goes to infinity. Therefore, $F\Lambda^i F' = A^i$ goes to zero as i goes to infinity.

9. In (1.17) y is known to be stationary because (1.17) was derived from (1.15). If it was not known that (1.17) came from (1.15), it would be necessary to show that the roots of the characteristic equation $\det(I-A(L)L) = \det(A^*(L)) = 0$ lie outside the unit circle.

10. For more details on the identification and estimation of VARMA models, see G. C. Tiao and G.E.P. Box (1981) and George C. Tiao and Ruey S. Tsay (1983).

11. The triangularity restriction on H and the restriction on the covariance matrix of the noise vector e are sufficient conditions for the uniqueness of (1.20) for a given ordering of the variables in y . See C.W.J. Granger and Paul Newbold (1977), pp. 222-224.

12. It should be noted that the large scale macroeconomic models suffer from this problem as well.

13. For a more comprehensive discussion of the types of forecasts that VARs can be used for see Thomas Sargent (1979, 1984) and Christopher Sims (1982).

14. That is, a necessary and sufficient condition for the exogeneity of the first variable of y , y_1 , is the condition that (2.20) is zero when $j \neq 1$ and 100 when $j = 1$ for all k greater than zero.

15. According to the exogeneity conditions discussed in section 2, y is exogenous with respect to x if and only if $c = 0$. Unless $r = 0$ in addition to $c = 0$, however, lagged values of x will appear in the y_t equation (3.9). Thus, because y is ordered second, y may be exogenous ($c=0$) even if lagged values of x appear in (3.9) due to nonzero contemporaneous correlation ($r \neq 0$) between the x and y errors. In terms of the variance decomposition of x , x is exogenous with respect to y if and only if the percentage of the k -step-ahead forecast error variance of x due to innovations in y is zero for all k greater than zero. But if y is exogenous, the percentage of the k -step-ahead forecast error variance of y due to innovations in x will be nonzero as long as there is a nonzero contemporaneous correlation between the x and y errors. For example, for $k = 2$ and $c = 0$ the percentage of the 2-step-ahead forecast error variance of y due to an innovation in x equals $100r^2$.

16. As noted in footnote 15, nonzero entries in the GM column in the middle section of the Table may be due to contemporaneous correlation between the innovations in GM and GY. Therefore, one can not say conclusively that GY is not exogenous.

17. For a more general and complete discussion of the relationship between time series and structural models see Zellner and Palm (1974) and Zellner (1979).

REFERENCES

Instead of simply listing references, we chose to list several references for each subsection. Not all references listed were referred to in the text.

1. VARS: A TECHNICAL DISCUSSION

1a. Univariate time series models

1. Box, G.E.P. and G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, Inc., San Francisco, 1976, chapters 1-9.
2. Fuller, Wayne A., Introduction to Statistical Time Series, John Wiley & Sons, New York, 1976, chapters 1,2,5,6,8,9.
3. Granger, C.W.J. and P. Newbold, Forecasting Economic Time Series, Academic Press, New York, 1977, chapters 1,3-6.
4. Nelson, C.R., Applied Time Series Analysis, Holden-Day, Inc., San Francisco, 1973, chapters 1-8.
5. Sargent, T.J., Macroeconomic Theory, Academic Press, New York, 1979a, chapters 9,11.
6. Whiteman, C.H., Linear Rational Expectations Models: A User's Guide, University of Minnesota Press, Minneapolis, 1983, chapter 1.

1b. Multivariate Time Series Models

1. Box and Jenkins (1976), chapters 10,11.
2. Box, G.E.P. and G.C. Tiao, "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, December 1981, pp. 802-816.
3. Gordon, R.J. and S.R. King, "The Output Cost of Disinflation in Traditional and Vector Autoregressive Models," Brookings Papers on Economic Activity, 1982:1, pp. 207-215.
4. Granger and Newbold (1977), chapter 7.
5. Sargent (1979a), chapter 16.16.

6. Sargent, T.J., "Estimating Vector Autoregressions Using Methods Not Based on Explicit Economic Theories," Quarterly Review, Federal Reserve Bank of Minneapolis, Summer 1979b, pp. 8-15.
7. Tiao, G.C. and Tsay, R.S., "Multiple Time Series Modeling and Extended Sample Cross-Correlations," Journal of Business and Economic Statistics, January 1983, pp. 43-56.
8. Whiteman (1983), chapter 4.

2. USES OF VARS

2a. Forecasting with a VAR: unconditional and conditional

1. Granger and Newbold (1977), chapters 4-8.
2. Nelson (1973), chapters 1,6.
3. Sargent, T.J., "Autoregressions, Expectations, and Advice," American Economic Review, Papers and Proceedings, May 1984, pp. 408-415.
4. Sargent (1979b).
5. Sims, C., "Policy Analysis with Econometric Models," Brookings Papers on Economic Activity, 1982:1, pp. 107-152.

2b. The moving average representation of a VAR

1. Gordon and King (1982).
2. Sims, C., "Macroeconomics and Reality," Econometrica, 48, January 1980a, pp. 1-48.

2c. Exogeneity

1. Granger and Newbold (1977), chapters 6-7.
2. Sargent (1979a), chapter 11.
3. Sims, C., "Money, Income and Causality," American Economic Review, September 1972, pp. 540-552.
4. Zellner, A. and F. Palm, "Time Series Analysis and Simultaneous Equation Econometric Models," Journal of Econometrics, May 1974, pp. 17-54.

2d. Variance decomposition and exogeneity

1. Gordon and King (1982).
2. Sims (1980).
3. A SIMPLE TWO-VARIABLE VAR EXAMPLE
4. ESTIMATION OF A SIMPLE VAR EXAMPLE
 1. King, S., Macroeconomic Activity and the Real Rate of Interest, Ph.D. dissertation, Northwestern University, 1984.
 2. Litterman, R.B. and L. Weiss, "Money, Real Interest Rates, and Output," Federal Reserve Bank of Minneapolis, working paper no. 179, September 1981.
 3. Sims, C., "A Comparison of Interwar and Postwar Cycles: Monetarism Reconsidered," American Economic Review, May 1980b, pp. 250-257.
 4. Sims (1972).

4a. Choice of lag length

1. Hsiao, C., "Autoregressive Modeling of Canadian Money and Income Data," Journal of the American Statistical Society, September 1979, pp. 553-560.
2. Judge, G.G., W.E. Griffiths, R.C. Hill and T. Lee, The Theory and Practice of Econometrics, John Wiley & Sons, New York, 1980, pp. 248-251.
3. Sims (1980).
5. THE RELATIONSHIP BETWEEN STANDARD STRUCTURAL ECONOMETRIC MODELS AND VAR MODELS
 1. Wallis, K.F., "Multiple Time Series Analysis and the Final Form of Econometric Models," Econometrica, September 1977, pp. 1481-1497.
 2. Zellner, A. and F. Palm (1974).
 3. Zellner, A., "Statistical Analysis of Econometric Models," Journal of the American Statistical Association, September 1974, pp. 628-643.

TOPICS NOT COVERED IN TEXT

1. Bayesian methods and VAR modeling

1. Litterman, R., "Specifying Vector Autoregressions for Macroeconomic Forecasting," Federal Reserve Bank of Minneapolis, working paper no. 208, July 1982.
2. Litterman, R., "Techniques of Forecasting Using Vector Autoregressions," Federal Reserve Bank of Minneapolis, working paper no. 115, November 1979.

2. Criticism of standard macroeconomic models

1. Lucas, R.E., Jr., "Econometric Policy Evaluation, a Critique," in The Phillips Curve and Labor Markets, Carnegie-Rochester Series on Public Policy, vol. 5, edited by K. Brunner and A. Meltzer, North-Holland, New York, pp. 19-46.
2. Lucas, R.E., Jr. and T.J. Sargent, "After Keynesian Macroeconomics," Quarterly Review, Federal Reserve Bank of Minneapolis, Spring 1979.

3. Causality

1. Granger, C.W.J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, July 1969, pp. 424-438.
2. Sargent (1979a).
3. Sims (1972).
4. Zellner, A., "Causality and Econometrics," in Three Aspects of Policy and Policymaking: Knowledge, Data and Institutions, Carnegie-Rochester Series on Public Policy, vol. 10, edited by K. Brunner and A. Meltzer, North-Holland, New York, pp. 9-54.

4. Computer programming

1. Doan, T.A. and R.B. Litterman, User's Manual RATS, Version 4.11,
September 1983.

APPENDIX A: METHOD USED TO CHOOSE THE DGM PARAMETERS

In this Appendix the choice of parameter values for the DGM is discussed. The variance-covariance matrix of the structural errors, equation (6.2a), is first discussed. Then, the exogenous process for z_1, \dots, z_4 , equations (6.3) and (6.4), is discussed. Finally, the parameter values in the structural model, equations (6.11) - (6.13), are discussed.

The steady-state values of y , r , and M were first chosen. These values were

$$\begin{aligned} y &= 1000 \\ M &= 150 \\ r &= 3 \end{aligned} \tag{A1}$$

The choice of the steady-state values influence later decisions.

The variance-covariance matrix of ERROR, equation (6.2), was chosen as follows. The unconditional standard error (SE^u) of each variable was chosen to equal 5 percent of the steady-state value. This implies

$$\begin{aligned} SE^u(y) &= 50 \\ SE^u(M) &= 7 \\ SE^u(r) &= .15 \end{aligned} \tag{A2}$$

The conditional variance of each variable, the diagonal elements in equation (6.2a), was chosen such that the R^2 of the reduced form equation equals .95. Since $R^2 = 1 - (\text{conditional variance}/\text{unconditional variance})$, the conditional variance is calculated as

$$\text{conditional variance} = (SE^u)^2 * (1 - R^2) \tag{A3}$$

For example, the variance of u_t is given by $(50)^2 * (1 - .95) = 125$. The correlation between the structural errors was arbitrarily chosen to equal 0.33. Given the correlation matrix and variances, the covariances are readily calculated.

The exogenous variables, z_1, \dots, z_4 , were chosen to be AR(1). The values of ρ were chosen such that the z 's were strongly autocorrelated. The particular values chosen, .75 to .90, were arbitrary.

The variance-covariance matrix of the errors in the exogenous processes, w_1, \dots, w_4 , were chosen as follows. The variances were arbitrarily chosen to be 25. The correlations were chosen to be positive and small. The correlation between the MS and MD exogenous process errors (w_2 and w_3) was chosen to be larger than the correlation with the exogenous variable in the IS curve. The absolute magnitudes were arbitrary.

The slope coefficients in the DGM, equations (6.11) - (6.13), were chosen first. For the IS curve, the coefficients on the current and lagged interest rate were chosen such that

$$\left| \frac{dy_t}{dr_t} \right| < \left| \frac{dy_t}{dr_{t-1}} \right| \quad (A4)$$

The absolute magnitudes were arbitrary. For the money supply function, a value of 1.0 was chosen, arbitrarily, for dM_t/dr_t . For the money demand function, the derivative of money with respect to the interest rate was chosen to be -.1, a small number. The steady-state slope of money demand with respect to income was chosen to be 1.0, while the short-run slope was chosen to be 0.7

The constant terms in equations (6.11) - (6.13) were chosen such that the steady state values of y , r , and M are given by (A1). The steady-state of (6.11) - (6.13) is obtained by setting $EXOG(t) = (1 \ 0 \ 0 \ 0 \ 0)$, $ERROR(t) = (0 \ 0 \ 0)$, A and B to their respective values, and $ENDOG(t) = ENDOG(t-1) = (1000, 3, 150)$.

APPENDIX B: COMPUTER PROGRAMS

```
CALENDAR 52 1 4
ALLOCATE 31 100,1 3 25
EQV 1 TO 6
  YEAR QTR P Y M INT
DATA(UNIT=INPUT,ORG=OBS) 52,1 83,4 YEAR QTR P Y M INT
52 1 57.58 593.783 121.4834 1.5666
```

```
83 4 218.53 1570.5 519.08 8.7985
```

```
NOTE 1
PRINT(DATES) 55,3 83,4 M Y P
```

```
NOTE 1
*****
* GENERATE GROWTH RATES *
```

```
NOTE 1
EQV 7 8 9
GP GY GM
SET GP 52,2 83,4 = LOG(P(T))-LOG(P(T-1))
SET GY 52,2 83,4 = LOG(Y(T))-LOG(Y(T-1))
SET GM 52,2 83,4 = LOG(M(T))-LOG(M(T-1))
```

```
*****
* PROGRAM TO CALCULATE LAG LENGTH. *
```

```
DECLARE VECTOR DETLOG(10)
IEVAL IRES1=11
IEVAL IRES2=12
IEVAL IRES3=13
IEVAL OBS1=(55,3)
IEVAL OBS2=(83,4)
  DO NLAG=1,10
    CLEAR IRES1 IRES2 IRES3
    OLS GP OBS1 OBS2 IRES1
    # CONSTANT -GP 1 NLAG -GM 1 NLAG -GY 1 NLAG
    OLS(SAME) GM OBS1 OBS2 IRES2
    OLS(SAME) GY OBS1 OBS2 IRES3
    VCV OBS1 OBS2
    # IRES1 IRES2 IRES3
    EVAL DETLOG(NLAG)=DETLN
  END DO NLAG
```

```
NOTE 1
  DO N2=2,10
    DO N1=1,(N2-1)
      IEVAL C=3*N2+1
      IEVAL DF=(N2-N1)*3*3
      EVAL CHI1=NOBS*(DETLOG(N1)-DETLOG(N2))
      EVAL CHI2=(NOBS-C)*(DETLOG(N1)-DETLOG(N2))
      WRITE N2 N1 C DF
      CDF CHISQ CHI1 DF
      CDF CHISQ CHI2 DF
    END DO N1
```

END DO N2

NOTE 1

```
*****
* ESTIMATION OF VAR, AS A SYSTEM OF EQUATIONS.      *
* LAG LENGTH = 8.                                  *
* EQUATION 1=GM, EQUATION 2=GY, EQUATION 3=GP.      *
* VARIANCE COVARIANCE MATRIX, FOR 1955:3 - 1983:4  *
* IS STORED IN VARCOV, RESIDUALS ARE STORED         *
* IN VARIABLES 11-13.                               *
*****
```

```
EQUATION(NOCONST,MORE) 1 GM
# CONSTANT -GM 1 8 -GY 1 8 -GP 1 8
EQUATION(NOCONST,MORE) 2 GY
# CONSTANT -GM 1 8 -GY 1 8 -GP 1 8
EQUATION(NOCONST,MORE) 3 GP
# CONSTANT -GM 1 8 -GY 1 8 -GP 1 8
SYSTEM 1 2 3
END(SYSTEM)
ESTIMATE 55,3 83,4 11
DECLARE SYMMETRIC VARCOV(3,3)
VCV(MATRIX=VARCOV) 55,3 83,4
# 11 12 13
```

NOTE 1

```
*****
* UNCONDITIONAL FORECAST OF VAR.                  *
* FORECAST IS FOR 64 STEPS, TO 1999:4.          *
*****
```

```
FORECAST(PRINT) 3 64 84,1
# 1 GM 84,1
# 2 GY 84,1
# 3 GP 84,1
```

NOTE 1

```
* THESE STATEMENTS CONVERT TO ANNUALIZED GROWTH RATES.
SET GM 55,3 99,4 = 400.0*GM(T)
SET GY 55,3 99,4 = 400.0*GY(T)
SET GP 55,3 99,4 = 400.0*GP(T)
```

NOTE 1

```
PRINT(DATES) 55,3 99,4 GM GY GP
```

NOTE 1

```
PLOT(DATES,CHARS) 3
# GM 55,3 99,4
# GY 55,3 99,4
# GP 55,3 99,4
# '*' 'X' '.'
```

NOTE 1

```
*****
* CONDITIONAL FORECAST OF A VAR.                  *
* FORECAST IS OF GY AND GP.                      *
* GM IS ASSUMED TO BE DECREASED TO 1%/YEAR BY   *
* 1988:1.                                         *
* FORECAST 64 STEPS AHEAD, TO 1999:4.          *
*****
```

```
* THESE STATEMENTS ARE TO RECONVERT ALL THE DATA
* BACK TO QUARTERLY GROWTH RATES, TO MATCH THE
```

```
* FORMAT USED IN ESTIMATION.
SET GM 55,3 99,4 = GM(T)/400.0
SET GY 55,3 99,4 = GY(T)/400.0
SET GP 55,3 99,4 = GP(T)/400.0
* THESE STATEMENTS INPUT THE ASSUMED PATH OF GM.
DATA(UNIT=INPUT,ORG=VAR) 84,1 88,4 GM
0.0129 0.0129 0.0129 0.0129
0.0103 0.0103 0.0103 0.0103
0.0077 0.0077 0.0077 0.0077
0.0051 0.0051 0.0051 0.0051
0.0025 0.0025 0.0025 0.0025
SET GM 89,1 99,4 = 0.0025
FORECAST(PRINT) 2 64 84,1
# 2 GY 84,1
# 3 GP 84,1
NOTE 1
* CONVERT TO ANNUALIZED GROWTH RATES.
SET GM 55,3 99,4 = 400.0*GM(T)
SET GY 55,3 99,4 = 400.0*GY(T)
SET GP 55,3 99,4 = 400.0*GP(T)
NOTE 1
PRINT(DATES) 55,3 99,4 GM GY GP
NOTE 1
PLOT(DATES, CHARS) 3
# GM 55,3 99,4
# GY 55,3 99,4
# GP 55,3 99,4
# '*' 'X' '.'
NOTE 1
```

```
*****
* CALCULATE THE MOVING AVERAGE REPRESENTATION, *
* ALSO KNOWN AS THE IMPULSE RESPONSE FUNCTION. *
* CALCULATION IS FOR 40-STEPS AHEAD. *
* THE ORDERING IS GM(1), GY(2), GP(3) *
* THE MA COEFFICIENTS ARE STORED IN VARIABLES *
* 14-16, 17-19, 20-22. *
* THE RESPONSE COEFFICIENTS ARE STORED IN *
* XXRESZ: "XX RESPONSE TO A Z INNOVATION. *
*****
* RESPONSE OF GM, GY, GP TO A SHOCK IN GM(EQUATION 1)
IMPULSE 3 40 1 VARCOV
# 1 14 1 1
# 2 15 1 2
# 3 16 1 3
* RESPONSE OF GM, GY, GP TO A SHOCK IN GY(EQUATION 2)
IMPULSE 3 40 2 VARCOV
# 1 17 1 1
# 2 18 1 2
# 3 19 1 3
* RESPONSE OF GM, GY, GP TO A SHOCK IN GP(EQUATION 3)
IMPULSE 3 40 3 VARCOV
# 1 20 1 1
# 2 21 1 2
# 3 22 1 3
NOTE 1
EQV 14 TO 22
GMRESM GYRESM GPRESM $
GMRESY GYRESY GPRESY $
GMRESP GYRESP GPRESP
NOTE 1
*****
* CALCULATE THE VARIANCE DECOMPOSITION OF *
* OF THE VAR, IN THE ORDER GM(1), GY(2), *
* GP(3). *
*****
ERRORS 3 40 VARCOV
# 1 0 0 1
# 2 0 0 2
# 3 0 0 3
NOTE 1
PLOT 3
# GMRESM 1 40
# GYRESM 1 40
# GPRESM 1 40
NOTE 1
PLOT 3
# GMRESY 1 40
# GYRESY 1 40
# GPRESY 1 40
NOTE 1
PLOT 3
# GMRESP 1 40
# GYRESP 1 40
# GPRESP 1 40
NOTE 1
```

```
*****  
* CALCULATE THE CUMULATED RESPONSE IN M, Y, P TO *  
* ONE-TIME INNOVATIONS IN GM, GY, GP. THE *  
* CUMULATED RESPONSES ARE STORED IN THE VARIABLE *  
* CXRESZ: "CUMULATED X RESPONSE TO AN INNOVATION *  
* IN Z." THESE RESPONSES ARE STORED IN VARIABLES *  
* 23-25, 26-28, 29-31. *  
*****
```

```
DO I=14, 22  
IEVAL CUM=I+9  
ACCUMULATE I 1 40 CUM 1  
END DO I  
EQV 23 TO 31  
CMRESM CYRESM CPRESM $  
CMRESY CYRESY CPRESY $  
CMRESP CYRESP CPRESP  
NOTE 1  
PLOT 3  
# CMRESM 1 40  
# CYRESM 1 40  
# CPRESM 1 40  
NOTE 1  
PLOT 3  
# CMRESY 1 40  
# CYRESY 1 40  
# CPRESY 1 40  
NOTE 1  
PLOT 3  
# CMRESP 1 40  
# CYRESP 1 40  
# CPRESP 1 40  
END
```