

## **VAR Priors : Success or lack of a decent macroeconomic theory ?**

Francisco F. R. Ramos

Faculty of Economics, University of Porto, 4200 Porto, Portugal.

Phone: +351-(0)2-5509720, Fax: +351-(0)2-5505050

E-mail: [framos@fep.up.pt](mailto:framos@fep.up.pt)

### **Abstract:**

The purpose of this paper is to demonstrate that the success of the Litterman prior in VAR forecasting is not due to the realism of the prior, but rather because the prior conveniently reduces forecast error variance in common cases of misspecification. Specifically, it is shown that the imposition of a random walk prior reduces forecast error variance in misspecifications involving (1) time-varying coefficients misspecified as constant coefficients, (2) serially correlated residuals misspecified as white noise, and (3) the inclusion of an irrelevant unit root process in the VAR.

**Classification system for journal articles:** C32, C53.

**Key words:** BVAR, Forecasting performance, Litterman prior, Misspecification  
Random-walk prior, VAR.

## 1. Introduction

Perhaps the most successful application of vector autoregression methods (VAR) in macroeconomics has been in the area of economic forecasting. Since 1980, VAR forecasts produced by R. Litterman at the Federal Reserve Bank of Minneapolis have consistently outperformed forecasts generated from conventional Keynesian macroeconometrics models <sup>1</sup>. The forecasting superiority of VAR's is particularly striking at long forecast horizons, and persists even after forecasts from conventional models have benefited from judgmental adjustment by their model builders. The enviable forecasting record compiled by Litterman strongly reflects the imposition of Bayesian prior restrictions on the VAR. Although the Bayesian prior Litterman has developed for VAR forecasting is not derived from economic theory, users of this methodology claim that the Litterman prior is successful because it is a realistic way of expressing uncertainty about the problem of predicting macroeconomics time series.

This paper critically examines the proposition that Bayesian priors for VAR forecasting are successful because they are realistic. We contend that their lack of realism, combined with the lack of a decent macroeconomic theory is the reason for the success of the Litterman prior. Through a series of theoretical examples, we demonstrate that classical VAR forecasting leads to high forecast error variance but that the forecast error variance is reduced substantially by the imposition of the Litterman random walk prior. The second section of this paper briefly reviews the Litterman random walk prior. In the third section, we develop three examples, which we argue are typical of actual macroeconomics models, that show how the "unrealistic" random walk prior reduces forecast error variance. The paper closes with a brief summary and conclusion.

---

<sup>1</sup> Litterman's forecast record is described in Litterman (1986b) and McKnees (1986).

## 2. A quick review of Litterman random walk prior

In a series of papers, Litterman (1979, 1986a, 1986b) has demonstrated that the imposition of a Bayesian random walk prior on VAR's substantially improves forecast accuracy. Through a series of forecasting comparisons, he shows that Bayesian VAR (BVAR) models are competitive with large scale addfactored commercial models such as DRI, Wharton Econometrics, and Chase Econometrics. Given Fair's demonstration (1979) that unrestricted VAR's can produce forecast errors that are four times greater than the errors from conventional models, Litterman's forecast record is very impressive.

Following Litterman, a VAR model of order  $p$  consists of a system of equations relating  $n$  time series  $Y_{1t}, Y_{2t}, \dots, Y_{nt}$ . Each equation in a VAR model can be viewed as a multiple regression equation in which  $Y_{it}$  is regressed on  $p$  lags of each of the variables in the system. Thus, the  $i$ th equation of a BVAR model can be written as

$$\mathbf{Y}_i = \mathbf{X}\beta_i + \mathbf{u}_i \quad (1)$$

In (1),  $\mathbf{Y}_i$  is a vector of observations on the dependent variable;  $\mathbf{X}$  is the known data matrix, consisting of lagged values of  $Y_{1t}, Y_{2t}, \dots, Y_{nt}$ ;  $\beta_i$  is a vector of coefficient parameters; and  $\mathbf{u}_i$  is a vector of random disturbances. The error terms,  $\mathbf{u}_i$ , are assumed to be normally distributed, so that

$$\mathbf{u}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}) \quad (2)$$

Note that if the error terms are assumed to be contemporaneously, but not serially, correlated, then the system of  $n$  equations in the VAR can be viewed as a system of Seemingly Unrelated Regression (SUR) equations; see Zellner(1962).

The BVAR approach restricts a VAR model by incorporating available information about the coefficients of the model into the estimation procedure. The prior information takes the form of stochastic constraints on the coefficient parameters. For each coefficient parameter, the prior information is used to supply a point estimate of the mean of the parameter and an estimate of the parameter's variation about its mean. The prior information used in the specification of the BVAR model is based on two general premises, rather than explicitly on economic theory. This represents a departure from the practice associated with structural econometric models, in which model restrictions are derived explicitly on economic theory. In the first premise, Litterman (1986b) suggests that a realistic approximation of the movement over time of many macroeconomic variables is a random walk around an unknown deterministic component. He suggests that this prior should be imposed on all equations in the VAR. For the  $i$ th equation this distribution is centered around the specification

$$Y_i = d_{it} + Y_{i,t-1} + u_i \quad (3)$$

The second premise is that more information may be derived about a variable at time  $t$  from recent lags of the variables in the system than from more distant lags. An implication of the first premise is that prior point estimates for the coefficients are all taken to be zero, except for the first lag on the dependent variable in each equation in the system, which is given a prior mean of 1. Adherence to the second premise means that the variation about the prior means is assumed to diminish as the lag length increases, with the result that as the lag length increases, the prior distributions of the coefficients become tighter about a mean of 0. In addition, the prior has larger standard deviations on lag coefficients of the dependent variable than on other variables in the system. A non-informative prior is imposed on the deterministic component, since it is suggested that little is known about the distribution of this parameter.

An important factor in the design of BVAR models is that the prior estimates of the coefficient parameters variation are constructed in a relatively mechanistic manner, indexed by two or more "hyperparameters". This permits a fairly general specification of prior information regarding the coefficient parameters in the model, based on relatively few prior parameters. This approach greatly facilitates the use of a model which incorporates information, in a Bayesian fashion, about the coefficients. A fully Bayesian approach to the estimation of the VAR model would require the specification of a prior distributional form (e.g., mean and variance, assuming normal priors, or appropriate non-informative priors) for each parameter in the model. Note that a VAR model including six lags on seven variables would have 296 coefficient parameters, seven constant terms, and 28 elements in the covariance structure of the error terms, for a total of 331 free parameters. A BVAR specification for the same model can be based on the specification of as few as two hyperparameters. The specification of the prior information included in the BVAR model is admittedly less general than that which can be included in a fully Bayesian approach, but the BVAR model appears to offer a practical compromise, in terms of ease of both specification and estimation, between an unrestricted VAR and a fully Bayesian model. The literature on the elicitation of prior information in terms of distributional forms and the necessary parameters of these distributions for use in a Bayesian framework shows that the specification of parameters of prior distributions, even by those with considerable expertise and familiarity with the models under consideration, can be quite difficult. The BVAR approach to model specification makes these models considerably more attractive by circumventing much of the need to specify individual prior parameters.

With the imposition of this prior, external information is allowed to enter each equation at the margin, but it is important to recognize that the prior will never let the equations of the BVAR deviate too far from independent random walk with drift.

Although it is claimed that the prior is a realistic one, it should be noted that it is not derived from any explicit economic theory.

While Litterman has exploited the use of this prior in generating unconditional (nonstructural) macroeconomic forecasts, the prior has also been used to draw structural inference from macroeconomic time series. Doan, Litterman and Sims (1984), use this prior in making conditional forecasts, while Sims(1982, 1986) suggests that the prior is useful in assessing macroeconomic policies. To justify the use of this prior in examining structural equations, these authors argue that the Litterman prior is a realistic one for analyzing macroeconomic time series, and in particular, may reveal empirical regularities that remain hidden to standard procedures.

### **3. Theoretical examples**

In this section, we argue that the success of BVAR forecasting does not depend on the realism of the Litterman prior. Instead, we suggest that classical VAR forecasting under common situations of model misspecification leads to high forecast error variance, and that the judicious use of a random walk prior in this situations improves forecasting performance.

Specifically, we develop three theoretical examples which show how the "unrealistic" Litterman prior improves forecasting performance in misspecified VAR's. The three sources of misspecification we consider are (1) constant coefficient estimation when the true coefficients are time-varying, (2) estimation under the assumption of white noise errors when the true disturbances are serially correlated, and (3) inclusion of a spurious variable which contains a unit root. Given the lack of consensus in macroeconomic theory, these sources of misspecification are likely to be common in macroeconomic modeling. Moreover, they all can significantly affect the consistency and ( or) efficiency properties of OLS estimation in VAR's.

Lucas (1976), and Belsey and Kuhn (1973), have suggested that the usual fixed coefficient specification used in macroeconomic modeling will often be rejected in favor of time-varying parameter structures. They both point out that the benefits for improved macroeconomic forecasts resulting from adjusting equation intercepts, or "add-factoring", directly reflects coefficient variability. Rosenberg (1973), demonstrates that this type of misspecification has significant statistical implications. He shows that estimation of fixed coefficients when the true model contains stochastic coefficients results in inefficient estimation. Specifically, OLS error variance rises to five times the efficient variance, and OLS sampling theory underestimates OLS error variance by a factor of at least 20.

The problem of serially correlated disturbance terms is recognized to be extremely common in macroeconomic models. Most VAR studies assume white noise errors, since autoregressions of sufficient order should produce serially uncorrelated residuals. Models with moving average errors, however, theoretically may require an infinite number of autoregressive parameters to deliver white noise errors. Therefore, it is not clear that the lag lengths commonly employed in VAR's will completely eliminate serial correlation in the residuals<sup>2</sup>. Moreover, OLS estimates are biased and inconsistent in models with lagged dependent variables and serially correlated errors [Griliches, (1967)]. In addition, Griliches (1967) argues that it can be quite difficult to distinguish between distributed lag models, such as VAR's, and models with serially correlated errors. In the absence of prior identifying restrictions, these two cases may be observationally equivalent.

Inclusion of irrelevant variables in a VAR will likely be common since VAR's do not employ exclusion restrictions. The specific case considered in this paper is the

---

<sup>2</sup> Lag lengths of one year are commonly employed in empirical VAR studies Sims, e.g. (1980).

inclusion of an irrelevant variable with a unit root. This is an interesting process to consider since unit roots appear to be very common among aggregate time series [Nelson and Plosser (1982)], and nonstationary time series can lead to problems in estimation and hypothesis testing. In particular, standard estimation and testing procedures are probably not useful in examining linear restrictions in a VAR which are not co-integrated with the rest of the model [Sims, Stock and Watson, (1990)].

### 3.1 Time varying parameters.

In this example, we analyze the effect of misspecifying time varying coefficients as fixed coefficients on VAR forecast error.

Consider the following model:

$$X_t = b_t X_{t-1} + a Y_{t-1} + u_{1t} \quad (4.1)$$

$$b_t = b_{t-1} + e_t \quad (4.2)$$

$$Y_t = d Y_{t-1} + u_{2t} \quad (4.3)$$

and  $Y_t = \sum_{i=0}^{\infty} d^i u_{2,t-i}$ .

Estimating (4.1) by OLS yields:

$$X_t = \hat{b} X_{t-1} + \hat{a} Y_{t-1} + v_t \quad (4.4)$$

The k-step ahead forecast is:

$$X_{t+k}^* = \hat{b} X_{t+k-1}^* + \hat{a} Y_{t+k-1}^* \quad (4.5)$$

The k-step ahead forecast error is:

$$X_{t+k}^* - X_{t+k} = \hat{b}X_{t+k-1}^* - b_{t+k-1}X_{t+k-1} + \hat{a}Y_{t+k-1}^* - aY_{t+k-1} \quad (4.6)$$

Decomposing the forecast error, we first analyze  $b_{t+k-1}X_{t+k-1}$  :

$$b_{t+k-1}X_{t+k-1} = b_{t+k-1}b_{t+k-2}X_{t+k-2} + b_{t+k-1}aY_{t+k-2} \quad (4.7)$$

$b_{t+k-1}X_{t+k-1}$  can now be represented as:

$$X_t \prod_{j=1}^{k-1} b_{t+j} + \prod_{j=1}^{k-1} b_{t+j} (aY_{t+j-2} + u_{1,t+j-1}) \quad (4.7a)$$

Since  $Y_t$  is a covariance stationary stochastic process,

$$b_{t+k-1}X_{t+k-1} = X_t \prod_{j=1}^{k-1} b_{t+j} + a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \sum_{i=0}^{\infty} d^i u_{2,t+j-i} + \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} \quad (4.7b)$$

now, decomposing  $\hat{b}X_{t+k-1}^*$  :

$$\begin{aligned} \hat{b}X_{t+k-1}^* &= \hat{b}^2 X_{t+k-2}^* + \hat{b}\hat{a}Y_{t+k-2}^* \\ &= \hat{b}^{k-1} X_t + \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} Y_{t+j} \end{aligned} \quad (4.8)$$

Therefore, (4.6) becomes :

$$\begin{aligned}
X_{t+k}^* - X_{t+k} &= \hat{b}^{k-1} X_t + \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} Y_{t+j} + \hat{a} Y_{t+k-1}^* - a Y_{t+k-1} - X_t \prod_{j=1}^{k-1} b_{t+j} - \\
&\quad a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \sum_{i=0}^{\infty} d^i u_{2,t+j-i} - \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} \\
&= X_t \left( - \prod_{j=1}^{k-1} b_{t+j} + \hat{b}^{k-1} \right) - \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} - a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \sum_{i=0}^{\infty} d^i u_{2,t+j-i} + \\
&\quad \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} Y_{t+j} + \hat{a} Y_{t+k-1}^* - a Y_{t+k-1}
\end{aligned} \tag{4.9}$$

Recall that  $Y_{t+k} = \sum_{i=0}^{\infty} d^i u_{2,t+k-i}$ . The best forecast for  $Y_{t+k}^*$  with  $d$  known is

$$Y_{t+k} = d^k \sum_{i=0}^{\infty} d^i u_{2,t-i} \tag{4.10}$$

substituting yields:

$$\begin{aligned}
X_{t+k}^* - X_{t+k} &= X_t \left( - \prod_{j=1}^{k-1} b_{t+j} + \hat{b}^{k-1} \right) - \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} - \\
&\quad a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \sum_{i=1}^{k-1} d^i u_{2,t+j-i} + \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} d^j \cdot \sum_{i=0}^{\infty} d^i u_{2,t-i} + \\
&\quad \hat{a} d^{k-1} \sum_{i=0}^{\infty} d^i u_{2,t-i} - a d^{k-1} \sum_{i=0}^{\infty} d^i u_{2,t-i}
\end{aligned} \tag{4.11}$$

rearranging (4.11), we may now write the forecast error as:

$$\begin{aligned}
X_{t+k}^* - X_{t+k} &= X_t \left( - \prod_{j=1}^{k-1} b_{t+j} + \hat{b}^{k-1} \right) - \\
&\quad \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} - a \sum_{i=1}^{k-1} d^i u_{2,t-i} + \\
&\quad d^{k-1} \sum_{i=0}^{\infty} d^i u_{2,t-i} (\hat{a} - a) - a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \sum_{i=1}^{k-1} d^i u_{2,t+j-i} + \\
&\quad \sum_{i=0}^{\infty} d^i u_{2,t-i} \left( \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} d^j - a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \right)
\end{aligned} \tag{4.11a}$$

Now, decomposing the forecast error  $X_{t+k}^* - X_{t+k} =$

$$= \underbrace{\left[ - \sum_{j=1}^{k-1} \left( \prod_{j=1}^{k-1} b_{t+j} u_{1,t+j} + a \prod_{j=1}^{k-1} b_{t+j} \sum_{i=1}^{k-1} d^i u_{2,t+j-i} + a d^j u_{2,t-j} \right) \right]}_{\text{forecast error due to unknown shocks}}$$

(4.12a)

$$+ \underbrace{d^{k-1} (\hat{a} - a) \sum_{i=0}^{\infty} d^i u_{2,t-i}}_{\text{forecast error due to estimation of } a} \tag{4.12b}$$

$$+ \underbrace{\left( \hat{b}^{k-1} - \prod_{j=1}^{k-1} b_{t+j} \right) X_t + \sum_{i=0}^{\infty} d^i u_{2,t-i} \left( \hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} d^j - a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j} \right)}_{\text{forecast error due to estimation of } b_t \text{ and model misspecification}}$$

(4.12c)

Analyzing the first component of (4.12c) :

$$\left( \hat{b}^{k-1} - \prod_{j=1}^{k-1} b_{t+j} \right) X_t = \hat{b}^{k-1} \left( 1 - \prod_{j=1}^{k-1} b_{t+j} \cdot \frac{1}{\hat{b}^{k-1}} \right) X_t$$

This term explodes in  $k$  unless  $\hat{b}$  goes to zero and  $\frac{b_{t+j}}{\hat{b}^{k-1}}$  goes to  $k$  less than infinity

or  $\prod_{j=1}^{k-1} \frac{b_{t+j}}{\hat{b}^{k-1}}$  goes to 1 faster than  $\hat{b}^{k-1}$  goes to infinity. Constraining  $\hat{b}$  to 1 (random

walk) when it is greater than 1 reduces this term and unequivocally reduces mean square forecast error. If  $\hat{b}$  is less than 1, constraining it to be 1 increases the mean forecast error, but this is less serious because the error is nonexplosive. The second term of (4.12c),  $\hat{a} \sum_{j=1}^{k-1} \hat{b}^{k-j} d^j$  is reduced by setting  $\hat{a}$  to zero. The third term of (4.12c)  $-a \sum_{j=1}^{k-1} \prod_{j=1}^{k-1} b_{t+j}$  is explosive, but no model parameters will reduce it.

### 3.2 Serially correlated residuals.

In this example, we analyze the effect of serial correlation misspecification on VAR forecast error. In particular, we examine the case in which substantial serial correlation in the error term is mistaken for serial correlation in the  $Y_t$ 's.

Consider the following autoregression with an AR(1) disturbance term:

$$Y_t = aY_{t-1} + bX_{t-1} + u_t \quad (5.1)$$

$$u_t = cu_{t-1} + v_t \quad (5.2)$$

$$X_t = dX_{t-1} + e_t \quad (5.3)$$

$$E(v_t \cdot e_t) = 0$$

but instead, white noise errors are assumed:

$$Y_t = aY_{t-1} + bX_{t-1} + w_t \quad (5.4)$$

While (4.1) can be written as an autoregression with a white noise error term, it will require the following second order terms and nonlinear restrictions:

$$Y_t = (a + c)Y_{t-1} - a \cdot cY_{t-2} + b(X_{t-1} - cX_{t-2}) + v_t \quad (5.1a)$$

A simple comparison between (5.1a) and (5.4) clearly indicates that ignoring serially correlated residuals significantly changes the structure of the VAR. In particular, note that the assumption of white noise errors in (5.1a) requires a coefficient estimate of  $(a + c)$ , rather than  $a$ , for optimal prediction of future  $Y$ . As Griliches (1967) has pointed out, OLS estimation of (5.4) yields a biased and inconsistent estimate of  $a$ . The large sample bias in  $\hat{a}$  has been calculated by Griliches (1967) as:

$$p \lim(\hat{a} - a) = \frac{c(1 - a^2)}{1 + ac} \cdot \frac{1}{1 + \frac{b^2 \text{Var}(X)(1 - d^2)(1 - ac)}{\text{Var}(u)(1 - ad)(1 + ac)}} \quad (5.5)$$

As long as  $c > 0$ , and  $a$  is less than one in absolute value,  $\hat{a}$  will overestimate  $a$ . This seems to be a fortunate result from a forecasting viewpoint, since we wish to estimate  $(a + c)$  rather than  $a$ . To determine if the OLS estimate  $\hat{a}$  approximates  $(a + c)$ , we calculate  $\hat{a}$  using the following parameter values:

$$a = 0.15 \quad b = 0.5 \quad c = 0.8 \quad d = 0.9 \quad \text{Var}(X_t) = \text{Var}(u_t)$$

Since our interest in this example is to determine the effect of substantial serial correlation misspecification on forecast error, we choose  $a$  and  $c$  to be positive, with  $a$  small relative to  $c$ . Note, however, that  $(a + c) = 0.95$  is stationary, but close to the unit root specification Nelson and Plosser (1982) have found to be common in aggregate time series. Similarly, we choose  $X_t$  also to be stationary but with a root close to the unit circle. Thus,  $X_t$  and  $Y_t$  will reasonably approximate many macroeconomic time series. The assumption that  $\text{Var}(X_t) = \text{Var}(u_t)$  is arbitrary; the larger the variance in  $X_t$  relative to  $u_t$ , the smaller the large sample bias in  $\hat{a}$ .

Calculating  $\hat{a}$  under these parameter values yields  $\hat{a} = 0.81$ . Thus, the OLS estimate is biased substantially upwards from the true value of 0.15, but it is significantly below the optimal value for prediction of  $Y_{t+k}$  ( $a + c = 0.95$ ).

Consider the imposition of the Litterman prior on (5.4). First, note setting  $\delta$  approximately equal one is much closer to the optimal value of 0.95 for the coefficient on the first lag of the dependent variable than the OLS estimate of  $\hat{a} = 0.81$ . Second, note that pushing other terms to zero does not seriously increase forecast error. To see this, write (5.1a) as:

$$Y_t = (a + c)Y_{t-1} - acY_{t-2} + b(dX_{t-2} - cX_{t-2}) + be_{t-1} + v_t \quad (5.1b)$$

For the parameter values we choose, note that  $ac$  is close to zero (0.12), and the term  $b(dX_{t-2} - cX_{t-2})$  is also close to zero (0.05), so a prior that sets these terms approximately to zero is not damaging. Other information ignored under the Litterman prior is  $be_{t-1}$ , the  $X_{t-1}$  innovation. In contrast, note that OLS estimation of (5.4) yields not only a downward biased estimate of  $(a + c)$ , but will incorrectly estimate the influence of the  $X_t$ 's on  $Y_t$  since it ignores the second order term from (5.1a),  $-bcX_{t-2}$ . The BVAR model will outperform the VAR in this case, since the random walk prior is closer to the true value than the OLS estimate  $a$ , and since OLS estimation of the VAR ignores more information about the  $X_t$ 's than the BVAR,  $-bcX_{t-2}$  vs.  $-be_{t-1}$ .

This example illustrates that the imposition of the Litterman prior in a VAR with serial correlation misspecification improves forecast accuracy relative to OLS estimation. Specifically, the random walk prior reduces coefficient bias from misspecification.

While the specific results of this example will not necessarily generalize, we can make the following observations. First, in VAR's with highly serially correlated residuals, the OLS estimate of the coefficient on the lagged dependent variable will be biased and inconsistent, and may not be optimal from a forecasting viewpoint. In this case, the imposition of a unit coefficient on the lagged dependent variable will correspond more closely to the coefficient for optimal prediction than the OLS estimate. Second, solving for serial correlation in VAR's, as was done in (5.1a), results in higher order autoregressions with nonlinear restrictions. For VAR's with more general orders of serial correlation, reducing the disturbance terms to white noise may significantly complicate the structure of the model. In these cases, a prior of pushing other coefficients to zero may be less harmful than attempting to estimate the additional parameters associated with higher order models and nonlinearities.

### 3.3 Inclusion of an irrelevant unit root.

In this example, we analyze the effect of including an irrelevant unit root process on VAR forecast error. We choose a unit root variable since it is well known that nonstationary time series can lead to problems in estimation and inference, and since unit root specifications appear to be good approximations for many macroeconomics time series (Nelson and Plosser, 1982).

Consider the following (true) model:

$$Y_t = aY_{t-1} + e_t \quad (6.1)$$

Instead, the estimated model includes an irrelevant random walk,  $Z_t$  :

$$Y_t = \hat{a}Y_{t-1} + \hat{b}Z_{t-1} + u_t \quad (6.1a)$$

$$Z_t = Z_{t-1} + e_t \quad (6.2)$$

For stationary  $Z_t$ , standard testing procedures are fine for testing whether  $\hat{b} = 0$  in (6.1a). However, testing  $\hat{b} = 0$  is more difficult when  $Z_t$  contains a unit root. While Sims, Stock, and Watson (1990) demonstrate that standard estimation and testing procedures are valid in many situations with nonstationary, but co-integrated regressors, tests for Granger-causality and neutrality (testing for  $\hat{b} = 0$  in (6.1a)) will have nonstandard limiting distributions if the model is not co-integrated, as in (6.1a). In these cases, classical tests will fail to reject  $\hat{b} = 0$  too often.

Given that we may include  $Z_{t-1}$  with a non-zero coefficient in (6.1a), the forecast error for  $Y_{t+k}$  grows without bound because of the presence of the unit root. The k-step ahead forecast  $Y_{t+k}^*$  conditional on information at  $(t-1)$  from (6.1a) is given by:

$$Y_{t+k}^* = \hat{a}Y_{t+k-1}^* + \hat{b}Z_{t+k-1}^*$$

(6.3)

We also have:

$$Y_{t+k} = \hat{a}Y_{t+k-1} + \hat{b}Z_{t+k-1} + u_{t+k-1}$$

(6.4)

Subtracting the forecast of  $Y_{t+k}$  ( $Y_{t+k}^*$ ) (6.3) from (6.4), and focusing on the spurious random walk  $Z_t$ , we obtain:

$$\hat{b}Z_{t+k-1} - bZ_{t+k-1}^* \quad (6.5)$$

Since  $Z_{t+k-1}^*$  is conditional on information at  $(t-1)$ , the forecast error (6.5) is equal to

$$\hat{b} \sum_{i=1}^{k-1} e_{t+i} \quad (6.6)$$

The variance of (6.6) is equal to  $\hat{b}^2(k-1)Var(e_t)$ . This term is unbounded in  $k$ , and results in forecast error variance of  $Y_{t+k}$  growing without bound. The imposition of the Litterman prior, however, will reduce forecast error variance, because it pushes the coefficient  $b$  to zero.

#### 4. Summary and Conclusions.

The purpose of this paper was to demonstrate that the success of the Litterman prior in VAR forecasting is not due to the realism of the prior, but rather because the prior conveniently reduces forecast error variance in common cases of misspecification. Specifically, it was shown that the imposition of a random walk prior reduces forecast error variance in misspecifications involving (1) time-varying coefficients misspecified as constant coefficients, (2) serially correlated residuals misspecified as white noise, and (3) the inclusion of an irrelevant unit root process in the VAR.

Given the lack of consensus in macroeconomic theory and the minimal structure imposed in VAR's, it seems to us that these and other types of misspecification will often be present in empirical macroeconomic models. Therefore, this analysis suggests that the Litterman prior is a useful way of restricting VAR's for economic forecasting. The substantial value added from the prior in generating forecasts, however, should not be mistaken as a signal of the prior's realism. Rather, it should indicate to the forecaster that specification uncertainty is a fundamental component of forecast error in

macroeconomic models, and that a random walk prior is a useful approach in these type of situations.

## REFERENCES

- Belsey, D.A. and E. Kuh, 1973, Time-varying parameters structures: An overview ,  
Annals of Economic and Social Measurement 2, 375-380.
- Doan, T., R. Litterman and C.Sims, 1984, Forecasting and conditional projection using  
realistic prior distributions, Econometric Reviews 3, 1-100.
- Fair, R., C., 1979, An analysis of the accuracy of four macroeconomics models, Journal  
of Political Economy 87, 701-718.
- Griliches, Z., 1967, Distributed lags: A survey, Econometrica 35, 16-49.
- Litterman, R., 1979, Techniques for forecasting using vector autoregressions, Ph.D.  
Thesis, University of Minnesota.
- Litterman, R., 1986a, Specifying vector autoregressions for macroeconomic forecasting,  
in: P. Goel and A. Zellner, eds., Bayesian Inference and Decision Techniques  
with applications: Essays in Honor of Bruno De Finetti (North-Holland,  
Amsterdam) 79-84.
- Litterman, R., 1986b, Forecasting with bayesian vector autoregressions: Five years of  
experience, Journal of Business and Economic Statistics 4, 25-37.
- Lucas, R. E., 1976, Econometric Policy Evaluation: A Critique, in: K. Brunner and A.  
Meltzer, eds., The Phillips Curve and Labor Markets, 19-46.
- McKnees, S., 1986, Forecasting accuracy of alternative techniques: A comparison of  
U.S. macroeconomic forecasts, Journal of Business and Economic Statistics 4,  
5-16.
- Nelson, C. R and C.I. Plosser, 1982, Trends and random walks in macroeconomic time  
series: Some evidence and implications, Journal of Monetary Economics 10,  
139-162.
- Rosenberg, B., 1973, A survey of stochastic parameter regression, Annals of Economic  
and Social Measurement 2, 381-397.
- Sims, C., 1980, Macroeconomics and reality, Econometrica 41, 775-788.

- Sims, C., 1982, Policy Analysis with Econometric Models, Brookings Papers on Economic Activity, 107-152.
- Sims, C., 1986, Are forecasting models useable for policy analysis ? , Federal Reserve Bank of Minneapolis Quarterly Review, Winter, 1-10.
- Sims, C., J. H. Stock, and M. W. Watson, 1990, Inference in linear time series with some unit roots, *Econometrica* 58, 113-144.
- Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348-368.