Abstract—In the existing literature, conditional forecasts in the vector autoregressive (VAR) framework have not been commonly presented with probability distributions. This paper develops Bayesian methods for computing the exact finite-sample distribution of conditional forecasts. It broadens the class of conditional forecasts to which the methods can be applied. The methods work for both structural and reduced-form VAR models and, in contrast to common practices, account for parameter uncertainty in finite samples. Empirical examples under both a flat prior and a reference prior are provided to show the use of these methods.

I. Introduction

In a dynamic multivariate system such as a vector autoregression (VAR) model, out-of-sample forecasts are often made with no conditions imposed on the future values of endogenous variables. In the existing literature, there are a number of classical results on the asymptotic distribution of unconditional forecasts under the assumption of stationarity (e.g., Goldberger et al., 1961; Schmidt, 1974, 1977; West, 1996). In recent papers, Sims and Zha (1998, 1999) showed how Bayesian methods can be used to simulate the exact finite-sample distribution of unconditional forecasts. Their methods apply to nonstationary models as well.

Finite-sample inference on forecasts conditional on the future values of endogenous variables, however, has remained a challenging problem. In empirical policy analysis, conditional forecasts of this sort are often used to answer questions like “How do the forecasts of other macroeconomic variables change if the federal funds rate follows a different path?” Since movements in the federal funds rate are mostly due to the endogenous responses of the monetary authority to the changing state of the economy, the funds rate should be treated as an endogenous variable within a system of equations (Leeper et al., 1996). But endogeneity makes the existing methods of Sims and Zha (1998, 1999) inapplicable to finite-sample inferences on conditional forecasts.

In this paper, we develop Bayesian methods for computing the exact finite-sample distribution of forecasts conditional on the future values of endogenous variables in the VAR framework. The methods work for both structural and reduced-form VARs and do not depend on the assumption of stationarity. One method deals with conditions that fix the future values of variables at single points. For example, the future funds rate is restricted to 5% in the next year. Such conditions have been often considered in the forecasting literature and are called hard conditions in this paper. The other method deals with conditions that restrict the future values within only a certain range (for example, a target range for the M2 growth rate or a restriction that future inflation is below 3%). These types of conditions are referred to as soft conditions in this paper.

Both methods take explicit account of two sources of uncertainty: uncertainty about the “true” parameters and uncertainty originating from exogenous random shocks in the system. Exact finite-sample inferences on the model’s parameters are computed based on the shape of the likelihood or posterior density. We show that ignoring parameter uncertainty in finite samples can result in potentially misleading conditional forecasts. This finding is complementary to the existing evidence on the importance of taking account of parameter uncertainty in unconditional forecasts (Schmidt, 1977; West, 1996; Sims & Zha, 1998, 1999).

When conditions are imposed on the future values of an endogenous variable such as the federal funds rate, the variable itself should continue to be treated as endogenous over the forecast period. One might adopt an easy approach to replace, say, the estimated federal funds rate equation with an equation that specifies the funds rate as an exogenous deterministic process over the forecast period. This way, all the existing methods in both the classical asymptotic and Bayesian finite-sample literatures can be applied to the modified system, for the variable conditioned on (in this case, the federal funds rate) is now treated as exogenous.

This approach, however, is not advocated for two reasons. First, there is no rationale for believing that the Federal Reserve, at each and every forecast date, will decide to stop responding to the state of the economy and begin to control the funds rate in an exogenous fashion. Because most of the variation in the federal funds rate as a policy instrument arises in response to movements of other macroeconomic variables (such as output and inflation), the forecasts conditional on an exogenous process of the funds rate are conceptually problematic. Second, even if one is willing to treat the federal funds rate as exogenous over the forecast period, the values of the parameters in other equations of the original system will be different, in general, from those estimated under the assumption that the funds rate has been endogenous up to the forecast date. This point is essentially the rational-expectations critique on such econometric exercises.

Received for publication September 21, 1998. Revision accepted for publication June 8, 1999.

* Federal Reserve Bank of Atlanta.

Detailed comments from three referees and the editors (James Stock and Ken West) have led to significant improvement over earlier drafts. We also thank Frank Diebold, Bob Eisenbeis, John Geweke, Lutz Kilian, Eric Leeper, Adrian Pagan, Will Robers, Matt Shapiro, Ellis Tallman, and especially John Robertson, Chris Sims, and Chuck Whiteman for valuable comments on earlier drafts. Bryan Acree and Jeff Johnson provided able research assistance. The views expressed herein are not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

The remainder of this paper is organized as follows. Section II lays out a general framework. Section III develops the theoretical foundation of our Bayesian methods for...
computing the probability distribution of conditional forecasts. Section IV provides empirical examples that show how these methods can be used to compute conditional forecasts and their probability distributions. Section V concludes the paper.

II. Conditional Forecasts

A. General Framework

The dynamic multivariate framework considered in this paper has the form:  

\[ \sum_{l=0}^{p} y_{t-l} A_l = d + \varepsilon_t, \quad \text{for } t = 1, \ldots, T, \]  

(1)

where \( T \) is the sample size, \( y_t \) is a vector of observations, \( A_l \) is the coefficient matrix of the \( l \)-th lag, \( p \) is the maximum lag length, \( d \) is a vector of constant terms, and \( \varepsilon_t \) is a vector of i.i.d. structural shocks that are Gaussian with  

\[ E(\varepsilon_t | y_{t-s}, s > 0) = I_{m \times m}, \quad E(\varepsilon_t | y_{t-s}, s > 0) = 0_{m \times m}, \quad \text{for all } t. \]  

(2)

Note that columns in \( A_t \) correspond to equations. This paper considers only linear restrictions on the contemporaneous coefficient matrix \( A_0 \), which is assumed to be nonsingular.

When model (1) is used for out-of-sample forecasting, it must be transformed to the reduced form:  

\[ y_t = c + \sum_{l=1}^{p} y_{t-l} B_l + \varepsilon_t A_0^{-1}, \quad \text{for all } t. \]  

(3)

The relationships between the reduced-form parameters and structural parameters are  

\[ c = dA_0^{-1} \quad \text{and} \quad B_l = -A_l A_0^{-1}, \quad \text{for } l = 1, \ldots, p. \]  

(4)

Given equation (3) and (4) and the data up to time \( T \), the \( n \)-step forecast at time \( T \) can be written as  

\[ y_{T+n} = cK_{n-1} + \sum_{l=1}^{p} y_{T+1-l} N_l(n) \]  

\[ + \sum_{j=1}^{n} \varepsilon_{T+j} M_{n-j} \quad n = 1, 2, \ldots \]  

(5)

where  

\[ K_0 = I, \quad K_i = I + \sum_{j=1}^{i} K_{i-j} B_j, \quad i = 1, 2, \ldots ; \]  

\[ N_l(1) = B_l, \quad l = 1, \ldots, p; \]  

\[ N_l(n) = \sum_{j=1}^{n-1} N(n-j) B_j + B_{n+l-1}, \quad l = 1, \ldots, p, \quad n = 2, 3, \ldots ; \]  

\[ M_0 = A_0^{-1}, \quad M_i = \sum_{j=1}^{i} M_{i-j} B_j \quad i = 1, 2, \ldots ; \]  

with the convention that \( B_j = 0 \) for \( j > p \).

Equation (5) is composed of two parts. The first part, consisting of the first two terms in (5), gives dynamic forecasts in the absence of shocks; the second part—the third term in (5)—is the dynamic impact of various structural shocks. These shocks affect the future realizations of variables through the impulse-response matrices \( M_i \). The definition of conditional forecast in this paper is restricted to conditions imposed on the values of endogenous variables \( y_{T+n} \). Traditionally, conditions concern the future values of only exogenous variables (Intriligator et al., 1996, pp. 518–532). In this case, the method of Sims and Zha (1998) is readily applicable to statistical inferences on point forecasts. When conditions are imposed on the future values of endogenous variables, however, it becomes both conceptually and numerically difficult to obtain the finite-sample probability distribution of conditional forecasts.

B. Conditional Forecasts

To make an efficient use of some notations, denote  

\[ a_0 = \text{vec}(A_0), \quad a_+ = \text{vec}(-A_1 \ldots -A_p)^T, \quad \text{and} \quad a = \begin{bmatrix} a_0 \\ a_+ \\ d \end{bmatrix}. \]

Consider a condition that restricts the value of the \( j \)-th endogenous variable in the model at time \( T + n \), denoted by \( y_{T+n}(j) \), in the range \((\widetilde{y}_{T+n}(j), \gamma_{T+n}(j))\). Equation (5) implies that this condition can be expressed in the following form:  

\[ \sum_{j=1}^{n} \varepsilon_{T+j} M_{n-j}(\cdot, j) \in (\gamma_{T+n}(j) - Z_{n,j}(a), \gamma_{T+n}(j) - Z_{n,j}(a)), \]  

(6)

\[ \text{where} \quad Z_{n,j}(a) = \sum_{l=1}^{n} N(n-l) B_j + B_{n+l-1}. \]
where

\[ Z_{n,j}(a) = cK_{n-1}J + \sum_{l=1}^{p} y_{T+1-l} N_l(n) \]

and the notation \((\cdot, j)\) denotes the \(j\)-th column of a matrix. A compact form of equation (6) that allows for multiple constraints across \(n\) or \(j\) or both is given by

\[ R(a) = B(a) \subseteq \mathbb{R}^q, \quad q \leq k = hm, \tag{7} \]

where \(h\) is the maximum forecast horizon, \(q\) is the total number of conditions, \(k = hm\) is the total number of future shocks, \(R(a)\) is a stacked matrix from the impulse responses \(M_{n-1}(\cdot, j)\), \(e\) is a vector corresponding to \(B_{T+i}\), and \(B(a)\) is the restricted set of outcomes corresponding to the right-hand term in (6).

5 The types of conditions implied by equation (7) are broader than those implied by (6) because conditions imposed directly on structural shocks can also be put in the form of (7). See Leeper and Zha (1999) for applications of macroeconomic forecasts conditional on monetary policy shocks. The methods described in this paper work for restrictions of both the form (6) and (7).

orthogonal transformation of the other. Proposition (1) therefore provides a theoretical rationale for the convention that \(A_0\) is parameterized to be triangular (Doan et al., 1984; Doan, 1992). When the distribution of the parameters is taken into account, two arbitrary identification schemes are generally not orthonormally transformable. But, for the class of exactly identified models where \(A_0\) is triangular, the distribution of conditional forecasts does not depend on how \(A_0\) is triangularized. For example, a lower triangular form of \(A_0\) can be transformed to an upper triangular form of \(A_0\) through two operations. One operation uses a sequence of orthonormal transformations discussed in proposition (1); the other interchanges the order of variables accordingly. Neither operation affects the distribution of conditional forecasts.

III. Simulation Methods for Probability Distributions

Probability distributions of conditional forecasts have not been commonly presented in the existing VAR literature (e.g., Sims, 1982; Doan et al., 1984; Miller & Roberds, 1991; Roberds & Whiteman, 1992). Since all forecasts contain errors, some of which are substantial, it is important to provide the probability distributions underlying these errors.

The forecast errors of \(y_{T+n}\) emanate from two sources of uncertainty. One source pertains to future shocks \(\varepsilon_{T+i}\) for \(i = 1, \ldots, n\), which are assumed to have a Gaussian distribution. The other source of uncertainty relates to the shape of the likelihood (posterior density) of the parameters \(a_0\) and \(a_+\). In the Bayesian framework, the exact posterior distribution of the model’s parameters can be easily obtained. Under the flat prior or the informative prior of Sims and Zha (1998), the posterior distribution of \(a\) has the form

\[ p(a; Y_T) = \pi(a_0)\pi(a_+|a_0), \tag{8} \]

where \(Y_T\) denotes the data matrix up to time \(T\),

\[ \pi(a_0) \propto |A_0|^T \exp \left( -\frac{1}{2} \text{trace}(A_0'SA_0) \right), \]

\[ \pi(a_+|a_0) = \varphi((I \otimes U)a_0; I \otimes V), \tag{9} \]

and \(\varphi(\mu; \Sigma)\) denotes the normal density function with mean \(\mu\) and variance \(\Sigma\). In equation (9), \(S, U,\) and \(V\) are matrix functions of the data \(Y_T\) (and the prior mean and variance when the informative prior of Sims and Zha (1998) is used). Depending on the type of linear restrictions imposed on \(A_0\), there are a number of Monte Carlo (MC) methods available for generating random draws of \(a\) from the posterior distribution (8) (Waggoner & Zha, 1999; Zha, 1999). These methods provide a first step towards obtaining the distribu-

6 See Kilian (1998a, 1998b) and Sims and Zha (1999) for detailed discussions on the difficulties associated with various classical approaches, especially for nonstationary VAR models.
tion of forecasts under both hard and soft conditions implied by equation (7).

A. Hard Conditions: A Gibbs Sampling Technique

The conditions discussed in the existing literature usually concern situations in which the value of $y_{T+n}(j)$ is restricted to a single value. Constraints (6) and (7) imply that $B(a)$ collapses to a $q \times 1$ vector of values. Denote the $q \times 1$ vector by $r(a)$. The set of conditions in constraint (7) can now be equivalently expressed as

$$R(a)' \epsilon = r(a), \quad q \leq k = mh.$$  

(10)

The set of conditions in equation (10) are called hard conditions. In order to derive a method for simulating the distribution of forecasts under hard conditions, we first establish the following proposition.

Proposition (2). Given the constraints in equation (10) and the value of the parameter vector $a$, the joint distribution of $y_{T+1}, \ldots, y_{T+h}$ is Gaussian with

$$p(y_{T+n}|a, Y_{T+n-1}) = \varphi(c + \sum_{i=1}^{p} y_{T+n-i} B_i + M (\epsilon_{T+n}) A_0^{-1}; A_0^{-1} V (\epsilon_{T+n}) X A_0^{-1})$$

(11)

where $Y_{T+n-1}$ is the data matrix up to time $T + n - 1$, $M (\epsilon_{T+n})$ and $V (\epsilon_{T+n})$ are the mean and variance of $\epsilon_{T+n}$, whose distribution is normal with the following form:

$$p(\epsilon|a, R(a)' \epsilon = r(a)) = \varphi(R(a)(R(a)' R(a))^{-1} r(a);$$

$$I - R(a)(R(a)' R(a))^{-1} R(a)' \).$$  

(12)

Proof. By assumption (2), the unconditional distribution of $\epsilon$ is normal with density $\varphi(0; I_{n \times n})$. Hence, constraint (10) implies that the conditional distribution of $\epsilon$ is given by equation (12). The marginal distribution of $\epsilon_{T+n}$ is also normal, and its mean and variance can be read off directly from (12). Given $a$, the conditional distribution (11) follows directly from (3). Q.E.D.

Proposition (2) provides the analytical form of the density function of conditional forecasts when the values of the parameters are taken as given; it is straightforward to construct draws from this distribution. The procedure used in previous work (Doan et al., 1984; Doan, 1992) derives a point-estimate forecast by minimizing $\epsilon' \epsilon$ subject to constraint (10). It can be easily seen that the solution to this optimization problem equals the conditional mean of $\epsilon$ in equation (12). So such a procedure ignores the uncertainty associated with future shocks.

If one wishes to take account of parameter uncertainty, simulations from the distribution of conditional forecasts become challenging. If one were to draw $a$ from the posterior distribution conditional on $Y_T$ and then condition on these draws to generate $y_{T+n}$ according to proposition (2), the resulting distribution of $y_{T+n}$ would be incorrect because these draws of $a$ ignore the set of conditions in equation (10). The correct marginal distribution of $a$ conditional on (10) must derive from the joint distribution of $a$ and $y_{T+n}$. The analytical form of this joint distribution is in general unknown because $R$ and $r$ are nonlinear functions of $a$ in (10). But this distribution can be simulated. In the following algorithm, we develop a Gibbs sampler technique for simulations.

Algorithm (1). Initialize an arbitrary value $a^{(0)}$ (e.g., the value at the peak of $p(a|Y_T)$) or a value randomly drawn from $p(a|Y_T))$. For $i = 1, 2, \ldots, N_1 + N_2$,

(a) generate $y_{T+1}^{(i)}, \ldots, y_{T+h}^{(i)}$ from $p(y_{T+1}, \ldots, y_{T+h}|a^{(i-1)}, Y_T)$ by proposition (2) (i.e., draw $\epsilon$ from (12) and then use (3) to obtain $y_{T+1}^{(i)}, \ldots, y_{T+h}^{(i)}$);
(b) generate $a^{(i)}$ from $p(a|y_{T+1}^{(i)}, \ldots, y_{T+h}^{(i)}, Y_T)$;
(c) repeat (a) and (b) until the sequence $a^{(1)}, y_{T+1}^{(1)}, \ldots, y_{T+h}^{(1)}, \ldots, a^{(N_1+N_2)}, y_{T+1}^{(N_1+N_2)}, \ldots, y_{T+h}^{(N_1+N_2)}$ is simulated;
(d) keep the last $N_2$ draws in the sequence. (In practice, $N_2$ is set to equal $N_1$.)

In step (a) of algorithm (1), because the forecast of $y_{T+n}$ is generated from equation (11) in proposition (2), it always satisfies constraint (10). In step (b), the density function $p(a|y_{T+1}^{(i)}, \ldots, y_{T+h}^{(i)}, Y_T)$ is the posterior density function conditional on the data extended to include $h$ additional simulated observations $y_{T+n}^{(i)}$ for $n = 1, 2, \ldots, h$. When $A_0$ is exactly identified, one can draw $a_0$ directly from $\pi(a_0)$ in equation (9) because $A_0^{-1} A_0^{-1}$ has a Wishart distribution (Sims & Zha, 1999; Zha, 1999). When $A_0$ is over-identified, one cannot simulate $\pi(a_0)$ directly but can use the MC method set forth by Waggoner and Zha (1999). Conditional on each draw of $a_0$, one can draw $a_{i+1}$ directly from the normal distribution specified in equation (9).

Step (b) of algorithm (1) is a crucial step for obtaining the correct finite-sample variation in parameters subject to a set of hard conditions in constraints (10). Because the distribution of parameters is simulated from the posterior density function, the prior plays an important role in determining the location of the parameters in finite samples. Under the flat prior, the posterior density is simply proportional to the likelihood function, which, in a typical VAR system, is often

---

7 Doan et al. (1984) arrives at this result under the assumption that model (3) is stationary. The stationarity assumption, however, is not required for proposition (2).

8 See Geweke (1996, 1999) for details of the Gibbs sampler and other Bayesian techniques.
flat around the peak in small samples. Moreover, maximum-likelihood estimates tend to attribute a large amount of variation to deterministic components (Sims & Zha, 1998). Such a bias, prevalent in dynamic multivariate models like VARs, is the other side of the well-known bias toward stationarity of least-squares estimates. These problems can have substantial effects on the distribution of conditional forecasts, as will be shown in section IV.

The informative prior of Sims and Zha (1998) (the SZ prior hereafter) is designed to eliminate erratic sampling errors in estimation by downweighting the influence of distant lags and the unreasonable degree of explosiveness in a system of multiple equations. As has been shown, the prior significantly improves out-of-sample forecasts relative to a system of multiple equations. As has been shown, the prior substantially reduces the degree of flatness in the shape of the likelihood. As will be shown in section IV, such a reduction helps remedy the shifty effects of parameter uncertainty on the distribution of conditional forecasts under the flat prior.

B. Soft Conditions: A Variance Reduction Technique

Since the future paths of endogenous variables are unknown, a set of hard conditions for variables can be very different from their eventual realizations. For this reason, researchers may be interested in restricting the future values of a variable (such as the federal funds rate, M2 growth, or CPI inflation) within certain ranges rather than to single values. Conditions of this sort are called soft conditions. Soft conditions imply that the set $B(a)$ in equation (7) has a positive measure in $R^q$.

When the interval $(y_{T+n}(j), \bar{y}_{T+n}(j))$ is very narrow—which implies that the measure of $B(a)$ is small—algorithm (1) can provide a reliable approximation if one uses the midpoint of this interval for a hard condition. When the interval is wide, however, the approximation using algorithm (1) is less reliable. For example, if the interval is unbounded, such as the case in which CPI inflation is restricted below 3%, a different method must be used.

As long as the probability that forecasts will satisfy the soft conditions in equation (7) is not too small, a straightforward way to simulate the distribution of conditional forecasts is simply to draw $a$ and $\epsilon$ independently and keep the draws that satisfy the conditions. For each kept draw, compute $y_{T+n}$ according to equation (5). The empirical distribution can be formed from the simulated draws of $y_{T+n}$. Because many draws may be discarded, it is important that the simulation be as fast as possible. One method for improving speed exploits the fact that draws of $a$ from the posterior distribution (conditional on $Y_T$) are in general more expensive than draws of $\epsilon$ from the standard normal distribution.

To determine the distribution of $y_{T+n}$ conditional on the constraints given by (10), one approach would be to approximate this distribution via a histogram. This entails estimating the probability that $R(a)|\epsilon \in I(a)$ for various $I(a) \subseteq B(a)$, which is a special case of estimating $E[\cdot]$ where $g$ is any function of $a$ and $\epsilon$. If $n_1$ draws of $a$ are made—and for each draw of $a$, $n_2$ draws of $\epsilon$ are made—then an estimator of $E[\cdot]$ is

$$G(n_1, n_2) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} g(a_i, \epsilon_{i,j}).$$

where each $a_i$ is an independent realization from the distribution of $a$ and each $\epsilon_{i,j}$ is an independent realization from the distribution of $\epsilon$. Moreover, the draws of $a$ and $\epsilon$ are also independent. The estimator $G$ is consistent and unbiased. The term $n_2$ is called the oversampling rate. The goal is to choose $n_1$ and $n_2$ so that $G$ can be computed quickly and has small variance—two conflicting objectives. The best estimator of $E[\cdot]$ is defined to be the one that minimizes computing time subject to the constraint that its variance be smaller than some fixed level—or, equivalently, the one that minimizes variance subject to the constraint that its computing time be smaller than some fixed amount. In particular, suppose that it takes one unit of time to draw $a$, $s$ units of time to draw $\epsilon$, and the total amount of time available is $t$.

The objective is to minimize the variance of $G(n_1, n_2)$, subject to the constraint that $n_1 (1 + sn_2) \leq t$. The following proposition, proved in appendix B, determines the optimal oversampling provided $t$ is sufficiently large.

**Proposition (3).** Let $n_2$ be the value of $n_2$ that minimizes the variance of $G(n_1, n_2)$ subject to the constraint $n_1 (1 + sn_2) \leq t$. If $\text{var}(E_{a}[g|a]) > 0$ and $t$ is sufficiently large, then

$$\sqrt{\frac{1 - \gamma}{s \gamma}} - 1 < n_2 < \sqrt{\frac{1 - \gamma}{s \gamma}} + 1,$$

where

$$\gamma = \text{var}(E_{a}[g|a])/\text{var}(g).$$

The expression $E_{a}[g|a]$ is the expectation, with respect to $a$, of the random variable $g$ conditional on $a$. This itself is a random variable with respect to $a$, so its variance can be taken. The term $\gamma$ can be loosely interpreted as the proportion of the variance of the random variable $g$ that is due to parameter uncertainty. From this expression, it is clear that there should be more draws of $\epsilon$ for each draw of $a$ if either $s$ is small or the amount of variance due to parameter uncertainty is small.

---

9 See Robertson and Tallman (1999a, 1996b) for other evidence.
10 In a recent paper, Sims (1999) argues that the location of parameters can be better characterized by the posterior density under a widely used informative prior than under an ignorant (flat) prior.
11 In an example discussed in the next section, $s$ is on the order of 0.01. In general, $s$ will depend heavily on the particular algorithms used to draw $a$ and $\epsilon$, the number of variables, and the forecast horizon, but not on the speed of the computer used in simulation.
Proposition (3) gives the optimal oversampling rate. The following proposition, which is also proved in appendix (2), shows how much of a variance reduction can be expected.

**Proposition (4).** Let $n_1 (1 + n_2) \leq t$. The percentage reduction in variance obtained by using $n_2$ draws of $\epsilon$ for each draw of $a$, relative to using one draw of $\epsilon$ for each draw of $a$, is approximately

$$100 \left(1 - \frac{1 + sn_2 1 + (n_2 - 1)\gamma}{1 + s n_2} \right),$$

for sufficiently large $t$.

To find the optimal oversampling rate or the reduction in variance, we must be able to compute $\gamma$. Although there is no convenient analytical expression for $\gamma$, its value can be estimated through simulation.

Once the oversampling rate is chosen, the following algorithm is designed to simulate the forecast distribution under soft conditions. As in algorithm (1), a method for drawing $a$ from the posterior distribution given by equation (9) is assumed to be available.

**Algorithm (2).** For $1 \leq i \leq n_1$,

(a) draw $a^{(i)}$ according to the posterior density function (9);
(b) for each $a^{(i)}$, draw $\epsilon^{(i,j)}$ independently from the standard normal distribution for $1 \leq j \leq n_2$;
(c) for each pair $(a^{(i)}, \epsilon^{(i,j)})$, use (5) to compute $y_{T+h}^{(i,j)}, \ldots, y_{T+h}^{(i,j)}$;
(d) repeat steps (a) through (c) until the sequence $y_{T+h}^{(i,j)}, \ldots, y_{T+h}^{(i,j)}$ for $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$ is completed;
(e) keep the simulated draws in the sequence that satisfy (7).

Algorithm (2) can be easily implemented because draws of $a$ are sampled independently of draws of $\epsilon$. As will be seen in the next section, algorithm (2) can provide a computationally efficient alternative to algorithm (1).

**IV Examples**

This section applies the methods developed in Section III to the VAR model used in Zha (1998) to show how these methods can be used for obtaining the finite-sample distribution of conditional forecasts out of sample. As shown in proposition (1), the distribution of conditional forecasts is invariant to orthonormal transformation of $A_0$. Following the convention, therefore, we restrict $A_0$ to be upper triangular.

Two cases are considered: the flat prior and the SZ prior. The model used in this section employs monthly data for the six macroeconomic variables: IMF’s index of world commodity prices (Pcm), M2, the federal funds rate (FFR), real gross domestic product (GDP), the consumer price index (CPI), and the unemployment rate (U). (See appendix A for precise descriptions.) All variables are logarithmic except the federal funds rate and the unemployment rate, which are expressed in percentages. The model includes thirteen lags. The data used to fit the model begin with 1959:1 and end at 1980:12.

The early 1980s is often considered a difficult period for forecasting macroeconomic variables. Inflation in 1980 reached the highest point since 1960 and declined rapidly thereafter. Real GDP growth in 1982 was very negative (−2.13%) and then increased rapidly in subsequent years (3.94% in 1983 and 7.02% in 1984). The model is used to predict, out of sample, the paths of these and other macroeconomic variables.

Since movements in the federal funds rate are often used to explain fluctuations in other macroeconomic variables, all examples in this section use conditions that restrict only the federal funds rate. The effects of such conditions on other endogenous variables over a four-year horizon are examined via conditional forecasts. The examples emphasize three main results. Under the flat prior, the distribution of conditional forecasts can shift when parameter uncertainty is taken into account. The SZ prior, by contrast, reduces the shift in the distribution while at the same time improving out-of-sample forecasts. The last result shows that the soft-condition method can provide a reasonable approximation to the hard-condition method.

**A. A Hard Condition under the Flat Prior**

With few exceptions, VAR models used in the macroeconomic literature do not impose informative priors (e.g., Christiano et al., 1999; Pagan & Robertston 1998). This subsection, therefore, focuses on the case of the flat prior. The imposed hard condition is that the federal funds rate follows the path of the actual annual average rates in 1981 through 1984. Algorithm (1) is used to simulate the distribution of conditional forecasts. In generating these forecasts, step (b) of algorithm (1) takes account of parameter uncertainty in finite samples. Figure 1 displays conditional forecasts with probability bands. The solid line represents the actual data; the dashed line represents the posterior means of forecasts; the two dashed and dotted lines...
around the dashed line represent the 16th and 84th percentiles so that the bands contain 0.68 probability. All variables are expressed in percentage changes in annual rates except the federal funds rate and unemployment, which are expressed in average percentage rates.

As clearly shown in the figure, the forecast bands do not capture the actual movements in many variables. Here, the patterns of these bands are briefly summarized. The lower forecast band of \( \text{Pcm} \) is close to the actual values in 1982–1983 but far from the actual values in 1981 and 1984; the recovery of GDP in 1981 is completely missed, and the 1982 GDP forecast signals a far more severe recession than the actual outcome; the forecast bands of CPI show a downward trend, but the actual values are still far away from the lower band; and the forecasts of \( U \) are above the actual values in 1982–84 by a large margin as measured by the probability bands. These forecasts will be compared to those under the SZ prior in the next subsection.

We now discuss the effects of parameter uncertainty on conditional forecasts. To examine these effects, algorithm (1) is used to generate the probability bands of conditional forecasts with the parameters fixed at the MLEs. Overall, these probability bands tend to be much narrower than those

---

15 All simulations in this paper use at least 6,000 draws that satisfy constraint (9); all probability bands are constructed to contain 0.68 probability. The bands demarcate the simulated marginal distributions of forecasts at each point of the time horizon, not for the horizon as a whole. (See Zha (1998) for examples of demarcating joint distributions of forecasts.) The small discrepancy between the dashed and solid lines for the forecast of the funds rate in figure 1 is due to the memory-conserving techniques used in storing the simulated draws.
in figure 1. Moreover, the marginal distributions of some conditional forecasts shift significantly. A telling example is given in figure 2, which displays the density functions of the 1983 forecast of M2 growth with and without parameter uncertainty. The vertical line marks the actual M2 growth rate in 1983. Clearly, the forecast distribution with the parameters fixed at the MLEs is unrealistically tight. By contrast, the forecast distribution with parameter uncertainty not only widens but also shifts to the left.

A possible shift in the distribution of a conditional forecast is in sharp contrast to classical asymptotic results. In the classical framework, since the “true” parameters are replaced by their estimates (MLEs in our case), the asymptotic forecast band allowing for parameter uncertainty will only widen the band with the “true” fixed values of parameters. But, in finite samples, the location of “true” parameters in a typical VAR model is quite uncertain. When the shape of the likelihood is not informative (i.e., flat) around the peak, the MLEs can change with the addition of a few new observations. As discussed before, the conditional forecasts are essentially equivalent to adding $h$ new observations in step (b) of algorithm (1). As a result, the peak of the unconditional likelihood that the distribution of a conditional forecast can shift when parameter uncertainty is taken into account. The example displayed in figure 2 underlines the finite-sample uncertainty about the model’s parameters as an important factor in statistical inferences on conditional forecasts. It also suggests that informative priors can improve finite-sample inferences on conditional forecasts.

B. A Hard Condition under the SZ Prior

The SZ prior is designed to influence the shape of the likelihood in directions that better characterize the behavior of macroeconomic time series. The prior contains components favoring unit roots and cointegration while avoiding the imposition of exact (but possibly false) restrictions. Such a prior is of reference nature because it is introduced to reflect widely held beliefs about the multivariate dynamics of macroeconomic time series among economists. The prior means of parameters are set to zero. The prior variances are controlled by the values of tightness hyperparameters, which equal those in Leeper and Zha (1999). Specifically, we follow the notation of Sims and Zha (1998) and let $\lambda_0 = 0.57$, $\lambda_1 = 0.13$, $\lambda_4 = 0.1$, $\mu_5 = 5$, and $\mu_6 = 5$. The value of $\lambda_0$ controls an overall tightness in the prior variances of all parameters; the value of $\lambda_1$ controls a relative tightness in parameters of lagged endogenous variables; $\lambda_4$ controls a relative tightness in constant terms; $\mu_5$ reflects the strength of a belief in unit roots; and $\mu_6$ reflects

---

16 Diebold et al. (1998) address the importance of density forecasts and suggest ways of evaluating such forecasts in a univariate case. Although it is beyond the purpose of this paper to select a model that provides the best forecast, it will be a challenging task in future research to evaluate density forecasts among different models in a multistep, multivariate framework.

the strength of a belief in stationarity and cointegration. As for the decay rate of lag length, denoted by $\lambda_3$, the value is usually set to 1 when quarterly data are used (e.g., Doan, 1992; Miller & Roberds, 1991). For the monthly model here, the lag decay rate declines in an exponential fashion so that the degree of decay in the thirteenth month matches that in the fifth quarter.\(^{18}\)

Applying the SZ prior to the VAR model, algorithm (1) is used to simulate the distribution of forecasts conditional on the actual annual average funds rates in 1981–1984. Figure 3 presents the forecasts along with probability bands. A comparison of figure 1 and figure 3 confirms that the SZ prior helps improve the overall accuracy of out-of-sample forecasts.\(^{19}\) In particular, the forecast bands for GDP growth in figure 3 look quite reasonable. The actual GDP growth rates are almost all within the probability bands; the 1982 recession is detected by the lower forecast band. The actual annual Pcm changes are within the probability bands as well. The forecast bands of CPI inflation capture the downward trend in actual inflation. Compared to figure 1, the actual inflation path is much closer to the lower forecast band in figure 3. Similarly, the lower forecast band for U in figure 3 is much closer to the actual path than that in figure 1.

We now address an important issue raised in the previous subsection: the effects of parameter uncertainty on condi-

\(^{18}\) See also Robertson and Tallman (1999a, 1999b) for details.

\(^{19}\) For comprehensive comparisons of out-of-sample forecasts under the SZ prior, the flat prior, and the Litterman (1986) prior, see Robertson and Tallman (1999a).
tional forecasts. In contrast to the flat prior, the SZ prior has effects on conditional forecasts to the extent that it only widens the probability distributions of forecasts. As an example, figure 4 displays the density functions of the 1984 U forecast with and without parameter uncertainty. Two results are worth discussion. First, figure 4 presents a clear case in which parameter uncertainty plays an important role in obtaining the distribution of a conditional forecast. As shown in the figure, the distribution of the 1984 U forecast with parameter uncertainty gives a higher density to the actual unemployment rate (marked by the vertical line) than does the distribution without parameter uncertainty. In other words, the distribution of the 1984 U forecast without parameter uncertainty is too confident about the actual realization.

The second result relates to a shift in distribution. Unlike figure 1, figure 4 shows no significant shift in the distribution of the forecast. This result accentuates the importance of an informative prior in finite-sample inferences. The SZ prior substantially reduces the degree of ill-behaved flatness in the likelihood. The peak of the resulting posterior density is unlikely to be affected by a few simulated observations in step (b) of algorithm (1). Consequently, the shape of the posterior density under the SZ prior is well behaved and informative relative to the likelihood shape under the flat prior.

C. A Soft Condition with the SZ Prior

This subsection applies the soft-condition method developed in section III.B, to an example similar to those examined in the previous sections. Here, the soft condition constrains the funds rate over 1981–1984 to be within plus-or-minus two percentage points of the actual annual

---

20 We did not choose the 1983 M2 growth forecast as an example because it does not show as a notable effect of parameter uncertainty as figure 4.

21 The probability distribution of forecasts without parameter uncertainty is simulated with the values of parameters fixed at the MLEs. Here, the MLEs are the generalized maximum likelihood estimates obtained at the peak of the posterior density function.
average rates. The SZ prior is used, and the oversampling rate is chosen using proposition (3).

For the six-variable model with a 48-month (four-year) forecast horizon, $s$ — the ratio of computing time required to draw $e$ to that of drawing $a$ — is about 0.01 in MATLAB. The parameter $\gamma$ depends on the random variable $g$. To compute the distribution of the conditional forecast, one must be able to estimate $E[g]$, where $g$ is the indicator function that assumes the value one if $R(a)\epsilon \subseteq \Pi(a) \subseteq B(a)$. Different $\Pi(a)$ will produce different values of $\gamma$, but simulations indicate that $\gamma$ is largest when $\Pi(a) = B(a)$. In this case, $\gamma$ is approximately 0.067, which corresponds to an oversampling rate of about 37. Figure 5 displays percentage reductions in variance as a function of the oversampling rate for various values of $\gamma$. The curves are plotted using proposition (4) with $s = 0.01$. In this example, $\gamma$ lies between 0 and 0.067, the top two lines in Figure 5. From this, one sees that using a oversampling rate of 37 will give close to optimal results for

**Figure 6.**—1980:12 Conditional Forecasts with Soft Condition and Parameter Uncertainty Under SZ Prior
all $g$ of interest. Other values of $\gamma$ included in figure 5, though not applicable in this example, give the reader an idea of the behavior of the variance reduction as $\gamma$ increases.

Simulations using algorithm (2) with 5,000 draws of $a$ and an oversampling rate of 37 took approximately 2.5 hours on a 266 Pentium II PC. Out of these 185,000 MC draws, about 8,000 draws satisfy the soft condition on FRR. If one draw of $\epsilon$ were taken for each draw of $a$, figure 5 implies that approximately sixteen hours would be needed to achieve the same accuracy for the simulated distribution of conditional forecasts.

Figure 6 reports simulated results with probability bands attached. Since all bands contain 0.68 probability, the actual data—e.g., for the federal funds rate—may lie outside the bands. Clearly, the results in figure 6 are quite close to those in figure 3 where the hard condition is imposed. Compared to figure 3, the bands in figure 6 are somewhat wider and shift slightly for a few forecasts. This example shows that, in addition to providing forecasts when the constraints are soft, the method of section III.B provides an efficient way to approximate the hard conditions.

V. Conclusion

Conditional forecasts are designed to answer many practical questions that cannot be answered by unconditional forecasts. Policymakers might want to know, for example, the effects of a contractionary monetary policy on the future state of the economy (Leeper & Zha, 1999). Forecasters might be interested in how a forecast changes if the federal funds rate or CPI inflation follows a certain path or range in the future. In real-time forecasting in which some data are released sooner than others, analysts would like to examine forecasts conditional on the released data.

To address these practical issues, this paper broadens the class of conditional forecasts in the VAR literature and develops methods for obtaining the exact finite-sample distribution of conditional forecasts. Empirical examples are used to show how the methods can be implemented and to highlight an important role of finite-sample inferences on parameters in conditional forecasts. It is hoped that the methods will help applied researchers analyze the effects on macroeconomic forecasts when conditions are imposed on endogenous variables in the model.

22 In contrast, computing time for the results in figure 3 is about thirteen hours. The demanding part of that computation is producing the singular value decomposition of the large $288 \times 288$ covariance matrix in equation (12) at each iteration.

REFERENCES


Appendix A

The empirical model estimated in this paper uses monthly data from 1959:1 to 1980:12 for the six macroeconomic variables:

- Pcm: International Monetary Fund’s Index of world commodity prices. Source: International Monetary Statistics.
- M2: M2 money stock, seasonally adjusted, billions of dollars. Source: Board of Governors of the Federal Reserve System (Board).
- FFR: Effective rate, monthly average. Source: Board.
- CPI: Consumer price index for urban consumers (CPI-U), seasonally adjusted. Source: BEA.
- U: Civilian unemployment rate (ages sixteen and over), seasonally adjusted. Source: Bureau of Labor Statistics.

Appendix B

This appendix provides proofs of propositions (3) and (4). These proofs depend on a careful expansion of the variance of \(G(n_1, n_2)\). Following the notation of section III.B,

\[
\text{var} (G(n_1, n_2)) = \frac{1}{n^2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij})))
\]

Splitting this sum into terms for which \(i = j\) and \(j = l\), for which \(i = k\) and \(j \neq l\), and for which \(i \neq k\), we obtain

\[
\text{var} (G(n_1, n_2)) = \frac{1}{n^2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij})))
\]

\[
+ \frac{2}{n^2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij})))
\]

\[
+ \frac{2}{n^2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij})))
\]

Since

\[
\text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij}))) = \text{var} (g),
\]

\[
\text{cov} (g(a_i, e_{ij}), (g(a_i, e_{ij}))) = \text{var} (E[g|a]),
\]

\[
\text{cov} (f(a_i, e_{ij}), (f(a_i, e_{ij}))) = 0,
\]

it follows that

\[
\text{var} (G(n_1, n_2)) = \frac{1}{n^2} n_1 n_2 \text{var} (g)
\]

\[
+ \frac{2}{n^2} n_1 (n_1 - 1) 2 \text{var} (E[g|a])
\]

\[
= \text{var} (g) + (n_2 - 1) \text{var} (E[g|a])
\]

\[
= \frac{n_1 n_2}{n_1} \text{var} (g) + (n_2 - 1) \gamma
\]

From this expression, it is easy to see that, given \(n_2\), the value of \(n_1\) that minimizes the variance of \(G(n_1, n_2)\), subject to the constraint \(n_1 (1 + \gamma n_2) \leq t\), is the integer less than or equal to \(t((1 + \gamma n_2)\), which is denoted by \(\hat{n}_1(n_2, t)\).

Proof of proposition (4)

The percentage reduction in variance is \(100(1 - \lambda(n_2, t))\), where

\[
\lambda(n_2, t) = \frac{\text{var} (G(\hat{n}_1(n_2, t), n_2))}{\text{var} (G(\hat{n}_1(1, t), 1))}.
\]

Proposition (4) will follow if it can be shown that

\[
\lim_{t \to \infty} \lambda(n_2, t) = \frac{1 + s n_2}{1 + s} \frac{1 + (n_2 - 1) \gamma}{n_2} \lim_{t \to \infty} \hat{n}_1(1, t) = \frac{1}{n_2}.
\]

Since

\[
t - (1 + s) < \frac{s}{1 + s} \hat{n}_1(1, t) \leq \frac{t}{1 + s}
\]

\[
t - (1 + s n_2) < \frac{1 + s n_2}{1 + s} \frac{1 + (n_2 - 1) \gamma}{n_2} \lim_{t \to \infty} \hat{n}_1(1, t) = \frac{1}{n_2}.
\]

it is easy to see that

\[
1 + s n_2 = \lim_{t \to \infty} \frac{t - (1 + s) + s n_2}{1 + s} \leq \lim_{t \to \infty} \frac{t}{1 + s} \hat{n}_1(n_2, t)
\]

\[
\leq \lim_{t \to \infty} \frac{t}{1 + s} \frac{1 + s n_2 - (1 + s n_2)}{1 + s} = 1 + s n_2.
\]

This completes the proof of the proposition. QED

Proof of proposition (3)

Minimizing the variance of \(G(n_1, n_2)\) subject to the constraint \(n_1 (1 + \gamma n_2) \leq t\) is equivalent to minimizing \(\lambda(n_2, t)\). But, by proposition (4), as \(t\) tends to infinity, \(\lambda(n_2, t)\) converges to

\[
\frac{1 + s n_2}{1 + s} \frac{1 + (n_2 - 1) \gamma}{n_2}.
\]

Furthermore, equation (14) implies that the rate of convergence from below is independent of \(n_2\). Treating \(n_2\) as a real variable and taking the derivative of equation (15) with respect to \(n_2\), one sees that the minimum of (15) occurs at

\[
1 - \frac{1}{2} (1/\gamma - 1).
\]

(13)

Thus, for sufficiently large \(t\), the value of \(n_2\) which minimizes \(\lambda(n_2, t)\) satisfies

\[
\frac{1}{2} (1/\gamma - 1) - 1 < n_2 < \sqrt{-\frac{1}{2} (1/\gamma - 1) + 1}.
\]

QED.