Staff Papers Series

Staff Paper P86-28 September 1986

A BEGINNER'S GUIDE TO VECTOR AUTOREGRESSION

by

Steve Ford

Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

Staff Paper P86-28 September 1986

A BEGINNER'S GUIDE TO VECTOR AUTOREGRESSION

by

Steve Ford

I. INTRODUCTION

The objective of this paper is to provide a starting point for those interested in using vector autoregression (VAR) techniques. It is assumed that the reader has a basic previous understanding of time series estimation and notation. With this purpose in mind, references will be made to VAR uses in the literature and the theory behind VAR's. This paper will also provide example code from the Regression Analysis of Time Series (RATS) econometric package to illustrate different types of VAR analysis. This paper, then, is essentially a companion guide to the sections of the RATS manual dealing with VAR techniques. Since this paper is applied in nature, only cursory discussion will be given to the development of VAR theory. References to several articles will be made instead.

Section II of this paper discusses the use of unrestricted VAR's and Section III discusses VAR's restricted with Bayesian priors. Section IV discusses forecasting with VAR's. Section V discusses extensions of VAR analysis beyond its forecasting use to impulse response analysis and decomposition of variance. Each of these sections includes discussions of the literature relevant to the topic at hand and annotated RATS code used for similar problems. Section VI concludes this paper with some ideas on future uses of VAR techniques. A list of variable definitions appears in the appendix.

II. UNRESTRICTED VAR'S

This section will briefly discuss the theory behind unrestricted VAR's and their intuitive appeal. Again, note that the purpose of this paper is not to fully develop the theory behind VAR's, but is instead to present an applied background for empirical applications of these techniques. References are made to important sources in the literature for those wishing to explore the underlying theory further.

A. Unrestricted VAR's

A vector autoregression is essentially a system of equations whose dependent variables are regressed on lagged observations of all the variables in the system. Exogenous or deterministic variables thought to be important can also be included in the system, however, these variables would not appear as regressands in the system.

Using Sargent's [28] notation, an M^{th} order (M lags) VAR for the (N x 1) wide sense stationary stochastic process, $\{Z_t\}$, is

$$Z_t = \sum_{j=1}^{M} D_j^M Z_{t-j} + \eta_t^M \qquad (1)$$

where the D_j^M's are (N x N) matrices, and the (N x 1) stochastic error process, η_t^M, satisfies the orthogonality conditions $E_t [\eta_t^M Z_{t-k}'] = 0$, $k=1,\ldots,M$[1].

The form of a VAR comes from the Wold Decomposition Theorem and the ability to model a moving average representative of a covariance

---

[1] There are t=1, ..., T observations of each of the N variables in Z_t in this VAR. The VAR is written with M lags.

stationary process with autoregression techniques. Good discussions of this theoretical development are found in Sargent [29], Orden [25], Litterman [15], and Nerlove, Grether and Carvalho [24]. These references explain the Wold theorem and show that any stochastic stationary process can be represented by the sum of deterministic and nondeterministic terms. The nondeterministic term is represented as a moving average. Litterman [15] shows that since this moving average representation, if invertible, can be represented as an infinite-order autoregression, then it can be approximated reasonably well by a finite order autoregression. This argument in support of univariate autoregression is then extended to vector analysis.

B. Intuitive Appeal of VAR's

A VAR is really a set of reduced form equations similar to what might be derived from a structural econometric model that relies heavily on economic theory. Such structural models make highly restrictive assumptions on the values of estimated parameters in their reduced form equations. These restrictions most often take the form of exclusions of variables or lags of variables from these models. This in effect restricts the estimated parameters of these "excluded" variables to zero. This implies that these variables have no predictive or explanatory power in the model. A VAR, however, will include some of these variables, relying on a much less restrictive concept of economic theory as it applies to these reduced form equations. Data is then allowed to determine the contribution of variables instead of a priori economic structure.

Borrowing a simple example from the literature (Fisher [10]), let the supply and demand equations for a commodity be:

$$Q_t + \alpha_{12}P_t^* + \gamma_{11}PI_t + \gamma_{13} = u_{1t} \text{ (supply)} \tag{2}$$

$$\beta_{21}Q_t + P_t + \gamma_{22}Y_t + \gamma_{23} = u_{2t} \text{ (demand)} \tag{3}$$

where Q_t = quantity supplied in period t

P_t^* = expected price in period t

PI_t = price of inputs in period t

P_t = actual price in period t

Y_t = disposable income.

u_{it} = error terms on equation i

Note that the market clears in each period.[2] Fisher develops this simple model in the context of rational expectations in P_t^*. The resulting reduced form equations are:

$$Q_t = \frac{-\alpha_{12}\beta_{21}}{w}\hat{PI}_t + \frac{\alpha_{12}\gamma_{22}}{w}\hat{Y}_t + \left[\left(\frac{\alpha_{12}\gamma_{23} - \alpha_{12}\beta_{21}\gamma_{13}}{w}\right) - \gamma_{13}\right] + v_{1t} \tag{4}$$

$$P_t = \frac{\beta_{21}^2\alpha_{12}}{w}\hat{PI}_t - \frac{\beta_{21}\alpha_{12}\gamma_{22}}{w}\hat{Y}_t + \beta_{21}\gamma_{11}PI_t - \gamma_{22}Y_t \cdot$$

$$+ \left[\left(\frac{\beta_{21}^2\alpha_{12}\gamma_{13} - \beta_{21}\alpha_{12}\gamma_{23}}{w}\right) + \beta_{21}\gamma_{13} - \gamma_{23}\right] + v_{2t} \tag{5}$$

---

[2] I have removed a fertilizer price subsidy variable from the supply equation to simplify the model.

where, $w = 1 - \beta_{21}\alpha_{12}$

v_{it} = reduced form disturbances on equation i

and $\hat{}$ denotes an expectation.

One can see that the rational expectations and the economic structure imposed on the form of the supply and demand equations have carried over into the reduced form equations in two ways: 1) each of these equations has expectations variables as regressors, expectations in this model being made at some prior time; and 2) there are important non-linear cross-equation restrictions on the model.

The forecasting ability of this model obviously depends on the economic structure imposed on the data a priori. This structure may not, however, represent exactly the true forces at work in the market. It also does not allow the dynamics of the market to enter the model other than in the expectations for the exogenous variables; input prices and disposable income. Additionally, the model is highly dependent on the method of generating expectations for these exogenous variables.

Vector autoregression offers an alternative to this structural model. For the model above, a four variable VAR including price, supply, input price, and income as the system variables could be estimated. Such a VAR would be consistent with the economic structure that was imposed on the model previously. Any cross-equation restrictions would be linearly approximated, in a sense, by the VAR, and would reflect the true but unknown relationships among the parameters as the data suggest. Also, expectations for all

variables are now generated within the model, taking advantage of any dynamic relationships that may exist among the variables.

A six lag VAR estimation of the above model would look like:

$$Q_t = a_1 + \sum_{i=1}^{6} b_{1i}^{1} Q_{t-i} + \sum_{i=1}^{6} b_{2i}^{1} P_{t-i} + \sum_{i=1}^{6} b_{3i}^{1} PI_{t-i} + \sum_{i=1}^{6} b_{4i}^{1} Y_{t-i} + v_{1t}$$

$$P_t = a_2 + \sum_{i=1}^{6} b_{1i}^{2} Q_{t-i} + \sum_{i=1}^{6} b_{2i}^{2} P_{t-i} + \sum_{i=1}^{6} b_{3i}^{2} PI_{t-i} + \sum_{i=1}^{6} b_{4i}^{2} Y_{t-i} + v_{2t}$$

$$PI_t = a_3 + \sum_{i=1}^{6} b_{1i}^{3} Q_{t-i} + \sum_{i=1}^{6} b_{2i}^{3} P_{t-i} + \sum_{i=1}^{6} b_{3i}^{3} PI_{t-i} + \sum_{i=1}^{6} b_{4i}^{3} Y_{t-i} + v_{3t}$$

$$Y_t = a_4 + \sum_{i=1}^{6} b_{1i}^{4} Q_{t-i} + \sum_{i=1}^{6} b_{2i}^{4} P_{t-i} + \sum_{i=1}^{6} b_{3i}^{4} PI_{t-i} + \sum_{i=1}^{6} b_{4i}^{4} Y_{t-i} + v_{4t}$$

Note that a constant has been included in the VAR system. There are N=4 variables with M=6 lags each.

The right hand sides of each of these four system equations contains exactly the same variables and lags of variables. Consequently, the moment matrix (X'X) will be the same for each equation. In this case where the right hand sides are the same and are predetermined the system can be estimated equation by equation using ordinary least squares (OLS) with no simultaneous equation bias. This means that no special software is required to estimate unrestricted VAR's. Each equation can be estimated independently using OLS with exactly the same results.

It is intuitively appealing to let market data show how particular components of the market, the variables, interact. However, it is important to note that VAR's are not totally devoid of economic theory and structure. The underlying theory and any hypothesized structure indicate to the economist which variables to include in the model and how many lags would be appropriate. The method of determining the appropriate lag length is still an important issue in the literature on unrestricted VAR's (Webb [39], Bessler [2], Doan and Litterman [7]).

C. Determination of Lag Length

There have been several methods proposed to deal with the problem of correctly determining the proper lag length for an unrestricted VAR. A lag structure of any particular length essentially states that lags from any longer period do not add information to the model and have parameter values of zero. Longer lag lengths also increase the number of estimated parameters, reduce degrees of freedom, and increase data requirements. Hence, it is important for the investigator to take care in choosing the lag length of the model.

The technique used by Sims [31] involves a likelihood ratio between models of different lag lengths. The ratio is of the form:

$$(T-C)\ [\ln |\Sigma_1| - \ln |\Sigma_2|]$$

where T = number of observations

C = a correction factor to bring the test statistic closer to its asymptotic distribution

$|\Sigma_i|$ = the determinant of Σ_i, the covariance matrix (cross products) of the residuals from VAR system i.

This statistic has the chi-squared distribution. Sims states, however, that this procedure is somewhat ad hoc in nature and more work must be done in this area.[3]

An alternative to Sims method is to first determine the statistic

$$M(k) = -(T - 1/2 - M \cdot N) (\ln |\Sigma_1| - \ln |\Sigma_2|)$$

where M is the autoregressive order, N is the number of variables, and all other variables remain the same as above (Brandt and Bessler [4]). This statistic is used to determine the lag length of the VAR and is asymptotically distributed chi-squared with m^2 degrees of freedom. Once the lag length is determined, the VAR is reestimated with statistically insignificant lags on the variables deleted from the model. While Sims includes all intermediate lags, this method removes many, and results in a model with fewer parameters.

The final method of lag length determination discussed here is that used by Webb [39]. Lag lengths are chosen to minimize the Akaike Information Criterion (AIC):

$$AIC = (T-p) \ln \sigma^2 + 2p$$

---

[3] Sims [31].

where T = number of observations

p = number of estimated parameters

σ^2 = estimate of the residual variance.

This criterion functions similarly to the adjusted R^2 statistic. As the number of estimated parameters increases, σ^2 is reduced, as is the first term in parentheses. However, the second term, 2p, increases. Consequently, the inclusion of additional parameters may actually increase the AIC in some instances as a reduction in residual variance is out-weighed by the increase in the number of parameters.

The actual procedure used by Webb is quite complex. Basically, a search procedure is used to avoid comparing pairwise the vast number of possible exclusions on intermediate lag lengths. The precise procedure used is described in the appendix to Webb's article.

D. Basic References

Good references on the statistical and time series theory behind unrestricted VAR's are Granger and Newbold [12] and Nerlove, Grether, and Carvalho [24]. Another more basic discussion of time series analysis appears in Nelson [23]. Good treatments of this subject matter also appear in Sargent [29].

Perhaps the most cited article dealing with VAR's is Sims' [31] original piece. Sims presents good intuitive arguments for the use of VAR's and a very rigorous treatment of the statistical theory. Beyond this piece, other good general treatments of VAR methodology appear in Orden [25], Webb [38], Brandt and Bessler [4], Hakkio and

Morris [13], the RATS manual [7], and Sargent [28].

Applications of VAR methodology have been made to: the Brazilian economy (Bessler [1]), the U.S. agricultural financial market (Chambers [5]), the U.S. hog market (Brandt and Bessler [4]), U.S. agricultural trade (Orden [25, 26, 27]), and the U.S. macro-economy (Hakkio and Morris [13]). Thus, a wide variety of applications of VAR methodology is possible, and many different types of questions can be answered with these techniques.

E. VAR's and RATS

This section will list the RATS code used for unrestricted VAR's and the determination of lag length as proposed by Sims. The code is annotated to clarify the discussion of these commands presented in the RATS manual. Not all options of the commands will be presented, and only the code required for the VAR's is presented. Data handling commands and the setting of variables are not discussed. These commands are discussed in the RATS manual.

The series of commands defining a four variable, six lag VAR with a constant appears below:

```
SYSTEM 1 TO 4
VARIABLES X1 X2 X3 X4
LAGS 1 TO 6
DET CONSTANT
END(SYSTEM)
```

The SYSTEM command defines the equation numbers involved in the VAR system. VARIABLES provides a list of the system variables and LAGS provides a list of the lag lengths to include in the system.

DET lists any deterministic or exogenous variables. These variables will be different from those listed on the VARIABLES command. This command may be omitted if no constant or other exogenous variable is to be included. The final command for the system definition sequence is END(SYSTEM). This is required. The discussion of system specification in the RATS manual is quite good.

Once the VAR is defined it is estimated with the ESTIMATE command. The basic form of this command is (for example):

```
ESTIMATE 70,1 80,4 9
```

This example estimates a previously defined system over the period from first quarter 1970 to fourth quarter 1980. Residuals from this estimation are put in several series beginning with series 9.

Caution is warranted on two points. First, it is important to allow for lag length when setting the range of the estimation. For a six lag system, the above ESTIMATE command would require data back to third quarter 1968 because of the lag structure.

Second, residuals must be saved to data series if they are required later for further analysis. In the above example the residuals are written to series nine. (This parameter is unnecessary if the residuals are not to be saved.) The 9 represents data series nine which must be free or the series will be overwritten, resulting in a loss of any previous series nine. It may be necessary to adjust the ALLOCATE command to create this series (See the RATS manual). A separate data series is required for the residuals from each equation

in the VAR system. If the VAR has four equations, then the above example would store the residuals from this system in series nine through twelve in the order of the variables listed in the VARIABLES command. It is important, then, to be careful in defining series of residuals so that series containing data are not overwritten.[4]

The RATIO command is used to test for the appropriate lag length of the model using the likelihood ratio method used by Sims [31]. It takes the form:

```
RATIO (DEGREES=16,MCORR=25) 70,1 80,4
#9 TO 12
#13 TO 16
```

This command would be used to test the difference between the system defined previously and the system defined below:

```
SYSTEM 5 TO 8
VARIABLES X1 X2 X3 X4
LAGS 1 TO 5
DET CONSTANT
END(SYSTEM)
ESTIMATE 70,1 80,4 13
```

The likelihood ratio is calculated over the same period as the estimation of each of the VAR's. It is testing for a significant difference between a five lag model and a six lag model. The series containing the residuals from the two VAR's are listed on the two

---

[4] The microcomputer version of RATS handles data series slightly differently. Care must still be taken to make sure residual series do not overwrite some other series.

supplementary lines. DEGREES tells RATS the number of restrictions placed on the model. In this example we have one lag of four variables in each of four equations restricted to zero. This results in sixteen restrictions. The MCORR parameter is the adjustment Sims includes in this test. It is the number of regressors in each unrestricted regression (four variables times six lags plus a constant).

There is a wide variety of options to what is presented above. The code listed above provides a good starting place for estimating generic unrestricted VAR's.

III. BAYESIAN VECTOR AUTOREGRESSION

The use of Bayesian priors on parameter values in a VAR simply involves the incorporation of our prior beliefs about parameter values and their distributions into the estimation technique. The exclusion of variables in a structural model really involves the imposition of our very strong prior beliefs that parameter values are zero with certainty. Incorporating such strong beliefs about parameter values, even if non-zero, creates the same problems encountered with a priori economic structure discussed in the previous section. Consequently we wish to make broader, more general assumptions about the distributions of parameter values. These priors have sometimes been referred to in the literature as Minnesota or Litterman priors.

A. Underlying Theory

A prior distribution can be created for each estimated parameter. With a quarterly VAR of five variables and six lags, this would involve thirty parameters per equation (assuming no constant) times five equations for a total of 150 prior distributions to specify. Although these can be specified individually, it is by no means an easy task. Also, the individual specification of each parameter's prior could be subject to critical arguments about its specification. Each parameter's prior would have to be justified. Instead, a systematic method to specify more general and accessible priors is developed that alleviates the task of specifying each prior individually, weakens the basis for criticism of individual priors, and allows reproduction by other investigators.

The general specification of the Litterman priors revolves around the assumption that each equation follows a random walk process; $Y_t = Y_{t-1} + v_t$. That is, the expectation of this period's value of the dependent variable is simply last period's observation of that value. This means that in the model presented earlier in equation (1), $E[D_1^1] = 1$ and $E[D_j^m] = 0$ for all $j \neq 1$ and $m \neq 1$. Thus, the prior mean for the parameter on the first lag of the dependent variable will be one and the prior mean on all other variables will be zero.

Since such naive forecasts are restrictive and unsatisfactory for a variety of reasons, prior distributions are placed around these means. One assumption of these distributions is that variable lags further into the past have less explanatory power than more recent lags. The resulting distributions are illustrated in Figure 1.

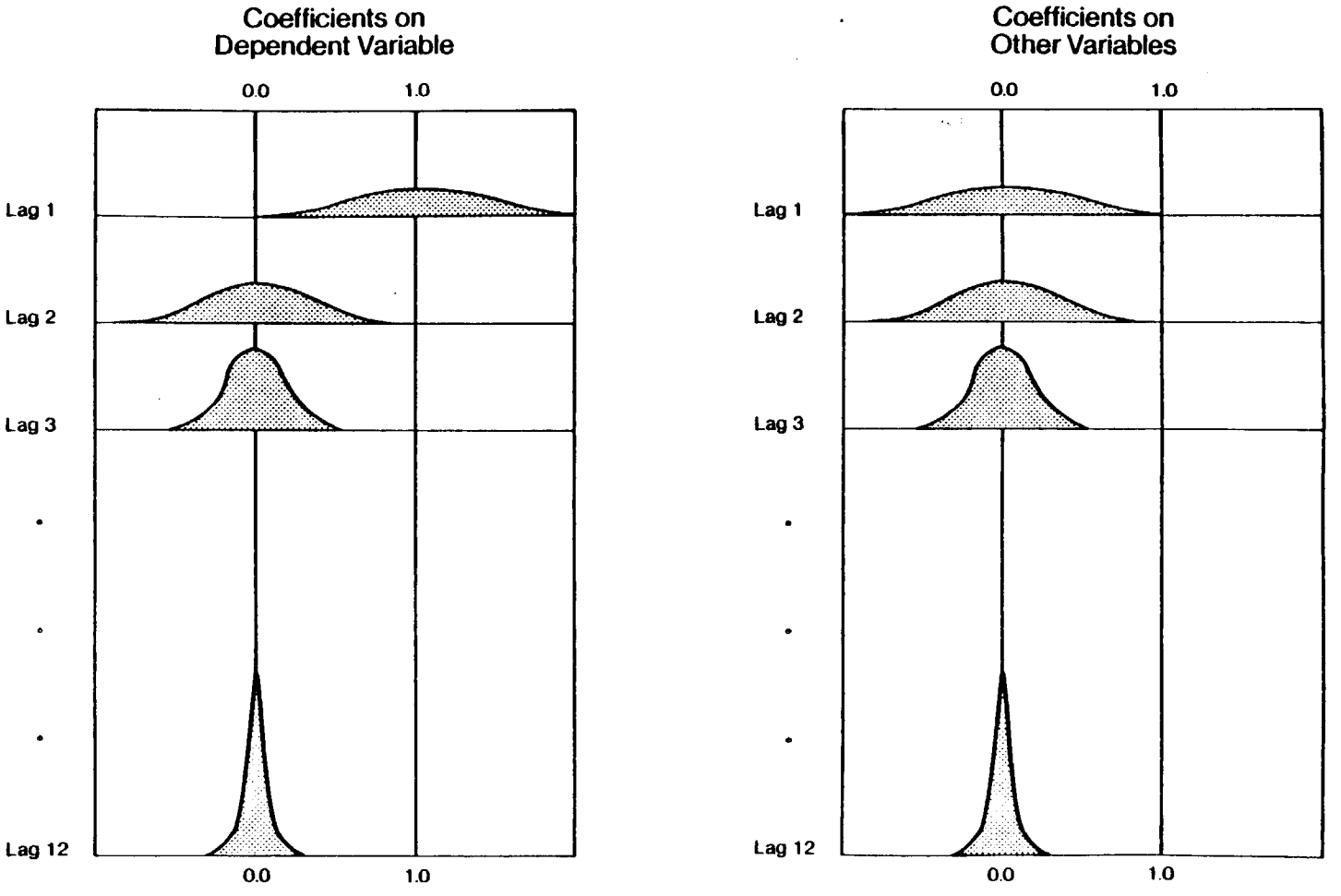
Distributions around the prior means must still be quantified. These distributions are specified with general priors imposed in the form of standard deviations of the estimated parameters. The Litterman prior on standard deviations is of the form[5]

$$\delta_{ij}^{\ell} = \begin{cases} \dfrac{\lambda}{\ell^{\gamma 1}} & \text{if } i = j \\ \dfrac{\lambda\ \gamma_2\ \hat{\sigma}_i}{\ell^{\gamma 1}\ \hat{\sigma}_j} & \text{if } i \neq j \end{cases}$$

[5] See Litterman [15], and Bessler and Kling [3]. Note that the RATS manual [Doan and Litterman] is incorrect in its specification of this prior. The subscripts on the scale factors, σ_i and σ_j, are reversed.

FIGURE 1

A Schematic Representation of the Prior *



* Source: Litterman [17]

where δ_{ij}^{ℓ} is the standard deviation of the coefficient for variable j with lag ℓ in equation i.

The investigator must specify three parameters (λ, γ_1, γ_2) to derive the standard deviations. λ is the constant overall tightness of the prior.[6] As this parameter is set very tightly (approaches zero), the estimated coefficients in the model approach their prior means since the distributions around these means become spiked.

γ_1 is a decay parameter which determines the rate at which lags farther back receive less weight (become tighter around their means). Bessler and Kling use a harmonic lag decay of the form $g(\ell) = \ell^{-\gamma_1}$. They note, however, that $\gamma_1 = 0$ gives better results than other values of γ_1. This means that there is no decay, and that past lags receive weights equal to more recent lags. This structure would not look like Figure 1. Instead all lags would have distributions similar to lag 1 around their respective means. An alternative to harmonic decay is geometric decay of the form $g(\ell) = \gamma_1^{(\ell-1)}$. This specification decays much more rapidly than the harmonic lag specification.

Finally, γ_2 is used to specify the relative weights of variables in each equation. Own lags of dependent variables typically carry a

---

[6] λ is described by Litterman [15] and Bessler and Kling [3] as the "constant standard deviation on the first lag of the dependent variable in each equation." For higher lags and other variables, λ is then adjusted by the decay term and the appropriate weights, $\gamma 2$. This interpretation is correct in the true form of the Litterman prior. However, if the prior structure is altered so that the random walk assumption is changed, λ should be reinterpreted as a general prior standard deviation for <u>all</u> lags which is modified for individual sets of parameters.

weight of 1.0. Other variables would be assigned weights ranging from 0.0 to 1.0. Other variables can receive the same weight in each equation or the weights can be more finely tuned to each individual equation (Bessler and Kling [3]).[7]

Before estimation, priors are scaled by a ratio of standard errors of univariate autoregressions of the same lag length to be estimated in the VAR. $\hat{\sigma}_i$ is the standard error of residuals from such an autoregression for variable i. This scaling is to ensure that individual variables do not receive inappropriate weights in their contributions to the VAR merely because of the magnitude of their units of measurement.

Once the prior means and standard deviations of the parameters are specified, the equation coefficients can be estimated using a form of Theil mixed estimation (Theil). This is a method of incorporating prior information about the equation parameters into the estimation procedure.

Let $y = X\beta + u$ be the general linear statistical model, and let our series of "dummy observations" on the parameter values be described as $r = R\beta + v$. ($var(u) = var(v) = \sigma^2$.) The Theil mixed estimator is

$$\beta_m = (X'X + R'R)^{-1} (X'y + R'r). \qquad (6)$$

Given the prior on an individual parameter as $\beta_i \sim N(b_i, \Theta_i^2)$, an individual restriction on βi is $r_i = R_i\beta_i$ where $R_i = \sigma/\Theta$ and $r_i =$

---

[7] See the RATS manual [Doan and Litterman] for "symmetric" and "general" prior specifications.

$(\sigma/\Theta_i)\ b_i$. b is the prior mean. For a set of these restrictions, in matrix notation, R becomes a diagonal matrix with σ/Θ_i in the ii^{th} entries along the diagonal. With mean 1 on the first lag of the dependent variable, r becomes a vector with σ/Θ as the entry in the $(\ell*(i-1)+1)^{th}$ cell and zeros everywhere else, where this would be the entry for the first lag of the dependent variable in equation i.[8]

Since σ^2 is unknown, we substitute an estimate, s^2, where s^2 is the variance of the residuals from a univariate autoregression of the same lag length as the VAR. We have specified our Θ's previously using the notation δ_{ij}^{ℓ}. The entries in the R and r arrays then become s/δ_{ij}^{ℓ}. This is a slightly different approach taken from that appearing in Litterman [15] or Bessler and Kling [3]. However, it is consistent with the RATS manual and the way RATS computes the priors.[9] The R and r arrays as specified above are then used to derive the Theil mixed estimation parameter estimates.

One further comment is necessary to clarify the degrees of freedom reported by RATS. Since dummy observations are added for lags of each system variable, the number of observations increases. The degrees of freedom are no longer T-(N*M)-W as one would expect in

---

[8] This notation is for single equation estimation. The arrays R, r, and b would expand if we thought of this procedure as estimating the system simultaneously.

[9] Bessler and Kling have misspecified the array r; in their paper. r is not simply the vector of means (ones and zeros). That vector is b. However, in the case of the Litterman priors λ is the standard deviation of the parameter of the first lag of the dependent variable and their specification of r works. A more general specification appears in Ford [11].

a normal regression, where T is the number of observations over time, N is the number of system variables, M is the number of lags on those variables, and W is the number of deterministic or non-system variables. Instead they are T-W.

B. References to Bayesian VAR's in the Literature

By far, the most important contribution to VAR estimation with Bayesian priors is that of Litterman [15]. Other work in this area has been contributed by Doan, Litterman, and Sims [8], Todd [37], and Doan and Litterman [7].

There have been few published applications of Bayesian vector autoregressions to agricultural markets. The article by Bessler and Kling [3] is the only instance at this time. Obviously, there are many opportunities open in this area for work with Bayesian VAR's.

C. Bayesian VAR's and RATS

Bayesian priors are included in VAR estimation in RATS through the inclusion of the SPECIFY command in the system definition. This appears as:

```
SYSTEM 1 TO 4
VARIABLES X1 X2 X3 X4
LAGS 1 TO 6
DET CONSTANT
SPECIFY (TIGHT=λ,DECAY=γ1,LAGTYPE=HARMONIC,TYPE=SYMMETRIC) γ2
END(SYSTEM)
```

where λ, γ_1, and γ_2 are defined as described previously.

The λ and γ_1 parameters require little further explanation. LAGTYPE can either be HARMONIC or GEOMETRIC as discussed previously. The greatest flexibility RATS provides is in the TYPE specification

and in γ_2. Scaling is done automatically by RATS.

The SPECIFY command above uses a symmetric prior where γ_2 is set the same for lags on the non-dependent variables. This is the SYMMETRIC specification. An alternative would be that used by Bessler and Kling in which a more specific determination of γ_2 is used. This is the TYPE=GENERAL specification. This takes the form:

```
SPECIFY (TIGHT=λ,DECAY=γ1,LAGTYPE=HARMONIC,TYPE=GENERAL)
# 1.0 .8 .7 .6 $
  .9 1.0 .7 .6 $
  .9 .8 1.0 .6 $
  .9 .8 .7 1.0   .
```

In this example, X1 would have a weight of $\gamma_2 = .9$ in all equations other than its own, X2 a weight of $\gamma_2 = .8$ in all equations other than its own and so on. In this manner the γ_2 parameter can be set to reflect the investigator's prior on the relative importance of each variable in each equation.

A third method of specifying the prior is to use the FULL option which takes the form:[10]

SPECIFY(FULL=ARRAY)

where ARRAY is a rectangular matrix containing entries of the δ_{ij}^{ℓ} in a particular format. The scaling can be left up to RATS if the SCALE option of the ESTIMATE command is used.

---

[10] A fourth method involves the Circle-Star type of prior specification. Litterman [15] and the RATS manual provide good explanations.

A final method of estimating a VAR with Bayesian priors is to use Theil mixed estimation. In this method, the R and r arrays are set with a prior structure and then the model is estimated equation by equation using equation (6). RATS can be programmed to do this estimation, however it is somewhat complicated and a good working knowledge of RATS is required (See the RATS manual). This method does, however, provide a greater degree of flexibility in choosing prior structure.

IV. FORECASTING WITH VAR'S

One of the advantages VAR models have over other types of models is their forecasting accuracy. This has been shown to be especially true when Bayesian priors are used (Bessler and Kling [3], Litterman [15]). Two types of forecasting can be done with VAR's. Both involve the use of the chain rule of forecasting. One method, however, updates equation parameters as the model progresses through time. These two methods will be discussed below, as will their use in RATS.

A. The Chain Rule of Forecasting

For our vector stochastic process (Z_t) and our estimated VAR with parameters D_j, a one step ahead forecast would be[11]

$$E_t[Z_{t+1}] = \sum_{j=1}^{m} D_j \, Z_{t+1-j}.$$

If forecasts were being made farther into the future to time t+2, for example, this method could not be used since Z_{t+1} is unknown. The chain rule of forecasting allows $E_t[Z_{t+1}]$ to be substituted for Z_{t+1}. The two step ahead forecast then would be

$$E_t[Z_{t+2}] = D_1 E_t[Z_{t+1}] + \sum_{j=2}^{m} D_j \, Z_{t+2-j}.$$

This process could feasibly be repeated for some k-step ahead forecasts. The general form of this method for a k-step ahead

---

[11] See Sargent [28].

forecast is

$$E_t[Z_{t+k}] = \sum_{j=1}^{M} D_j \, E_t[Z_{t+k-j}], \quad k \geq 1$$

where $E_t[Z_{t-j}] = Z_{t-j}$ for $j \geq 0$.

The chain rule of forecasting allows forecasts to be made to arbitrarily long intervals into the future. The accuracy of these forecasts of course diminishes with longer forecast intervals. Note, however, that this forecast mechanism works only for VAR's that are "self-contained." Forecasts of all system variables can be fed back into the system representing future values of these variables in forecasts of longer intervals. If exogenous or non-system variables for which no estimating equations exist are included in the model, then only one-step ahead forecasts can be made. If future values of these variables are known, are generated by some external statistical process, or are assumed to be at some particular level, then these values can be inserted into the system as forecasts are made, and the forecasts will be conditioned on these values.

B. Forecasting With the Kalman Filter

The Kalman Filter offers an inexpensive method of updating the estimated parameters of an equation as information is added over time. For example, in the previous discussion on the chain rule of forecasting, $E_t[Z_{t+1}]$ is treated as an actual observation of Z_{t+1} in forecasts beyond time t+1. If for some a priori reasons the investigator feels that the estimated parameters change over time

from structural change or some other occurrence, then $E_t[Z_{t+1}]$ would be used as an actual observation in the reestimation of the VAR. New estimates would then be used for a one step ahead forecast and the process would contrive for longer step ahead forecasts.

Given the general linear statistical model $y_t = X_t\beta_t + u_t$ with $var(u_t) = \gamma_t$, two additional assumptions are necessary.

1) The coefficient vector follows a random walk, i.e.,

$$\beta_t = A_t\beta_{t-1} + V_t \qquad \text{with } var(V_t) = M_t.$$

2) u_t and V_t are independent.

The case which is most often used is where $A_t=I$ and $M_t=0$. This essentially indicates that the true parameters themselves do not change, but as more information becomes available better estimates of them can be made.

The Kalman Filter <u>update</u> estimator is:

$$\beta_t = \beta_{t-1} + \Sigma_t X_t' \eta_t^{-1}(y_t - x_t\beta_{t-1}) \tag{7}$$

where,

$$\Sigma_t = S_t - S_t X_t'(X_t S_t X_t' \eta_t)^{-1} X_t S_t$$

$$S_t = \Sigma_{t-1} + M_t$$

$$\Sigma_{t-1} = \text{cov}\ (\beta_{t-1})$$

The new estimate of β_t in equation (7) can be decomposed into three parts. It is the previous period's parameter estimate altered by a

term composed of the filter gain, $\Sigma_t X_t^1 \eta_t^{-1}$, and the one step ahead forecast error, $y_t - X_t\beta_{t-1}$. This is better seen as:

$$E[\beta_t|y_t, Y_{t-1}] = E[\beta_t|Y_{t-1}] + k_t(y_t - E[y_t|Y_{t-1}])$$

where given information through t-1 in the form of the data, Y_{t-1}, the expected value of β_t with added information, y_t, is equal to β_{t-1} (since $E[\beta_t|Y_{t-1}] = \beta_{t-1}$) plus a coefficient times the deviation of Y_t from its conditional mean. The Kalman filter, is the coefficient k_t.[12]

Again, the advantage of using the Kalman filter is that it is unnecessary to re-regress an entire system as observations are added. This will become an obvious time-saver when forecast statistics are discussed in the next sections.

C. Testing Forecasting Ability

Two forecast statistics will be discussed here. The first is the root mean squared forecast error of k-step ahead forecasts. It takes the form

$$RMSE = +[\Sigma_t (F_{t+k} - A_{t+k})^2/n]^{1/2}$$

where F_i = k step ahead forecast

A_i = actual value of the forecast variable

n = number of forecasts in the sample

---

[12] A thorough discussion of the Kalman filter is given in Chow [6].

This statistic will be discussed further in the next section. Note at this point, however, that this statistic gives a good estimate of the magnitude of forecast errors in the units of the forecast variable.

The second test statistic is the Theil U statistic. It takes the form:

$$U = + \left(\frac{\sum_t (F_{t+k} - A_{t+k})^2/n}{\sum_t (A_t - A_{t+k})^2/n} \right)^{1/2}$$

where the variables are defined as for RMSE. This statistic compares the forecasting ability of the model, measured in RMSE, to the naive forecast of no change. This statistic is useful because it is not unit dependent and therefore cross equation comparisons of forecasting ability can be made. Theil U values between zero and one imply that the model does a better job of forecasting than the naive forecast.

D. Testing Model Selection Through Forecasting Ability With RATS

The primary command used to calculate forecast statistics in RATS is THEIL. The output from this command includes the mean forecast error, the mean absolute forecast error, the root mean squared forecast error, and the Theil U statistic for each of the k-step ahead forecasts requested. Example RATS code for the system defined in section III is:

```
(1)  THEIL(SETUP) 4 12 80,4
(2)  #1 TO 4
(3)  ESTIMATE(NOPRINT,NOFTESTS) 70,1 75,4
(4)  DO DATE=(76,1),(80,4)
(5)  THEIL DATE
(6)  KALMAN 0 0 DATE.EQ.(80,4)
(7)  END DO
(8)  THEIL(DUMP)
```

These statements are numbered for the convenience of discussion.

The first statement sets up the procedure to calculate and compile the forecast statistics. This example tells RATS that there are 4 equations, statistics are desired for up to 12-step ahead forecasts, and the sample data ends with fourth quarter, 1980. The second statement indicates which four equation numbers are to be used, as defined in the SYSTEM definition.

Statement (3) computes the initial equation parameter estimates of the previously defined system. This estimate is over a subset of the entire data series. Forecast statistics will be calculated from this initial estimate and then the parameter estiamtes will be updated with the Kalman filter and new statistics will be calculated. It is convenient to include the NOPRINT and NOFTEST options of the ESTIMATE command. Otherwise, parameter estimates and F-tests of this initial model will be printed for each new observation.

Statements (4) through (7) form a loop which does the statistical calculation and parameter estimate updates over the period (76,1) to (80,4). THEIL DATE calculates the forecast statistics on forecasts beginning with the period associated with DATE in each loop. Statement (6) updates the parameter estimates. The two 0's indicate that residuals and parameter values are not to

be saved to series. The regression output will be printed when DATE equals 80,4. The final statement, THEIL(DUMP), lists the compiled forecast statistics.

If this series of statements is used to generate forecast statistics as model selection criteria, selection is based, then, on the performance of forecasts from models estimated on subsets of the data sample. In actuality it is not so much the final parameter estimates which are being compared between two models, but the specification of the models instead. The comparison is really of chosen variables, lag length, priors, and deterministic variables in the models under consideration. The forecasting ability of such a specification through time becomes the issue of importance.

Mention should be made of other forecasting command in RATS. RATS offers a wide variety of forecasting options. The two most important of these commands for forecasting VAR's are FORECAST and SIMULATE.

FORECAST applies the chain role of forecasting to the estimated VAR. It is used as:

```
FORECAST 4 20 81,1
#1 X1 81,1
#2 X2 81,1
#3 X3 81,1
#4 X4 81,1
```

This set of commands indicates that RATS is to forecast a set of four equations for twenty periods beginning in the first quarter, 1981. The four equations are then listed with forecasts for equation 1

being entered in series X1 beginning with 81,1. It is important to remember that as forecasts are generated and written to a series, the original series may be overwritten. Care must be taken to ensure this is done correctly. Also, the ALLOCATE series may have to be altered to accommodate series generated into the future. Any future exogenous or deterministic variables necessary to be forecast must be provided as well.

SIMULATE does the same thing as FORECAST except random shocks are added to the system equations. FORECAST generates only deterministic forecasts. Rather than describe this command here, the RATS manual presents the command in sufficient clarity.

Two additional forecasting tools, of a kind, are ERRORS and IMPULSE. These commands are discussed in the following section.

V. IMPULSE RESPONSE ANALYSIS AND DECOMPOSITION OF VARIANCE

Two uses of VAR models other than for forecasting are decomposition of variance and impulse response analysis. These tools can be used to analyze the data in terms of the relationships among variables, given the estimated model.

A. Impulse Response Analysis

Impulse response analysis looks at the effects to a system of an exogenous shock to one of the variables in that system. For example, it may be used to predict what will happen to hog prices if the price of corn increases because of some force outside the model.

This analysis is based on the moving average transformation of the autoregressive model. If our autoregressive model is

$$Zt = \sum_{j=1}^{m} D_j Z_{t-j} + \eta_t$$

then the moving average representation is

$$Z_t = \sum_{j=0}^{\infty} H_j \, \eta_{t-j}.$$

Thus, the effects of any unexpected shock to the system can be traced through deviations of the shocked time paths from the expected time paths given by the model. This technique is quite useful in certain types of policy analysis and sensitivity analysis. It is also useful in analyzing the dynamic interrelationships among variables in the model.

One problem that often arises in this type of analysis and in the decomposition of variance is the contemporaneous correlation of forecast errors; i.e., the covariance matrix of the error terms $\Sigma = E\eta_t\eta_t^1$ is not diagonal. Because of this non orthogonality, the obvious step is to orthogonalize the innovations (errors).

Although there are many methods one can use to orthogonalize these innovations, the most common method used in VAR work has been the Choleski decomposition. This technique factors the covariance matrix of the errors such that $\text{cov}(\eta) = \Sigma = SS'$ so that $\eta = Sv$ and $\text{var}(v) = I$. v, then, is the orthogonalized error vector. These orthogonal innovations are then traced over time in response to shocks to the system. More complete treatments of this technique appear in Litterman [15], Doan and Litterman [7], Orden [25,26] and Bessler [2].

B. Decomposition of Variance

The decomposition of forecast variance allows one to attribute portions of the forecast variance to particular variables in the system. Again, because of the contemporaneous correlation in the error terms, an orthogonalization technique must be used before the variance can be decomposed.

Working from the moving average representation and the orthogonalization in the previous section, the orthogonalized moving average representation becomes:

$$Z_t = \sum_{j=0}^{\infty} H_j S v_{t-j}.$$

where the H_j are the moving average parameters. The k-step ahead forecast error variance is

$$\text{var}(Z_{t+k} - E_t[Z_{t+k}]) = \sum_{k=j=0}^{k-1} (H_kS)(H_kS)'.$$

Let h_kS_{ij} be the ij^{th} element of H_kS. Then $(h_kS_{ij})^2$, j=1, ..., m is the i^{th} diagonal element of $(H_kS)(H_kS)'$.

The k-step ahead forecast variance of the i^{th} variable is then given by

$$\sum_{k=0}^{k-1} \sum_{j} (h_kS_{ij})^2$$

and the percentage of that variance from equation i accounted for by variable J is

$$100 \cdot \frac{\sum_{k=0}^{k-1} (h_kS_{iJ})^2}{\sum_{k=0}^{k-1} \sum_{j} (h_kS_{ij})^2}$$

There are two important points to remember in this discussion. First, this decomposed variance is highly dependent on the ordering of the variables prior to the decomposition. Second, the resulting

decomposition is of this orthogonalized variance.

The most complete source dealing with this material is Litterman [15]. Other sources that discuss this technique are: Orden [25,26], Hakkio and Morris [13], and Bessler [2].

This technique can also be used for the historic decomposition of variance; to investigate possible shocks to the economy from a historic viewpoint. This technique will not be discussed in this paper. Discussion of historic decomposition of variance can be found in Orden [27].

C. Impulse Responses and Variance Decomposition in RATS

The two primary commands used in RATS to generate impulse responses and the decomposition of variance are IMPULSE and ERRORS. Only the simplest options of these commands will be discussed here.

Impulse responses to exogenous shocks to a VAR system are generated with the IMPULSE command. This command would be placed in the RATS code some place after the system is estimated. Using the example presented in the previous section, the code to generate impulse responses would look like:

```
ESTIMATE 70,1 80,4 5
DECLARE SYMMETRIC V(4,4)
VCV(MATRIX=V) 70,1 80,4
#5 TO 8
IMPULSE 4 20 0 V
#1 0 0 1
#2 0 0 2
#3 0 0 3
#4 0 0 4.
```

The system is estimated with residuals being stored in data series beginning with series 5 and continuing through series 8 (since there are four equations). A 4x4 symmetric matrix, V, is declared into which the variance-covariance matrix of the residuals will be written. This is done by the statements DECLARE and then VCV. The VCV statement requires a supplementary card listing the series continuing the residuals.

IMPULSE indicates that innovations for four equations over twenty periods are to be generated. This example indicates by the zero value of the third parameter that responses to shocks to each equation will be computed in turn. If this parameter is non-zero it must give the equation number (from the order in the accompanying supplemental cards) to be shocked. The fourth parameter, V, is the array name from which the error covariance terms are supplied. If this parameter is omitted, impulses from non orthogonal innovations will be computed. The Choleski decomposition is the default in RATS, although options exist to enable the use of other decomposition methods.

Finally, the supplemental cards are listed, one for each equation, in the desired order of orthogonalization. Equations for variables with no expected predictive value for the other variables are generally placed last in the ordering. In this example X1 is placed first and X4 is placed last.

Impulse responses can also be computed with the IMPULSE option of ERRORS. ERRORS is used to decompose the forecast variance into its variable components. If both types of analysis are desired it is

more convenient to use just the ERRORS command with the IMPULSE option for the simplest forms of these analyses.

The ERRORS command is used in the same way as IMPULSE. It appears as:

```
ERRORS 4 20 V
#1 0 0 1
#2 0 0 2
#3 0 0 3
#4 0 0 4.
```

where forecast errors for four equations are decomposed for twenty steps given the covariance matrix V. Everything else is as it appears for IMPULSE.

VI. SUMMARY AND CLOSING REMARKS

This paper has presented a basic introduction to vector autoregression theory and methodology for those familiar with time series methodology. Its intent has been to present a cursory discussion of the theory underlying VAR's, identify important references in the literature using VAR techniques, and provide a complement to the RATS user's guide.

What is most striking about VAR methodology is the wide variety of situations to which it may be applied. Much like what occurred with the incorporation of risk into economic analysis, there exists the opportunity to take another look at a wide range of previously studied topics where conventional econometric techniques were applied.

VAR techniques have only recently been introduced to agricultural studies, especially those using Bayesian priors. VAR's could be used to study individual agricultural commodity markets or more aggregate sectoral models. There may be a need to use VAR's to generate more accurate forecasts from large scale econometric models of the agricultural sector. There are also policy applications of VAR methodology, although there is still some controversy over the use of VAR's in this arena (Sims [33], Sargent [30]. Hansen and Sargent [14]).

Finally, there are applied methodological issues concerning VAR's to be addressed. Some issues are: seasonality problems with VAR's, the use of VAR's with panel data, and the appropriate choice of Bayesian priors.

One can see, then, that vector autoregression techniques are a valued, and relatively new asset in an economist's "tool box." It is appropriate, however, to end with a note of caution. Users of vector autoregression must remain aware that an appropriate methodology must be used for the analysis at hand. Although vector autoregression is a powerful analytical tool, its use may not be advisable in many situations.

APPENDIX

List of variables and their definitions.

This list omits some of the obvious variables found in the paper. Most of these variables are defined when used, however, this notation is fairly standard throughout the paper.

Z_t	=	a wide sense stationary stochastic process
M	=	autoregressive order (number of lags)
N	=	number of system variables
T	=	number of observations over time
η	=	error term
D_j^M	=	coefficients in VAR
δ	=	standard deviations on prior means
ℓ	=	lag
λ	=	overall tightness of prior standard deviation
γ_1	=	decay parameter of prior standard deviation
γ_2	=	weight of variables in prior standard deviation
β_m	=	estimated parameter in Theil mixed estimation
R,r	=	arrays of parameter restrictions
Θ	=	a generic δ
σ	=	actual standard deviation of regression residuals
s	=	estimated standard deviation of regression residuals
Σ	=	covariance matrix
H	=	parameter estimates in moving average representation

REFERENCES

1. Bessler, D., "Relative Prices and Money: A Vector Autoregression on Brazilian Data," American Journal of Agricultural Economics, Vol. 66, No. 1, February 1984.

2. Bessler, D., "An Analysis of Dynamic Economic Relationships: An Application to the U.S. Hog Market," Canadian Journal of Agricultural Economics, Vol. 32, March 1984.

3. Bessler, D., and J. Kling, "Forecasting Vector Autoregressions with Bayesian Priors," American Journal of Agricultural Economics, Vol. 68, Number 1, February 1986.

4. Brandt, J., and D. Bessler, "Forecasting with Vector Autoregressions Versus a Univariate Arima Process: An Empirical Example with U.S. Hog Prices," North Central Journal of Agricultural Economics, Vol. 6, No. 2, July 1984.

5. Chambers, R., "Agricultural and Financial Market Interdependence in the Short Run," American Journal of Agricultural Economics, Vol. 66, No. 1, February 1984.

6. Chow, G., Econometrics, McGraw-Hill Book Company, New York, 1983.

7. Doan, T. and R. Litterman, User's Manual: Regression Analysis of Time Series, VAR Econometrics, 1984.

8. Doan, T., R. Litterman and C. Sims, "Forecasting and Conditional Projection Using Realistic Prior Distributions," Working Paper #243, Federal Reserve Bank of Minneapolis, 1983.

9. Fair, R., "An Analysis of the Accuracy of Four Macroeconometric Models," Journal of Political Economy, Vol. 87, No. 4, 1979.

10. Fisher, B., "Rational Expectations in Agricultural Economics Research and Policy Analysis," American Journal of Agricultural Economics, Vol. 64, No. 2, May 1982.

11. Ford, S., "An Application of Bayesian Vector Autoregression to the U.S. Turkey Industry," Staff Paper P86-29, Department of Agricultural and Applied Economics, University of Minnesota, September 1986.

12. Granger, C., and P. Newbold, Forecasting Economic Time Series, Academic Press, Inc., Orlando, 1977.

13. Hakkio, C. and C. Morris, "Vector Autoregressions: A User's Guide," Research Working Paper, RWP 84-10, Research Division, Federal Reserve Bank of Kansas City, November 1984.

14. Hansen, L. and T. Sargent, "Two Difficulties in Interpreting Vector Autoregressions," Working Paper #227, Research Department, Federal Reserve Bank of Minneapolis, September 1984.

15. Litterman, R., "Techniques of Forecasting Using Vector Autoregressions," Working Paper #115, Federal Reserve Bank of Minneapolis, 1979.

16. Litterman, R., "A Bayesian Procedure for Forecasting with Vector Autoregressions, " MIT Working Paper, 1980.

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

18. Litterman, R., "Forecasting and Policy Analysis with Bayesian Vector Autoregression Models," Quarterly Review, Federal Reserve Bank of Minneapolis, Fall 1984.

19. Litterman, R., "Forecasting with Bayesian Vector Autoregression - Five Years of Experience," Working Paper #274, Federal Reserve Bank of Minneapolis, 1985.

20. Litterman, R., "Forecasting with Bayesian Vector Autoregression - Five Years of Experience," Journal of Business and Economic Statistics, January 1986.

21. Litterman, R., "A Statistical Approach to Economic Forecasting," Journal of Business and Economic Statistics, January 1986.

22. Litterman, R and T. Supel, "Using Vector Autoregressions to Measure the Uncertainty in Minnesota's Revenue Forecasts," Quarterly Review, Federal Reserve Bank of Minneapolis, Spring 1983.

23. Nelson, C., Applied Time Series Analysis for Managerial Forecasting, Holden-Day, Inc., San Francisco, 1973.

24. Nerlove, M., D. Grether and J. Carvalho, Analysis of Economic Time Series, New York, Academic Press, 1979.

25. Orden, D., "Preliminary Empirical Evidence Concerning an Asset Theory Model of Markets for Storable Agricultural Commodities," Staff Paper P82-17, Department of Agricultural and Applied Economics, University of Minnesota, 1982.

26. Orden, D., "Capital Flows, the Exchange Rate, and Agricultural Commodity Markets," Ph.D. Thesis, Department of Agricultural and Applied Economics, University of Minnesota, November, 1984.

27. Orden, D., "The Macroeconomics of Trade Sectors:" The Case of U.S. Agriculture," Virginia Polytechnic Institute and State University, December 1985.

28. Sargent, T., "Estimating Vector Autoregressions Using Methods Not Based on Explicit Economic Theories," Quarterly Review, Federal Reserve Bank of Minneapolis, Summer 1979.

29. Sargent, T., Macroeconomic Theory, New York, Academic Press, 1979.

30. Sargent, T., "Autoregressions, Expectations, and Advice," American Economic Review, AEA Papers and Proceedings, Vol. 74, No. 2, May 1984.

31. Sims, C., "Macroeconomics and Reality," Econometrica, Vol 48, No. 1, January 1980.

32. Sims, C., "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," American Economic Review, Vol. 70, No. 2, May 1980.

33. Sims, C., "Policy Analysis with Econometric Models," Brookings Papers on Economic Activity, 1:1982.

34. Sims, C. "Are Forecasting Models Usable for Policy Analysis?," Quarterly Review, Federal Reserve Bank of Minneapolis, Winter 1986.

35. Theil, H., "On the use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, Vol. 58, June 1963.

36. Tiao, G. and G. Box, "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, Vol. 76, No. 376, Theory and Methods Section, December 1981.

37. Todd, R., "Improving Economic Forecasting with Bayesian Vector Autoregression," Quarterly Review, Federal Reserve Bank of Minneapolis, Fall 1984.

38. Webb, R., "Vector Autoregression as tool for Forecast Evaluation," Economic Review, Federal Reserve Bank of Richmond, January/February 1984.

39. Webb, R., "Toward More Accurate Macroeconomic Forecasts from Vector Autoregressions," Economic Review, Federal Reserve Bank of Richmond, July/August 1985.