Conditional Forecasts in Dynamic Multivariate Models

Daniel F. Waggoner and Tao Zha

Federal Reserve Bank of Atlanta
Working Paper 98-22
December 1998

Abstract: In the existing literature, conditional forecasts in the vector autoregressive (VAR) framework have not been commonly presented with probability distributions or error bands. This paper develops Bayesian methods for computing such distributions or bands. 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 the parameter uncertainty in small samples. Empirical examples under the flat prior and under the reference prior of Sims and Zha (1998) are provided to show the use of these methods.

JEL classification: C32, E17, C53

Key words: conditional forecasts, hard and soft conditions, Bayesian methods, probability distribution, error bands, likelihood

---

Conditional Forecasts in Dynamic Multivariate Models

1. Introduction

In policy analysis, it is believed that monetary policy has long and variable effects on the overall economy. To capture such complex interactions between policy variables and the economy as a whole, macroeconomic forecasting becomes indispensable in actual policy making (Kohn 1995, Blinder 1997, and Diebold 1998). In a recent paper, Sims and Zha (1998) introduced Bayesian methods to vector autoregressive (VAR) models to improve the accuracy of out-of-sample forecasts in a dynamic multivariate framework. They showed how to compute Bayesian probability distributions or error bands around out-of-sample forecasts. Their methods apply only to forecasts with no conditions on future variables or future structural shocks. These are often called unconditional forecasts in the forecasting literature.

Forecasters, as well as policy analysts, however, are often interested in questions like "how do the forecasts of other macroeconomic variables change if the federal funds rate in the next two to three years follows different paths?" In the framework of dynamic multivariate models such as VARs, these kinds of questions require one to impose conditions, prior to forecasting, on the future values of certain endogenous variables such as the federal funds rate. Forecasts associated with such conditions are called conditional forecasts. Although Doan, Litterman, and Sims (1984, DLS hereafter) showed how to calculate point conditional forecasts in a Bayesian framework,[1] probability distributions or error bands around conditional forecasts in VAR models have not been commonly discussed or presented in the existing literature. If conditional

---

[1] The algorithm is available from the software package RATS (Doan 1992).

forecasts are used to guide policy decisions, it is important that a probability assessment, rather than simply point forecasts, be provided.

This paper develops two Bayesian methods for computing probability distributions of conditional forecasts. Both methods work with structural VARs as well as reduced-form VARs. One method relates to conditions that fix the future values of variables at single points. For example, the future funds rate is restricted to, say, 5% in the next year. These conditions have been often considered in the forecasting literature and are called hard conditions in this paper. The other method deals with conditions that only restrict the future values within a certain range. The future values pertain to either variables or structural shocks or both. Examples of such conditions are a certain range of the funds rate path, a target range of M2 growth rate, and a contractionary region for monetary policy shocks. The concept of conditioning directly on structural shocks is introduced by Leeper and Zha (1998) and is closely related to the recent work in structural VAR literature. These types of conditions are called soft conditions in this paper. Even in the case of unconditional forecasts, the soft-condition method proves more efficient than the approach used in Sims and Zha (1998).

The main thrust of both the hard-condition and soft-condition methods is their approach to accounting for parameter uncertainty in small samples via the shape of the likelihood or posterior density. The common practice is to fix parameters at, say, the maximum likelihood estimates (MLEs) in out-of-sample forecasts, thus ignoring parameter uncertainty. Yet to what extent parameter uncertainty matters for forecast errors is an important empirical question and can be answered by simulation. This paper provides empirical methods to answer this question. It is shown that ignoring parameter uncertainty can result in potentially misleading results. Thus, it is

1

important that the effect of the parameter uncertainty on forecasts be examined *before* one proceeds to forecast with parameters fixed at some estimate.

The remainder of this paper is organized as follows. Section 2 lays out a general framework and discusses conceptual differences between the types of conditional forecasts. Section 3 develops the theoretical foundation of Bayesian methods for computing probability distributions or error bands around conditional forecasts. Section 4 provides empirical examples to show how to use these methods by focusing on conditions imposed on future values of variables over the forecast horizon. Section 5 concludes this paper.

2. Conditional Forecasts

2.1. General Framework

The dynamic multivariate framework considered in this paper has the general, structural form[2]:

$$\sum_{\ell=0}^{p} \underset{1\times m}{y_{t-\ell}} \underset{m\times m}{A_\ell} = \underset{1\times m}{d} + \underset{1\times m}{\varepsilon_t}, \quad t=1,\cdots,T, \tag{1}$$

where T is the sample size, y_t is a vector of observations, A_ℓ is the coefficient matrix of the ℓth lag, p is the maximum number of lags, d is a vector of constant terms, and ε_t is a vector of i.i.d. structural shocks that are Gaussian with

$$E\left(\varepsilon_t' \varepsilon_t \mid y_{t-s}, s>0\right) = \underset{m\times m}{\mathbf{I}} \quad \text{and} \quad E\left(\varepsilon_t \mid y_{t-s}, s>0\right) = \underset{1\times m}{\mathbf{0}}, \text{ for all } t. \tag{2}$$

This paper considers only linear restrictions on the contemporaneous coefficient matrix A_0 which is assumed to be non-singular.

---

[2] Columns in A_ℓ correspond to equations.

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

$$y_t = c + \sum_{\ell=1}^{p} y_{t-\ell} B_\ell + \varepsilon_t A_0^{-1}, \text{ for all } t. \tag{3}$$

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

$$c = dA_0^{-1} \text{ and } B_\ell = A_\ell A_0^{-1} \quad \ell = 1, \cdots, p. \tag{4}$$

Given (3), (4), and the data up to time T, the h-step out-of-sample forecast at time T can be written as

$$y_{T+h} = c\mathrm{K}_{h-1} + \sum_{\ell=1}^{p} y_{T+1-\ell} N_\ell(h) + \sum_{j=1}^{h} \varepsilon_{T+j} M_{h-j}, \quad h = 1, 2, \ldots \tag{5}$$

where

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

$$N_\ell(1) = B_\ell, \quad \ell = 1, \cdots, p;$$

$$N_\ell(h) = \sum_{j=1}^{h-1} N_\ell(h-j) B_j + B_{h+\ell-1}, \quad \ell = 1, \cdots, p, \quad h = 2, 3, \cdots;$$

$$M_0 = A_0^{-1}, \quad M_i = \sum_{j=1}^{i} M_{i-j} B_j \quad i = 1, 2, \cdots;$$

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 M_i which is known as the matrix of impulse responses. When there are conditions or constraints imposed on future values of the variables, the forecast produced

through (5) is called a conditional forecast. When there are no such conditions, the forecast is called an unconditional forecast. The concept of conditional forecast in this paper is different from the traditional one. Traditionally, conditions imposed on the future values concern only exogenous variables (Intriligator, Bodkin, and Hsiao 1996, pp.518-532). In dynamic multivariate models like (1), conditions are imposed directly on the future values of endogenous variables or structural shocks.[3] Such conditions make it conceptually and numerically difficult to obtain the error bands on conditional forecasts.

2.2. *Distributions of Conditional Forecasts*

Two sources of uncertainty generate the errors associated with forecast y_{T+h}. One source pertains to unpredictable disturbances ε_{T+j}, $j = 1, \cdots, h$, which are assumed to have a Gaussian distribution. The other source of uncertainty relates to the likelihood shape of model parameters (d, A_ℓ). In the Bayesian framework, the exact small-sample property (likelihood) of parameters can be conveniently explored through the posterior distribution.[4] Specifically, let

$$a = \begin{pmatrix} a_0 \\ a_+ \end{pmatrix}, \text{ where } a_0 = vec(A_0), \text{ and } a_+ = vec\begin{pmatrix} -A_1 \\ \vdots \\ -A_p \\ d \end{pmatrix}.$$

With the flat prior or the informative prior of Sims and Zha (1998), the posterior distribution of a has the form:

$$p(a|\mathbf{Y}_T) = \pi(a_0)\pi(a_+|a_0), \tag{6}$$

where $\mathbf{Y}_T$ denotes the data matrix up to time T and

---

[3] Although structural shocks are exogenous to the model, they are stochastic. Traditionally, exogenous variables in conditional forecasts are held at fixed values.

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

$$\pi(a_+|a_0) = \varphi\big((I \otimes U)a_0; I \otimes V\big) \tag{7}$$

Here, $\varphi(\mu; \Sigma)$ denotes the normal density function with mean μ and variance Σ. In (7), S, U, and V are matrix functions of the data $\mathbf{Y}_T$ (and the prior mean and variance when the informative prior of Sims and Zha (1998) is introduced).[5] Depending on the 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 (6) (see Waggoner and Zha 1997 and Zha 1997).

In the existing VAR literature, interest has focused on the future values of variables y_{T+h}. Consider the following example. Suppose that the value of the jth variable $y_{T+h}(j)$ is constrained to be in the range $\left[\underline{y}_{T+h}(j), \overline{y}_{T+h}(j)\right]$. From (5), this constraint implies the following condition:

$$\sum_{n=1}^{h} \varepsilon_{T+n} M_{h-n}(\cdot, j) + cK_{h-1}(\cdot, j) + \sum_{\ell=1}^{p} y_{T+1-\ell} N_\ell(h)(\cdot, j) \in \left[\underline{y}_{T+h}(j), \overline{y}_{T+h}(j)\right], \tag{8}$$

where the notation $(\cdot, j)$ denotes the jth column of the matrix. Moving the last two terms on the left hand side of (8) to the right hand side, condition (8) can be generalized to have the compact form encompassing multiple conditions:

$$\underset{q \times k}{R(a)'} \underset{k \times 1}{\varepsilon} \subseteq B(a) \subseteq \mathbf{R}^q, \quad q \le k = hm, \tag{9}$$

where q is the number of conditions or constraints, k is the total number of future shocks, $R(a)'$ is a stacked matrix from impulse responses $M_{h-n}(\cdot, j)$, ε is a vector correspondingly

---

[4] See Sims and Zha (1995) for detailed discussions about the difficulties associated with various classical approaches.

[5] See Sims and Zha (1998) and Waggoner and Zha (1997) for details.

stacked from ε_{t+n}, and B is the restricted set in the q-dimentional Euclidean space. It is clear

from (8) that both R and B may depend on the values of parameters a. The unconditional

forecast is simply a special case of the conditional forecast when B is set to be the unrestricted

Euclidean space $\mathbf{R}^q$ in (9).

Before developing simulation methods for conditional forecasts, let us broaden the concept

of a condition discussed above. As introduced by Leeper and Zha (1998), a projection of the

out-of-sample effects of monetary policy in the structural VAR framework often relies on

conditions pertaining to future values of structural shocks ε_{T+h} rather than variables.

Mathematically, conditions on shocks can be put in the exact form of constraint (9) except that in

this situation neither R nor B in (9) depends on structural parameters a.

Although the methods developed later in this paper apply to conditions on both variables and

shocks, these conditions have distinct implications for conditional forecasts. Because structural

shocks such as a monetary policy shock have clear economic interpretations, conditions imposed

on future shocks deliver different distributions of conditional forecasts for different identifying

restrictions in A_0 (no matter whether A_0 is overidentified or exactly identified). This conclusion

can be easily seen through (5). Since paths of impulse responses M_i depend on particular A_0,

distributions of forecasts y_{t+h} conditioned on the range of variations in future shocks will depend

on A_0 as well.

On the other hand, with conditions imposed on future variables as DLS proposed, the

forecast distribution is invariant to orthonormal transformation of system (1). The following

proposition formally establishes this result.

Proposition 1. The marginal distribution of y_{t+h} subject to constraint (8) is invariant to

orthonormal transformation of system (1).

Proof. An orthonormal transformation of A_0 is equivalent to post-multiplying system (1) by an orthogonal matrix P. Denote $\varepsilon_t P$ by δ_t. Because there are no constraints on ε_t, δ_t remains Gaussian and satisfies assumption (2). Obviously from (4) this transformation leaves reduced-form parameters c and B_ℓ ($\ell = 1,...,p$) unaffected. Since K_{h-1} and $N_\ell(h)$ in (5) are simply functions of B_ℓ, these terms are not affected either. The only term that is affected by the transformation is M_{h-j}. According to (5) and (8), this term enters the forecast of y_{T+h} through:

$$\sum_{n=1}^{h}(\delta_{T+n}P')M_{h-n} .$$

Since $\delta_{T+n} = \varepsilon_{T+n}P$, $\delta_{T+n}P'$ has the same distribution as ε_{T+n} and is independent of M_{h-n}. Thus, the marginal distribution of conditional forecast y_{T+h} is invariant to the transformation P.

Q.E.D.

A special case of Proportion 1 relates to exactly identified models. If system (1) is exactly identified, Proposition 1 implies that the forecast distribution with constraints on future *variables* does not depend on any particular identification of A_0.[6] In sharp contrast, the forecast distribution conditioned on future *shocks* depends on a particular identification of A_0. Because of this reduced-form nature for forecasts conditioned on future variables, the convention is to parameterize A_0 to be triangular (see DLS).

3. Simulation Methods for Probability Distributions

Probability distributions of conditional forecasts or even error bands around conditional forecasts have not been commonly (in fact, not at all to our knowledge) presented in the existing

---

[6] In the situation whereby linear restrictions relate to lag structure A_ℓ, this invariance property fails to hold because there does not exist, in general, an orthonormal transformation from the system with upper triangular A_0 to the system with lower triangular A_0.

VAR literature (e.g., Sims 1982, DLS, Miller and Roberds 1991, and Roberds and Whiteman 1992). Since all forecasts contain errors, some of which are substantial, it is important that error bands or marginal probability distributions be provided for the assessment of forecast errors. The methods developed in this section focus on conditions imposed on the future values of variables ((8) or (9)). When conditions are imposed directly on future structural shocks so that ε_{T+h} is restricted to a fixed value or within a certain range, they can be expressed also in form (9). Thus the following methods, without alteration, apply to this type of conditions as well.

3.1. Hard Conditions: A Gibbs Technique

The conditions discussed in the existing literature concern the situation in which the value of, say, $y_{T+n}(j)$ is restricted to be a single value so that $\underline{y}_{T+n}(j) = \overline{y}_{T+n}(j) = y^*_{T+n}(j)$. Specifically,

$$y_{T+n}(j) = y^*_{T+n}(j), \quad \begin{matrix} n \in N \subseteq (1,2,\ldots,h) \\ j \in J \subseteq (1,2,\ldots,m) \end{matrix}. \tag{10}$$

According to (8) and (9), the set of conditions in (10) implies that $B(a)$ collapses to a $q \times 1$ vector of values. Denote this $q \times 1$ vector by r. The set (10) can be equivalently expressed as

$$\underset{q \times k}{R(a)'} \underset{k \times 1}{\varepsilon} = \underset{q \times 1}{r(a)}, \quad q \le k = mh. \tag{11}$$

This paper calls the exact conditions in (10) or (11) "hard conditions." To derive a method for generating the distribution of $y_{T+h}(j)$ under these hard conditions, let us first establish the following proposition.

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

$$p\left(y_{T+n}|a, \mathbf{Y}_{T+n-1}\right) = \varphi\left(c + \sum_{\ell=1}^{p} y_{T+n-\ell} B_{\ell} + \mathsf{M}\left(\varepsilon_{T+n}\right) A_0^{-1}; A_0^{-1'} \mathsf{V}\left(\varepsilon_{T+n}\right) A_0^{-1}\right), \quad n = 1,2,\ldots,h, \tag{12}$$

8

where $\mathbf{Y}_{T+n-1}$ is the data matrix up to time $T+n-1$, $\mathsf{M}(\varepsilon_{T+n})$ and $\mathsf{V}(\varepsilon_{T+n})$ are the mean and

variance of ε_{T+n} whose distribution is normal with the following form:

$$p\big(\varepsilon|a, R(a)'\varepsilon = r\big) = \varphi\Big(R(a)\big(R(a)'R(a)\big)^{-1}r; I - R(a)\big(R(a)'R(a)\big)^{-1}R(a)\Big). \qquad (13)$$

Proof. By assumption (2), the unconditional distribution of ε is standard normal. Hence,

constraint (11) implies the conditional distribution of ε is given by (13). As a result, the

marginal distribution of ε_{T+n} is normal as well and its mean and variance can be read off directly

from (13). Given a, it is clear from (3) that the mapping between y_{T+n} ($n = 1,2,\ldots,h$) and ε is

linear and one-on-one. Thus, the conditional distribution (12) follows directly from (3).

Q.E.D.

Proposition 2 gives the distributions of the conditional forecasts when the values of

parameters are taken as given. It implies the existence of infinite paths of forecasts y_{T+n} that

satisfy the set of conditions in (10) or (11). The procedure used in previous work is to derive a

single path of forecasts by minimizing the objective function $\varepsilon'\varepsilon$ subject to (11) (see DLS and

Doan 1992). It can be easily seen that the solution to this optimization problem is exactly the

conditional mean of ε in (13).[7] While the procedure in the existing literature offers the most

likely path of conditional forecasts, it ignores the errors associated with the uncertainty of future

shocks. The method laid out in Proposition 2 can be easily implemented to take account of this

source of uncertainty.

The aforementioned method takes the values of parameters as fixed. If one also takes

account of the uncertainty about model parameters, simulating the distributions of conditional

---

[7] In DLS, this result is derived under the assumption that model (3) is stationary. This assumption is never required for Proposition 2. In other words, Proposition 2 is valid no matter whether the model is stationary or not.

forecasts becomes a challenging task. It is tempting, though, to draw parameters a from the

posterior distribution (conditional on $\mathbf{Y}_T$) and then condition on these draws to generate y_{T+n} by

Proposition 2. Sensible though this procedure might seem, the probability distribution of y_{T+n}

thus computed is incorrect because draws of a from the posterior ignore constraint (10) or (11)

in which R and r are nonlinear functions of a. The correct marginal distribution of a

conditioned on (10) must derive from the joint distribution of a and y_{T+n}. Although the joint

distribution of a and y_{T+n} ($n=1,2,\ldots,h$) conditional on (10) is in general unknown, it is feasible

to simulate this distribution using a Gibbs sampler technique.[8] The following algorithm lays out

the detail of this simulation.

Algorithm 1. Initialize an arbitrary value $a^{(0)}$ (e.g., the value at the peak of $p(a|\mathbf{Y}_T)$ or any

value randomly drawn from density $p(a|\mathbf{Y}_T)$). For $i=1,2,\ldots,N_1+N_2$,

    (a) generate $y_{T+n}^{(i)}$ ($n=1,2,\ldots,h$) from $p\left(y_{T+1},\ldots,y_{T+h}\big|a^{(i-1)},\mathbf{Y}_T\right)$ by Proposition 2 (i.e., draw

        ε from (13) and then use (3) to obtain y_{T+n} ($n=1,2,\ldots,h$));

    (b) generate $a^{(i)}$ from density $p\left(a\big|y_{T+1}^{(i)},\ldots,y_{T+h}^{(i)},\mathbf{Y}_T\right)$;

    (c) repeat (a) and (b) until the sequence $\left\{a^{(1)},y_{T+1}^{(1)},\ldots,y_{T+h}^{(1)},\ldots,a^{(N_1+N_2)},y_{T+1}^{(N_1+N_2)}\ldots,y_{T+h}^{(N_1+N_2)}\right\}$ is

        simulated;

    (d) keep only last N_2 values of the sequence (in practice, N_2 is usually set to equal N_1).

    In Algorithm 1, because $y_{T+n}^{(i)}$ is generated from (12) in step (a), it always satisfies constraint

(10) or (11). In step (b), density $p\left(a\big|y_{T+1}^{(i)},\ldots,y_{T+h}^{(i)},\mathbf{Y}_T\right)$ at each (ith) loop can be treated as the

---

[8] The reader who is interested in the detail of Gibbs sampler and other Bayesian techniques can
consult Geweke (1994).

posterior density (6) but with the data $\mathbf{Y}_T$ extended to include additional h observations $y_{T+n}^{(i)}$

$(n = 1, 2, \ldots, h)$. When A_0 is exactly identified, one can draw directly from $\pi(a_0)$ because

$A_0^{-1\prime} A_0^{-1}$ has a Wishart distribution (Sims and Zha 1995 and Zha 1997). When A_0 is

overidentified, one cannot draw directly from $\pi(a_0)$ but can use the Metropolis method set out

by Waggoner and Zha (1997).

3.2. *Soft Conditions: A Theory*

In policy projections, the future paths of endogenous variables are unknown. If a set of hard

conditions in (10) turns out to be far away from what will eventually occur in the future,

conditional forecasts can be misleading. For this reason, researchers may be interested

restricting future variables (such as the federal funds rate, M2 growth, or CPI inflation) to lie in

some range, rather fixing their exact path. Such conditions are called soft conditions in this

paper. Soft conditions imply that the set $B(a)$ in (9) has a positive measure in $\mathbf{R}^q$.

When the measure of set $B(a)$ is very small (in other words, when the interval

$(\underline{y}_{T+n}(j), \overline{y}_{T+n}(j))$ is very narrow), Algorithm 1 may be a good approximation by taking the

midpoint of this interval as y_{T+n}^*. When this interval is wide, the application of Algorithm 1

becomes problematic. One naïve approach is to begin with an equally spaced grid in

$(\underline{y}_{T+n}(j), \overline{y}_{T+n}(j))$. Different points on this grid correspond to different r's in the notation of

(11). At each such point, Algorithm 1 can be used to generate the distributions of conditional

forecasts. The distributions are then weighted along the grid so as to attain the overall

distributions of conditional forecasts.

While this naïve approach seems, prima facie, sensible, it is in fact impractical in most cases.

There are at least two fundamental problems associated with this approach. First, there is no

efficient way to specify a fine grid that can well approximate the soft conditions. Second, this approach is circular in the sense that how different points on the grid should be weighted depends on the underlying distribution of conditional forecasts, which is what one intends to simulate in the first place.

When the probability of the set $B(a)$ is non-zero, a straightforward way of simulating the distribution of the conditional forecast is simply to draw a and ε independently and keep the draws that satisfy the set of soft conditions in (9). For each kept draw, compute y_{T+n} according to (5). The empirical distribution can be formed from these simulated samples of y_{T+n}. This method of simulating effective samples of y_{T+n} is, however, inefficient because draws of a from its posterior distribution (conditional on $\mathbf{Y}_T$) are much more expensive than draws of ε from the standard normal distribution. The rest of this subsection lays out a theory that explains why such a method is inefficient and derives analytical results for a better method.

Let us introduce some new notation. Let k_1 be the total number of model parameters $(m^2 p + m)$ and k_2 be the total number of future shocks under consideration ($h \cdot m$). The vector of parameters a is an element of $\mathbf{R}^{k_1}$ and the vector of future shocks ε is an element of $\mathbf{R}^{k_2}$. Let P be the probability measure on $\mathbf{R}^{k_1}$ induced by the posterior distribution and let Q be the probability measure on $\mathbf{R}^{k_2}$ induced by the standard normal distribution. Unconditionally, a and ε are independent, so the joint distribution of (a, ε) induces the product measure $P \times Q$ on $\mathbf{R}^{k_1} \times \mathbf{R}^{k_2}$. Let $\Theta \subseteq \mathbf{R}^{k_1} \times \mathbf{R}^{k_2}$ be the set of all (a, ε) which satisfy constraint (9). For any set measurable set $A \subset \Theta$, the probability of $(a, \varepsilon) \in A$ needs to be estimated. Operationally, one can simply draw a and ε independently from their marginal distributions and keep track of the proportion that lies in A. The strong law of large numbers guarantees that this proportion will

12

converge to the probability that $(a,\varepsilon) \in A$ as the sample size increases. Since it is significantly

more expensive to draw a than ε, the accuracy of simulated probability that $(a,\varepsilon) \in A$ can be

vastly improved by an alternative approach established by the following proposition.

Proposition 4 gives the extent of the efficiency gain from this approach.

Proposition 3. Suppose that n_1 draws of a are sampled from probability distribution P and for

each draw of a, n_2 independent draws of ε are sampled from probability distribution Q. As

$n_1 \to \infty$, the proportion of pairs (a,ε) that lie in A converges to the probability that $(a,\varepsilon) \in A$.

Proof. Let χ_A be the indicator function on A. Define a random variable f on $\mathbf{R}^{k_1} \times (\mathbf{R}^{k_2})^{n_2}$ by

$$f(a,\varepsilon_1,\cdots,\varepsilon_{n_2}) = \frac{1}{n_2} \sum_{i=1}^{n_2} \chi_A(a,\varepsilon_i).$$

By the strong law of large numbers, this sampling scheme converges to the expected value of f

as $n_1 \to \infty$. Since a and the ε_i's are independent,

$$E[f] = \int_{\mathbf{R}^{k_1}} \int_{\mathbf{R}^{k_2}} \cdots \int_{\mathbf{R}^{k_2}} \frac{1}{n_2} \sum_{i=1}^{n_2} \chi_A(a,\varepsilon_i) dQ(\varepsilon_1) \cdots dQ(\varepsilon_n) dP(a)$$

$$= \frac{1}{n_2} \sum_{i=1}^{n_2} \int_{\mathbf{R}^{k_1}} \int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon_i) dQ(\varepsilon_i) dP(a) = E[\chi_A]$$

The above result implies that the proportion of pairs (a,ε) that lie in A converges to the

probability that $(a,\varepsilon) \in A$ by the strong law of large numbers.

Q.E.D.

To show the efficiency gain from the approach established in Proposition 3, suppose it takes 1 unit of time to draw a and $r\,(<1)$ units of time to draw ε. The total amount of computational time is $t = n_1(1+n_2 r)$. The estimate of the probability of $(a,\varepsilon)\in A$ is given by

$$\frac{\sum_{j=1}^{n_1} f\left(a_j,\varepsilon_{j1},\cdots,\varepsilon_{jn_2}\right)}{n_1}, \tag{14}$$

where a_j is the jth independent draw of a and ε_{ji} is the ith independent draw of ε for the jth draw of a.

Given Proposition 3, the purpose is to choose n_2 so that estimate (14) is obtained as accurately as possible for fixed time t. By "accurate", we mean minimum variance. The variance of (14) is $\mathrm{var}[f]/n_1$. But

$$\mathrm{var}[f] = \int_{\mathbf{R}^{k_1}}\int_{\mathbf{R}^{k_2}}\cdots\int_{\mathbf{R}^{k_2}}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\chi_A(a,\varepsilon_i)-E[f]\right)^2 dQ(\varepsilon_1)\cdots dQ(\varepsilon_n)dP(a)$$

$$= \left(\frac{1}{n_2^2}\sum_{i=1}^{n_2}\sum_{j=1}^{n_2}\int_{\mathbf{R}^{k_1}}\int_{\mathbf{R}^{k_2}}\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon_i)\chi_A\left(a,\varepsilon_j\right)dQ(\varepsilon_i)dQ(\varepsilon_j)dP(a)\right)-\left(E[f]\right)^2$$

$$= \frac{1}{n_2}\int_{\mathbf{R}^{k_1}}\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon)^2 dQ(\varepsilon)dP(a) + \frac{n_2-1}{n_2}\int_{\mathbf{R}^{k_1}}\left(\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon)dQ(\varepsilon)\right)^2 dP(a)-\left(E[f]\right)^2$$

$$= \frac{1}{n_2}\left(\int_{\mathbf{R}^{k_1}}\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon)^2 dQ(\varepsilon)dP(a)-\left(E[f]\right)^2\right)$$

$$\qquad + \frac{n_2-1}{n_2}\left(\int_{\mathbf{R}^{k_1}}\left(\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon)dQ(\varepsilon)\right)^2 dP(a)-\left(E[f]\right)^2\right)$$

Denote $E[f]$ by p and let

$$q = \frac{\int_{\mathbf{R}^{k_1}}\left(\int_{\mathbf{R}^{k_2}}\chi_A(a,\varepsilon)dQ(\varepsilon)\right)^2 dP(a)}{p}. \tag{15}$$

Since χ_A is a Bernoulli random variable with the probability of success $E[\chi_A]=E[f]=p$,

14

$$\int_{\mathbf{R}^{k_1}} \int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon)^2 dQ(\varepsilon) dP(a) - \left(E[f]\right)^2 = \mathrm{var}[\chi_A] = p(1-p).$$

Thus, the variance of our estimate is

$$\frac{n_2 r + 1}{t n_2} p\big((1-p) + (n_2 - 1)(q - p)\big). \tag{16}$$

Proposition 4. Let $n_2 \geq 1$ and $p \geq 0$. The value of n_2 that minimizes (16) is

$$n_2 = \max\left\{1, \sqrt{\frac{1}{r}\frac{1-q}{q-p}}\right\}. \tag{17}$$

Proof. By examining the derivative of (16) with respect to n_2, one easily sees that (17)

minimizes (16), so long as both $1-q$ and $q-p$ are non-negative. Since $\int g^2 d\mu \geq \left(\int g d\mu\right)^2$ for

all square integrable functions g, we have that

$$1 - q = 1 - \frac{\int_{\mathbf{R}^{k_1}} \left(\int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon) dQ(\varepsilon)\right)^2 dP(a)}{p} \geq 1 - \frac{\int_{\mathbf{R}^{k_1}} \int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon)^2 dQ(\varepsilon) dP(a)}{p} = 0,$$

and

$$q - p = \frac{\int_{\mathbf{R}^{k_1}} \left(\int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon) dQ(\varepsilon)\right)^2 dP(a)}{p} - p \geq \frac{\left(\int_{\mathbf{R}^{k_1}} \int_{\mathbf{R}^{k_2}} \chi_A(a,\varepsilon) dQ(\varepsilon) dP(a)\right)^2}{p} - p = 0.$$

Q.E.D.

The optimal value of n_2 in Proposition 4 is, in general, a real number, while this simulation

method makes sense only for positive integer values of n_2. In practice, an integer close to (17)

will minimize the lattice problem as long as $n_2 \gg 1$.

To understand how different factors influence the optimal value of n_2, the following

corollary is in order.

Corollary 5. Let $A = A_a \times A_\varepsilon$, $p_a = P(a \in A_a)$, and $p_\varepsilon = P(\varepsilon \in A_\varepsilon)$. Then, the optimal value of n_2 is

$$n_2 = \sqrt{\frac{1}{r} \frac{1 - p_\varepsilon}{p_\varepsilon (1 - p_a)}} \quad . \tag{18}$$

Proof. Note that

$$p = E[f] = P((a, \varepsilon) \in A_a \times A_\varepsilon) = p_a p_\varepsilon. \tag{19}$$

From (15) it follows that

$$
\begin{aligned}
qp &= \int_{\mathbf{R}^{k_1}} \left(\int_{\mathbf{R}^{k_2}} \chi_A(a, \varepsilon) dQ(\varepsilon) \right)^2 dP(a) = \int_{\mathbf{R}^{k_1}} \left(\int_{\mathbf{R}^{k_2}} \chi_{A_a}(a) \chi_{A_\varepsilon}(\varepsilon) dQ(\varepsilon) \right)^2 dP(a) \\
&= \int_{\mathbf{R}^{k_1}} \chi_{A_a}(a)^2 \left(\int_{\mathbf{R}^{k_2}} \chi_{A_\varepsilon}(\varepsilon) dQ(\varepsilon) \right)^2 dP(a) \\
&= p_\varepsilon^2 \int_{\mathbf{R}^{k_1}} \chi_{A_a}(a)^2 dP(a) = p_\varepsilon^2 \int_{\mathbf{R}^{k_1}} \chi_{A_a}(a) dP(a) = p_\varepsilon^2 p_a.
\end{aligned}
$$

Thus,

$$q = p_\varepsilon. \tag{20}$$

Clearly, (18) derives from (17), (19), and (20).

$$\text{Q.E.D.}$$

From Corollary it can be easily seen that n_2 tends to increase when

1) drawing ε is much faster than drawing a;

2) the probability of drawing $a \in A_a$ is high;

3) the probability of drawing $\varepsilon \in A_\varepsilon$ is low.

When n_2 draws of ε are chosen for every draw of a, the variance of the estimate (relative to the probability of $(a, \varepsilon) \in A$), by (16), is

$$\left[\frac{r+1}{t} (1-p) \right] \left\{ \frac{n_2 r + 1}{n_2 (r+1)} \left(\frac{1-q}{1-p} + \frac{n_2 (q-p)}{1-p} \right) \right\}. \tag{21}$$

16

In (21) the term in the square brackets is the variance of the estimate when $n_2 = 1$ is chosen; the term in the curly braces can be interpreted as either the relative decrease in variance or the relative reduction in computing time for a given level of variance. Thus, when the optimal value of n_2 is much larger than one, the improvement factor is measured by

$$\frac{n_2 r + 1}{n_2 (r+1)} \left(\frac{1-q}{1-p} + \frac{n_2 (q-p)}{1-p} \right). \tag{22}$$

If $A = A_a \times A_\varepsilon$, the improvement factor (22) becomes

$$\frac{n_2 r + 1}{n_2 (r+1)} \left(\frac{1-p_\varepsilon}{1-p_a p_\varepsilon} + \frac{n_2 p_\varepsilon (1-p_a)}{1-p_a p_\varepsilon} \right). \tag{23}$$

Once the value of n_2 is chosen, the algorithm, following, can be easily implemented.

Algorithm 2. The simulation of forecast distributions conditional on constraint (8) or (9) involves the three steps:

   (a) draw a according to the posterior density function (6);

   (b) for each draw of a, draw n_2 sets of ε independently from the standard normal

   distribution;

   (c) for each pair (a, ε), compute y_{T+n} ($n = 1, 2, \ldots, h$) according to (5) and keep y_{T+n} in the

   simulated sample only if it satisfies (8).

The advantage of Algorithm 2 is that draws of ε are very inexpensive and thus a greater accuracy can be achieved with less computing time. Clearly, Algorithm 2 can be easily implemented and applies to not only conditional forecasts but unconditional forecasts as well. In contrast to the hard-condition method, draws of a are independent of draws of ε. Thus, in situations where the soft-condition method proves more efficient, it can be used as an approximation to the hard-condition method. For example, if one is interested in restricting the

17

future federal funds rate at, say, 5%, a narrow range around this value (e.g., from 4.50% to 5.50%) will be a good approximation in most cases.

On the other hand, when the set that contains soft conditions in (8) or (9) is very small, it may be more efficient to choose a single point in this set and use the Gibbs sampler technique described in Section 3.1 as an approximate method. Thus, Algorithms 1 and 2 should be viewed as complementary to each other.

4. Examples

This section applies the methods developed in previous sections to the VAR model used in Leeper and Zha (1998).[9] The purpose is to display empirical results for conditional forecasts out of sample using these methods. The discussion concerns only conditions on variables in a simple exactly identified case.[10] By Proposition 1, how A_0 is triangularized does not affect the values of conditional forecasts. Thus, the parameterization of A_0 follows the convention to be lower triangular.

Two cases are considered: one with a flat prior and the other with the informative prior of Sims and Zha (1998). It is shown that the parameter uncertainty can have substantial effects on results and that the Sims and Zha prior helps reduce such effects. In the last part of this section, the efficiency gain from the method in Section 3.2 and Algorithm 2 is shown to give quite reasonable results as compared to Algorithm 1.

The multivariate model used in this section employs monthly data with six macroeconomic variables: the IMF's index of world commodity prices (Pcm), M2, the federal funds rate (FFR), real GDP (GDP), the consumer price index (CPI), and the unemployment rate (U) (see Data

---

[9] See also Zha (1998).

Appendix for a precise description).[11] All variables are logarithmic except the federal funds rate

and the unemployment rate, which are in decimal percentage. The maximum lag length is 13.

Examples focus on the period of the early 1980s. The sample begins at 1959:1 and ends at

1980:12. In 1980 inflation reached its highest peak and then slowed rapidly, making this a

turning point for inflation. Also, a severe recession occurred in 1982 (-2.13% in real GDP

growth), followed by speedy recovery in subsequent years (3.94% and 7.02% real GDP growth

in 1983 and 1984, respectively). Thus, the early 1980s is considered to be a difficult period for

forecasting macroeconomic variables.

   All examples in this section use conditions that concern only actual federal funds rates as

movements in the federal funds rate are often used to explain fluctuations in other

macroeconomic variables. The effects of such conditions on other endogenous variables over a

4-year horizon are examined via conditional forecasts. In Sections 4.1 and 4.2, the federal funds

rate is constrained to follow the path of actual annual average rates in 1981-84. Section 4.3

considers the soft condition that constrains the federal funds rate to be within ±2 percentage

points of actual annual average rates in 1981-84.

4.1. A Hard Condition under the Flat Prior

   Until recently, the macroeconomic literature has studied VAR models with flat priors (e.g.,

Christiano, Eichenbaum, and Evans 1997 and Pagan and Robertston 1998). Thus, this

subsection focuses on the flat prior case. Let us begin with the common practice, suggested in

standard textbooks, of ignoring parameter uncertainty (e.g., Judge et al 1980 and Canies 1988).

The idea of computing the forecast error or distribution by considering only the randomness of

---

[10] The reader who is interested in the application of identified VAR models with conditions
imposed directly on structural shocks can consult Leeper and Zha (1998).

ε, while fixing parameters, dates back at least to Klein (1971) and is now captured by Step (a) in Algorithm 1. Specifically, the values of parameters are fixed at their maximum likelihood estimates, which we denote by $\hat{a}$. The distributions of conditional forecasts are generated by step (a) in Algorithm 1 with $a^{(i-1)} = \hat{a}$. Figure 1 displays such forecasts as of 1980:12 over the next 4 years, conditional on actual annual average funds rates in 1981-84. The solid line represents actual data; the dashed line represents the posterior mean of forecast; the two dashed and dotted lines around the dashed line represent 16th and 84th percentiles so that the bands contain .68 probability.[12] All variables are expressed in annual rates of change in percent, except the federal funds rate and the unemployment rate which are expressed in levels as average percentages.[13]

Both the bands and mean forecasts fail to capture the movements in the actual data.[14] The forecast of Pcm is far from the actual. The M2 forecast misses significantly in 1983-84. The recovery of GDP in 1981 is not detected by the forecast. The 1982 GDP forecast indicates a far more severe recession than the actual outcome. The forecasts of both CPI and U miss the actual data by a large margin in all forecast years. All error bands tend to be quite tight, partly due to the well-known bias of the MLEs toward stationarity[15].

Although the error bands considered here are sufficient for most purposes pertinent to policy analysis, it is sometimes useful to know the entire distribution or likelihood that a particular

---

[11] Monthly GDP is interpolated from quarterly GDP using the procedure described in Leeper, Sims, and Zha (1996).

[12] All simulations done in this paper use 6000 effective draws. And all error bands are constructed to contain .68 probability individually and pointwise. They demarcate the simulated marginal distributions of forecasts at each point of the time horizon, not for the horizon as a whole. For examples of demarcating the joint distributions of forecasts, see Zha (1998).

[13] In policy decisions, what policymakers are usually interested in are not month-to-month variations but rather annual changes in key macroeconomic variables.

[14] Note that the FFR forecast is the same as the actual by construction.

forecast is going to be realized.[16] Figure 2 provides an example with the marginal p.d.f. for the

1982 GDP forecast. The vertical line marks the actual GDP growth rate in 1982, which is near

the tail of the forecast distribution.

It is not uncommon to forecast with fixed values of parameters at their MLEs. When the

sample is small (which is usually the case in empirical macroeconomics), however, the

information about the location of the true parameters is contained in not only the peak of the

likelihood but also the shape of the likelihood as a whole. Indeed, when the shape of the

likelihood is explored via step (b) in Algorithm 1, the 1980:12 forecasts look drastically

different. Figure 3 displays the results by taking explicit account of the parameter uncertainty,

simulated through Algorithm 1. These results look quite reasonable as compared to those in

Figure 1. Most of the actual data lie in or close to the .68 probability bands. The use of error

bands is important because the bands demarcate high and low probability regions in which actual

outcome may or may not occur. It is therefore expected that actual data lie outside the bands at

times (although less frequently). Examples are the M2 forecasts in 1981 and 1984 and the U

forecast in 1984. The actual inflation is close to the lower bound of the error band; as a whole,

the error band captures the downward trend of inflation.

The 1982 recession is detected by the GDP forecast. In contrast to Figure 1, the actual GDP

growth is within or close to the error band. Figure 4 displays the entire p.d.f. of the 1982 GDP

forecast. The height of the p.d.f. predicts that actual GDP growth of –2.13% (marked by the

solid vertical line) is likely to occur, contrary to what is implied by Figure 2.

---

[15] Bias in impulse response functions has been addressed by Kilian (1998).
[16] Diebold, Gunther, and Tay (1997) 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 multi-step, multivariate setup.

In comparison with Figures 1 and 2, Figures 3 and 4 clearly show that the parameter

uncertainty not only widens the error band of the conditional forecast, but more importantly

shifts the distribution of the forecast, which in turn leads to a different mean forecast. This

phenomenon has been given little attention to in the literature and can be revealed only by

accounting for the parameter uncertainty explicitly. Heuristically, one can see why such a

phenomenon is possible.[17] When the shape of the likelihood is such that there is non-trivial

probability for parameters in a nonstationary neighborhood, the randomness of future shocks (ε)

tends to drive forecasts into the region in which forecasts with the stationary values of

parameters are unlikely to fall. As a result, the forecast distribution is likely to shift.[18] This

example implies that ignoring the shape of the likelihood may lead to misleading results (as

shown in Figures 1-4). Thus, it is important that the effects of parameter uncertainty be

examined through the shape of the likelihood before one proceeds to forecast with fixed

parameter values.

4.2. A Hard Condition under Informative Prior

It is well known that the mean forecast with parameters fixed at the MLEs under no prior (as

shown in Figures 1 and 2) tends to be poor for various reasons (Litterman 1986). A recent paper

by Sims and Zha (1998) introduces prior information that aims at eliminating unreasonable and

erratic sampling errors in estimation to improve out-of-sample forecasting.[19] The prior

introduced by Sims and Zha (the SZ prior hereafter) downweights the influence of distant lags.

It also contains components favoring unit roots and cointegration while avoiding the imposition

---

[17] A thorough theoretical analysis on this issue is clearly a subject for future research.

[18] Lutkepohl (1991) uses an asymptotic distribution of parameters as an approximation for the parameter uncertainty. In his framework (pp.85-89), the forecast distribution will never shift. This conclusion is in sharp contrast of our exact small-sample results.

of exact, but possibly false, unit roots and cointegrated relationships. Such a prior is of reference

nature because it likely fits widely held beliefs among applied macroeconomists.[20]

Following the exact notion in Sims and Zha (1998), the tightness of the SZ prior is set as

$\lambda_0 = 0.57$, $\lambda_1 = 0.13$, $\lambda_4 = 0.1$, $\mu_5 = 5$, and $\mu_6 = 5$. When quarterly data are used, the decay rate λ_3

of the lag length is usually set to 1 (e.g., Doan 1992, Miller and Roberds 1991). Thus, for the

monthly model, the lag decay is specified to decline smoothly in an exponential fashion so that

the degree of decay in the thirteenth month matches that for the fifth quarter.

Applying this prior to the same model as in Figure 1, Figure 5 shows the mean forecasts and

error bands with the values of parameters fixed at the MLEs.[21] Clearly, the comparison between

Figure 1 and Figure 5 suggests that the SZ prior effectively reduces the downward bias in the

estimation. The prior also improves the accuracy of the out-of-sample forecasts.[22] The GDP and

Pcm forecasts look quite reasonable. The forecasts of M2, CPI, and U also show notable

improvement over those presented in Figure 1, although the actual data still tend to be far outside

of the error bands for longer forecast horizons. For example, Figure 6 displays the p.d.f. of 1984

U forecast -- one of the worst cases. The distribution assigns almost no probability to the actual

unemployment rate (marked by the solid vertical line). Of course, the early 1980s are a difficult

period for macroeconomic forecasting, particularly when looking up to 4 years ahead.

---

[19] For detailed discussion of the fundamental differences between the Sims and Zha prior and the Litterman prior, consult Sims and Zha (1998).

[20] For classical perspectives, see, for example, Stock and Watson (1996) and Christofferson and Diebold (1997).

[21] Here, the MLEs are the generalized maximum likelihood estimates which are obtained at the peak of the posterior density function.

[22] For comprehensive comparisons between the SZ prior and no prior as well as between the SZ prior and the Litterman prior, see Leeper and Zha (1998). Using the criterion of RMSE, they also document overall improvement in out-of-sample forecasts under the SZ prior.

Note that Figures 5 and 6 treat the parameters as fixed. As shown in Figures 1-4, this assumption may affect the results. To examine such effects under the SZ prior, Algorithm 1 is used to generate forecasts by exploring the shape of the likelihood and the posterior density. Figure 7 presents the results, which are very similar to those in Figure 3 (under the flat prior). Since the SZ prior favors unit roots and cointegration, the similarity between Figure 3 and Figure 7 implies that the shape of the likelihood under the flat prior gives nontrivial probability to the nonstationary region in the parameter space.

The notable differences between Figures 5 and 7, albeit small relative to those between Figures 1 and 3, show up mostly in the widths of the error bands. When parameter uncertainty is taken into account, the 0.68 probability bands in Figure 7 look more reasonable than those in Figure 5. For example, actual inflation in 1984 is now captured by the lower bound of the forecast band for 1984. The distribution of 1984 U forecast, displayed in Figure 8, assigns a greater probability to the area around the actual unemployment rate (marked by the solid vertical line) than does the distribution presented in Figure 6. Thus, parameter uncertainty effects the results, even under the SZ prior.

4.3. A Soft Condition with Informative Prior

Although the traditional approach to conditional forecasting focuses on hard conditions, there is a wide range of applications for soft conditions. A prominent example relates closely to the structural VAR literature in which conditions often revolve around the question of whether monetary policy shocks are expansionary or contractionary. In this case, structural shocks are restricted to a certain range rather than a single value.[23] Another example of interest to policy makers, is the restriction of M2 growth to some target range. The analysis of the effect of a

---

[23] See Leeper and Zha (1998) for details.

target range on the economy can be provided by a model projection conditioned on the range. A

third example is the distribution of an unconditional forecast, which can be simulated more

efficiently with Algorithm 2 than with the approach of Sims and Zha (1998).[24]

Since the purpose of this subsection is to demonstrate how the method of Section 3.2 can be

applied in practical problems and what efficiency gain this method can achieve, the chosen

example is consistent with exercises in Sections 4.1 and 4.2. Specifically, the 1980:12 forecast

uses the SZ prior and imposes the condition that the federal funds rate (FFR) falls in the range of

± 2 percentage points around actual annual average funds rates in 1981-1982. This condition is

imposed in the form of constraint (8) and thus can be equivalently put in the form of constraint

(9). Constraint (9) restricts draws of a and ε to be in the set Θ.

With this soft condition, it is crucial that one obtains a good approximation of the probability

of $(a,\varepsilon) \in A$ for every measurable $A \subset \Theta$. Consider the case where $A = A_a \times A_\varepsilon$, so that

$p_a = p_\varepsilon = \sqrt{p}$ (see (19)). By Corollary, the optimal value of n_2 is

$$n_2 = \frac{1}{\sqrt{p^{1/2}r}} \, . \tag{24}$$

From (23) it can be seen that the improvement (reduction) factor is

$$D = \frac{\left(1+n_2 r\right)\left(1+n_2 p^{1/2}\right)}{n_2(1+r)(1+p^{1/2})} . \tag{25}$$

Since probability p ranges from 0 to 1, (24) implies that the lower bound for the optimal n_2 is

$1/\sqrt{r}$. On other hand, one would like to choose as large n_2 as possible as $p \to 0$. But the cost

of choosing too large a number of n_2 can quickly become overwhelming. This argument can be

---

[24] Clearly, unconditional forecast is simply a special case in Section 3.2.

seen clearly from (25). For any positive p, the improvement factor becomes a "deteriorating"

factor because D goes to infinity as $n_2 \to \infty$.[25]

Then what should the optimal n_2 be for all $p \in [0,1]$? For the 6-variable model here with the

48-month (4-year) forecast horizon, r -- the ratio of computing time of drawing ε to that of

drawing a -- is about 1/95. Thus, the lower bound for optimal n_2 is about 10. Figure 9 plots

the improvement factor under different p's. Except for $p < 0.1$, D does not improve much

further for $n_2 > 10$ and in fact quickly deteriorates (turns up from the trough) when n_2 is greater

than 20. In principle, there does not exist an optimal value of n_2 for all p. But in practice the

lower bound ($n_2 = 10$) is close enough to be the optimal number for almost all p.[26]

With $n_2 = 10$ and under the soft condition that FFR is within ± 2 percentage points around

actual average funds rates in 1981-84, Algorithm 2 is used to simulate forecasts as of 1980:12.

The forecast distribution is simulated with $n_1 = 40,000$ and $n_2 = 10$. Among these Monte Carlo

draws, about $6,000 \times 10$ effective draws satisfy the soft condition on the federal funds rate.

Computing time is about 3.6 hours on a 266 Pentium II PC.[27] Suppose that only one draw of ε

is taken for each draw of a. To achieve the same accuracy of the simulated distribution, Figure

9 implies that at least 6.12 hours would be needed if one cares only about the rough shape of the

forecast distribution ($p \approx 1$) and about 36 hours would be required if one is concerned with small

probability events ($p \approx 0$).

---

[25] Recall that D must be less than 1 to reduce the variance of the estimate.

[26] Such a good approximation of the lower bound to the optimal value is true even for larger
VAR models in which r is much smaller than 1/95.

[27] In contrast, the computing time for the results in Figure 7 is about 13 hours. The demanding
part is the computation of a single value decomposition of the large covariance matrix in (13) for
each loop. In our example, the size of the covariance matrix is $mh = 288$.

Figure 10 reports the simulated results with error bands attached. Since all these bands contain 0.68 probability, the actual data, even for constrained FFR, may lie outside the band. Clearly, the results in Figure 10 are quite close to those under the hard condition (Figure 7), although the bands are slightly wider and shift somewhat. This example shows that the soft-condition method (Section 3.2) can be a reasonable approximation to the hard-condition method (Section 3.1), although these two methods are designed to address different questions.

5. Conclusion

Conditional forecasts are designed to answer many practical questions that cannot be answered by unconditional forecasts. Policymakers as well as analysts often have a priori knowledge outside the model about the future paths of variables or the future stance of policy and would like to evaluate the effects on the forecast when such knowledge is imposed on the model. Policymakers might want to know, for example, the effects of contractionary monetary policy in the near future on the overall economy. Policy analysts might be interested in knowing how the forecast changes if the federal funds rate or CPI inflation follows a certain path in the future. If analysts are not sure of particular paths of future variables but have some idea of the range in which variables may fluctuate in the future (e.g., the funds rate will not be higher than 6% in the next year or inflation will be no less than 1% but no more than 3% in the next four years), they would like to analyse the effect of such a range condition on the forecast.

To address these practical issues, this paper broadens the class of conditional forecasts in the existing literature and has developed tools that enable one to tackle these issues in a multivariate framework. The methods developed here provide theoretical underpinnings for exact small-sample properties regarding parameter uncertainty and can be feasibly implemented to compute probability distributions or error bands associated with conditional forecasts. Concrete examples

27

are used to show the importance of accounting for parameter uncertainty in out-of-sample

conditional forecasts. It is hoped that the methods will help applied researchers analyze the

effects on macroeconomic forecasts under different conditions.

Data Appendix

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

Pcm: International Monetary Fund's Index of world commodity prices. Source: International Financial 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.

GDP: Real GDP, seasonally adjusted, billions of chain 1992 dollars. Monthly real GDP is interpolated using the procedure described in Leeper, Sims, and Zha (1996). Source: Bureau of Economic Analysis, the Department of Commerce (BEA).

CPI: Consumer price index for urban consumers (CPI-U), seasonally adjusted. Source: BEA.

U: Civilian unemployment rate (ages 16 and over), seasonally adjusted. Source: Bureau of Labor Statistics.

References

Blinder, Alan S., 1997. "What Central Bankers Could Learn from Academics—and Vice Versa," *Journal of Economic Perspectives 11*, 3-19.

Caines, Peter E., 1988. *Linear Stochastic Systems*, John Wiley & Sons, New York.

Christiano, Lawrence J., Martin Eichenbaum, Charles Evans, 1997. "Monetary Policy Shocks: What Have We Learned and To What End?" in *Handbook of Macroeconomics* (Eds. John Taylor and Michael Woodford, forthcoming).

Christofferson, Peter F. and Francis X. Diebold, 1997. "Cointegration and Long-Horizon Forecasting," *Journal of Business and Economic Statistics* (forthcoming).

Diebold, Francis X., 1998. "The Past, Present, and Future of Macroeconomic Forecasting," *Journal of Economic Perspective 12* (Spring), 175-192.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay, 1997. "Evaluating Density Forecasts," *NBER* Technical Working Paper 215, 1-27.

Doan, Thomas A., 1992. *RATS User's Manual Version 4*, Estima.

Doan, Thomas A., Robert B. Litterman, and Christopher A. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Review 3*, 1-100.

Geweke, John, 1994. "Monte Carlo Simulation and Numerical Integration," in *Handbook of Computational Economics* (Eds. Hans Amman, David Kendrick, and John Rust, forthcoming), North-Holland.

Intriligator, Michael D., Ronald G. Bodkin and Cheng Hsiao, 1996. *Econometric Models, Techniques, and Applications*, Second Edition, Prentice-Hall International, Inc. (Upper Saddle River, New Jersey).

Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl and Tsoung-Chao Lee, 1980. *The Theory and Practice of Econometrics*, John Wiley and Sons (New York).

Kilian, Lutz, 1998. "Small-Sample Confidence Intervals for Impulse Response Functions," *The Review of Economics and Statistics*, 218-230.

Klein, Lawrence R., 1971. *An Essay on the Theory of Economic Prediction*, Markham Publishing Company (Chicago).

Kohn, Donald L., 1995. "Comment on 'Inflation Indicators and Inflation Policy' by Cecchetti," *NBER Macro Annual 1995*, 227-35.

Leeper, Eric M., Christopher A. Sims and Tao Zha, 1996. "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity 2*, 1-63.

Leeper, Eric M. and Tao Zha, 1998. "Unifying Policy Analysis and Forecasting: The Cowels Commission Revisited," *manuscript*.

Litterman, Robert B., 1986. "Forecasting With Bayesian Vector Autoregressions – Five Years of Experience," *Journal of Business and Economic Statistics 4*, 25-38.

Lutkepohl, Helmut, 1991. *Introduction to Multiple Time Series Analysis*, Springer-Verlag (Berlin, Germany).

Miller, Preston J. and Roberds, William, 1991. "The Quantitative Significance of the Lucas Critique," *Journal of Business and Economic Statistics 9 (4)*, 361-387.

Pagan, Adrian R. and John C. Robertson, 1998. "Structural Models of the Liquidity Effect," *Review of Economics and Statistics*

Roberds, William and Charles H. Whiteman, 1992. "Monetary Aggregates as Monetary Targets: A Statistical Investigation," *Journal of Money, Credit, and Banking 24(2)*, 564-78.

Sims, Christopher A., 1982. "Policy Analysis with Econometric Models," *Brookings Papers on Economic Activity 1*, 107-64.

Sims, Christopher A. and Tao Zha, 1998. "Error Bands for Impulse Responses," *Econometrica*, (forthcoming).

_____, 1998. "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review 39*, 949-968.

Stock, James H. and Mark W. Watson, 1996. "Confidence Sets in Regressions with Highly Serially Correlated Regressors," *manuscript*.

Waggoner, Daniel F. and Tao Zha, 1997. "Normalization, Probability Distribution, and Impulse Responses," Federal Reserve Bank of Atlanta *Working Paper 97-11*.

Zha, Tao, 1997. "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics* (forthcoming).

_____, 1998. "A Dynamic Multivariate Model for Use in Formulating Policy," Federal Reserve Bank of Atlanta *Economic Review* (First Quarter), 16-29.

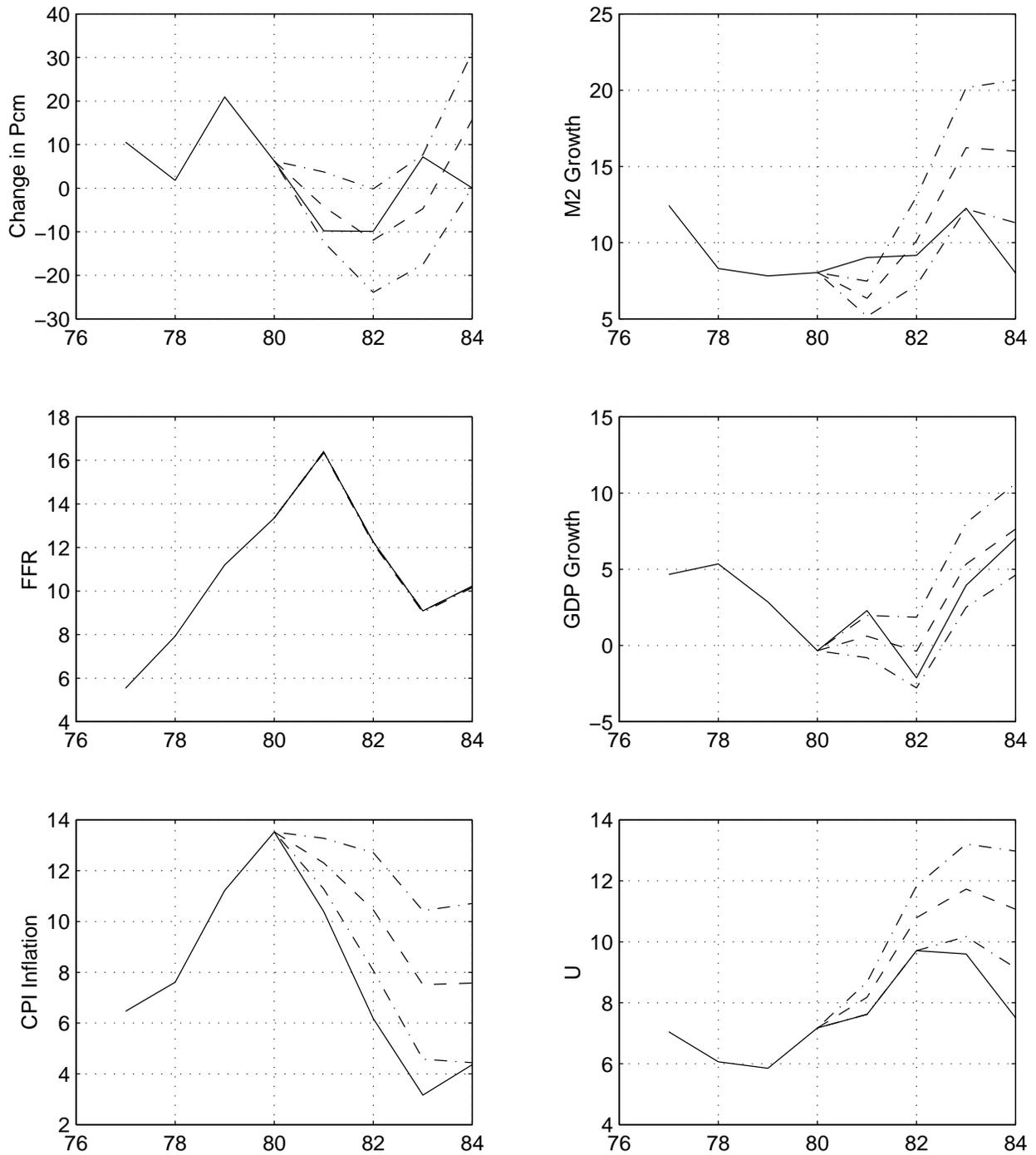**Figure 1.** 1980:12 Conditional Forecasts with MLEs under Flat Prior.
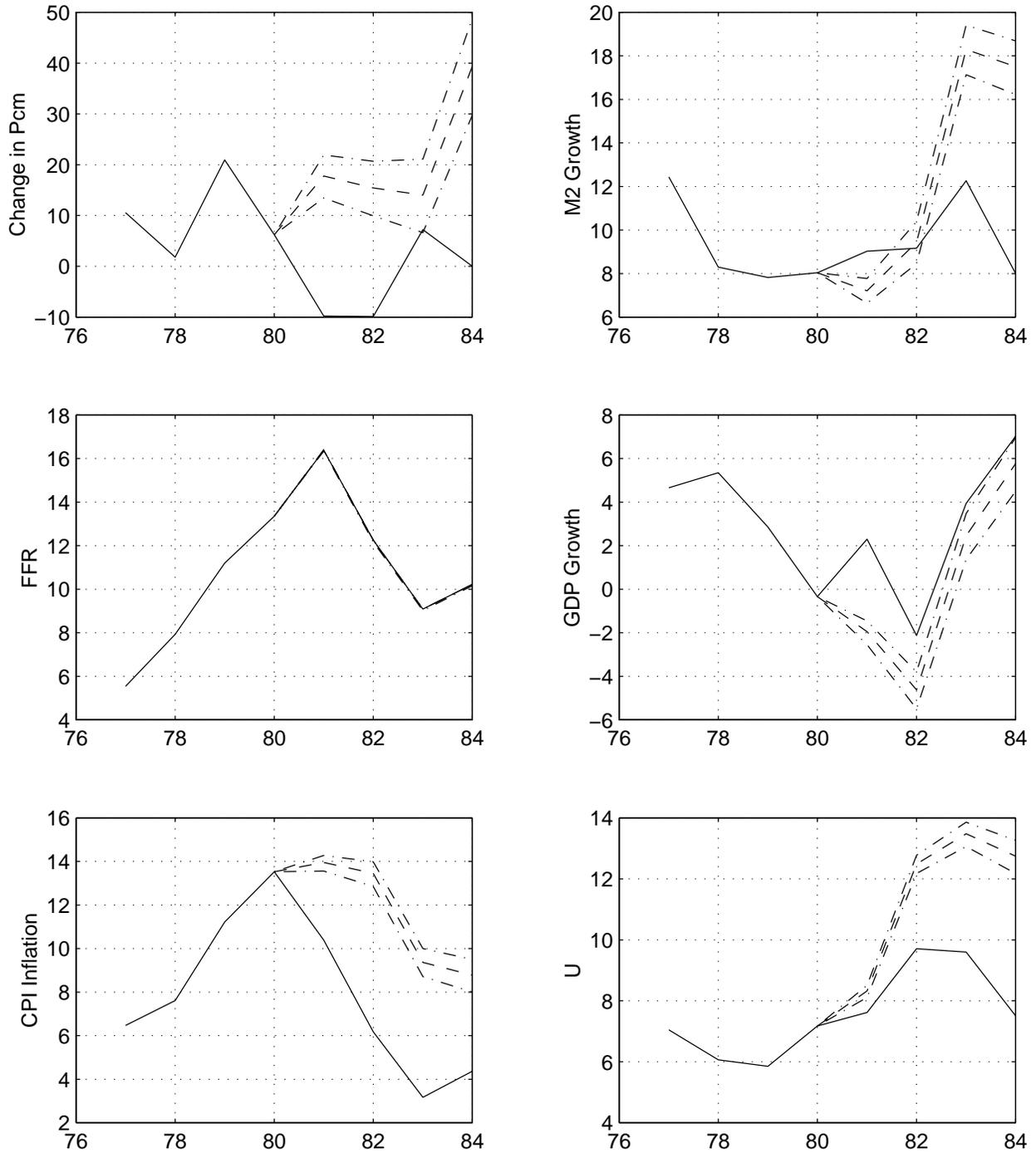
Solid line: actual; dashed line: posterior mean of forecast; dashed and dotted line: bounds of .68 probability bands.

Figure 2. Marginal p.d.f. of the 1982 GDP Forecast with MLEs under Flat Prior.
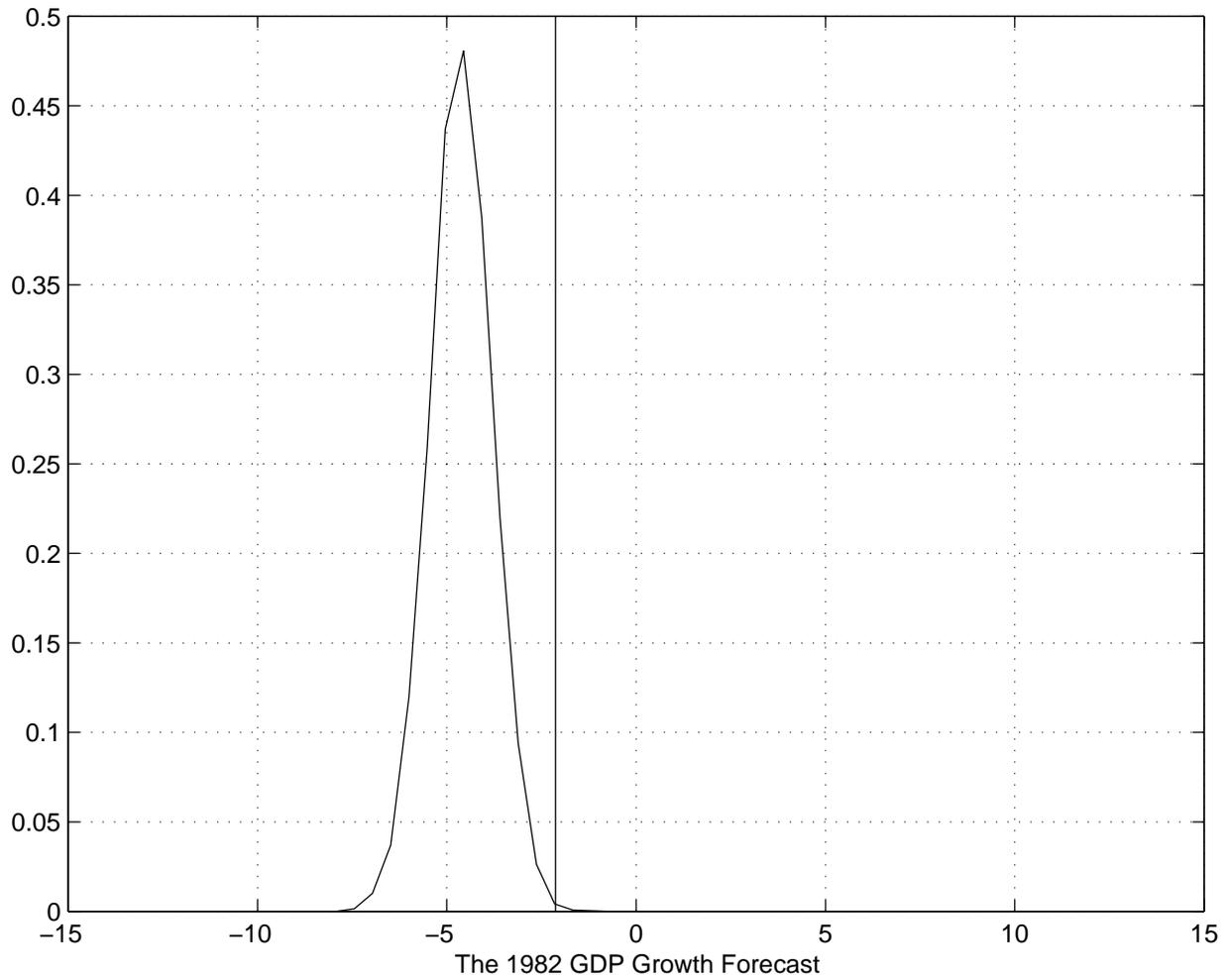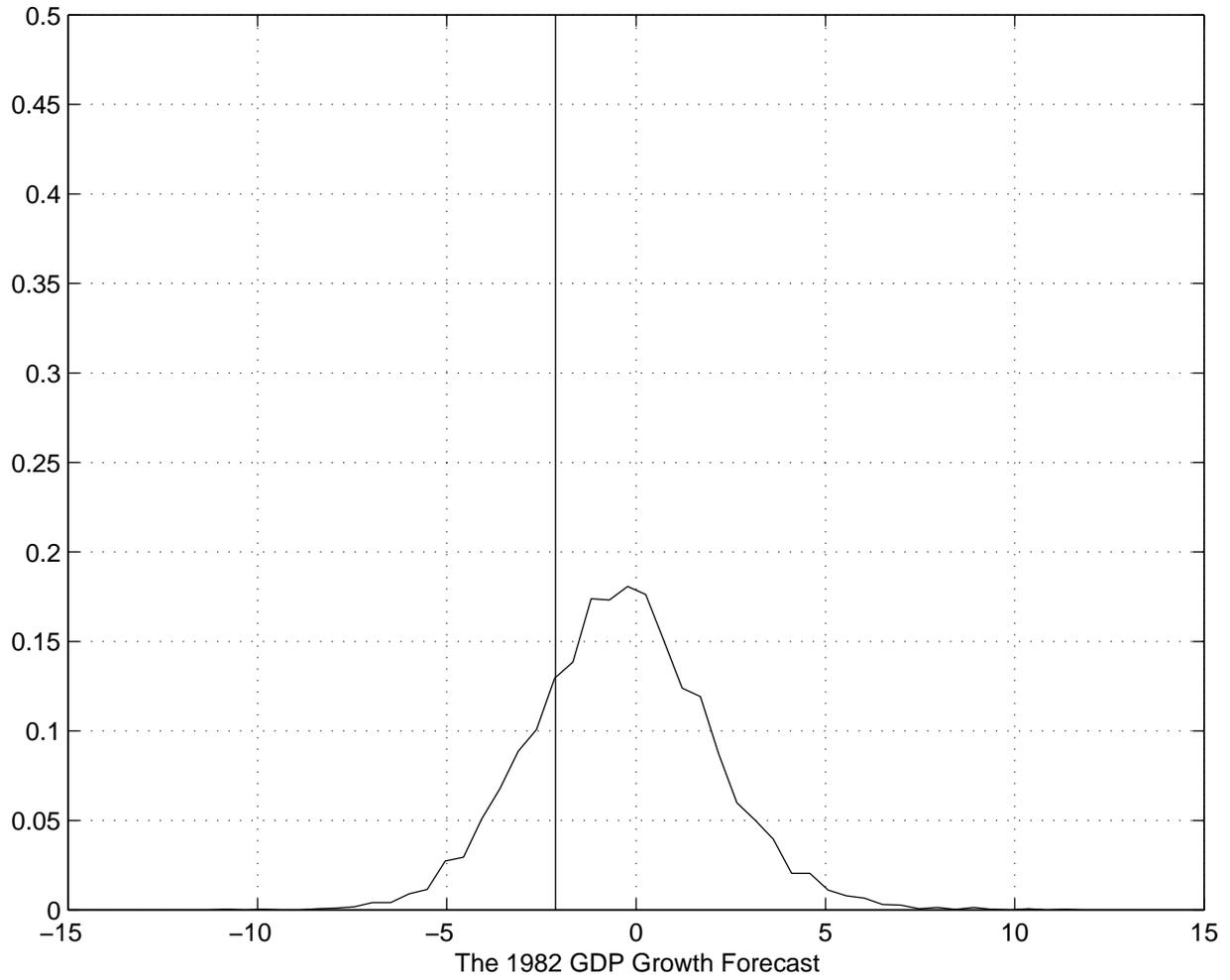

Solid vertical line: actual.

Figure 3. 1980:12 Conditional Forecasts with Parameter Uncertainty under Flat Prior.
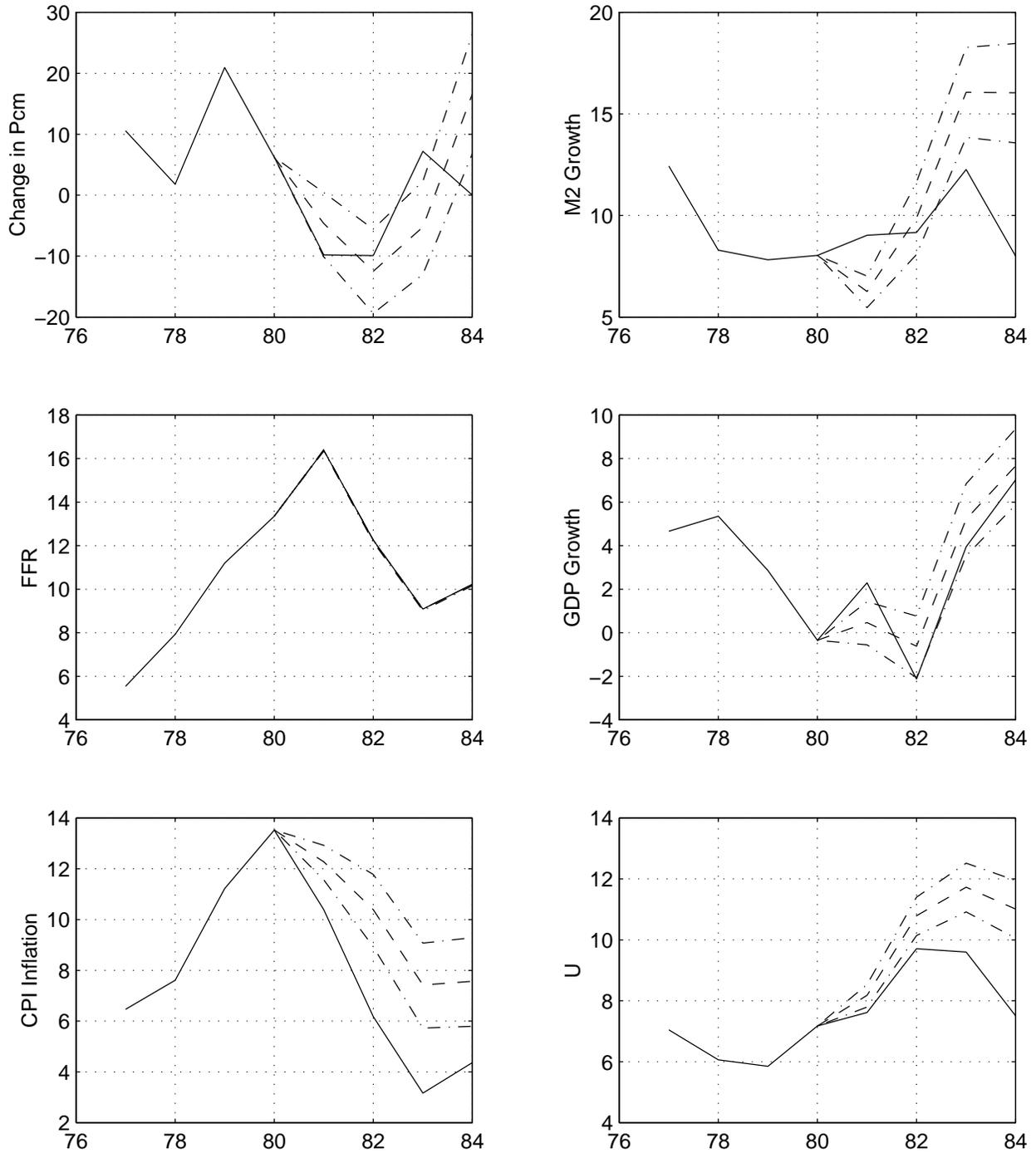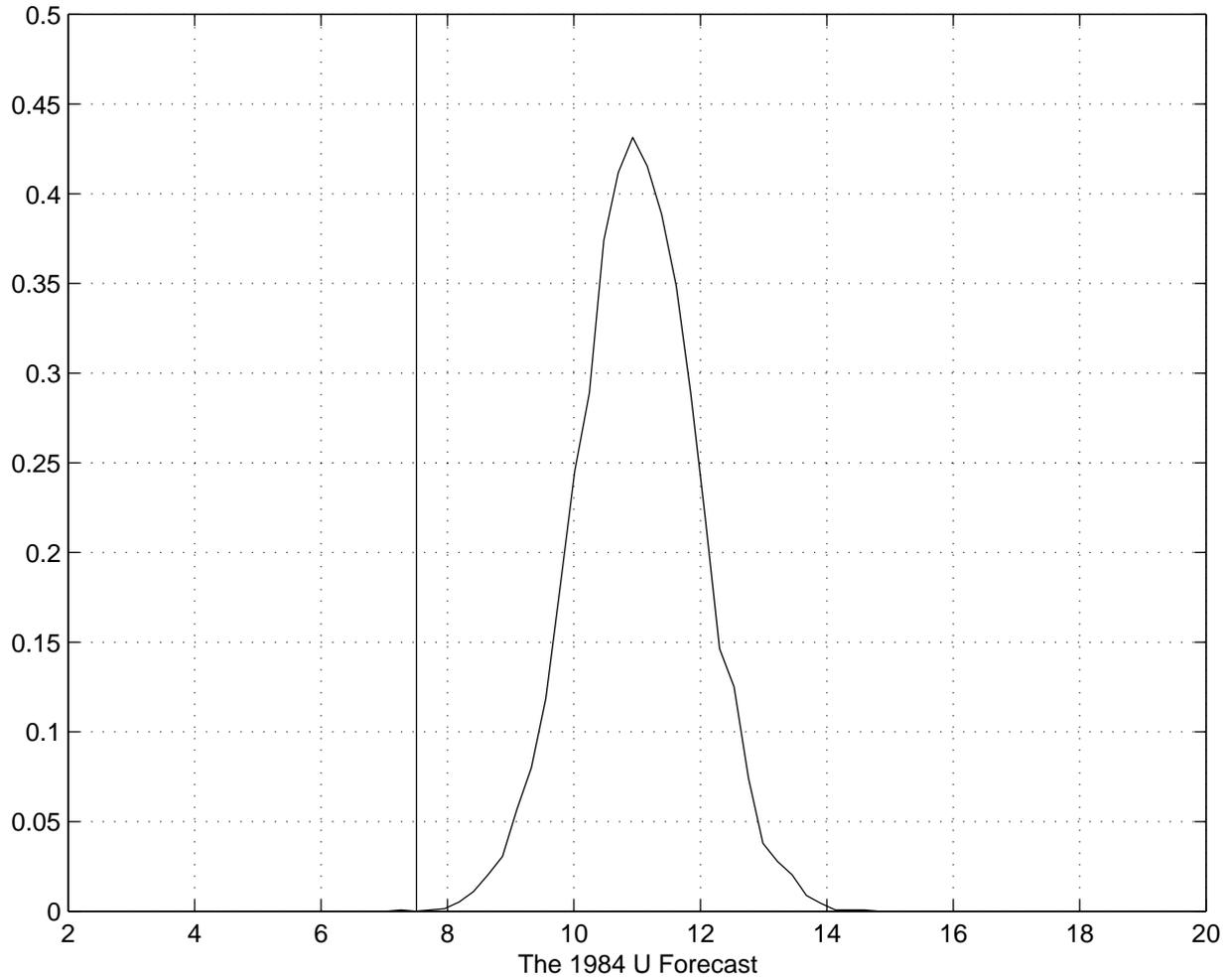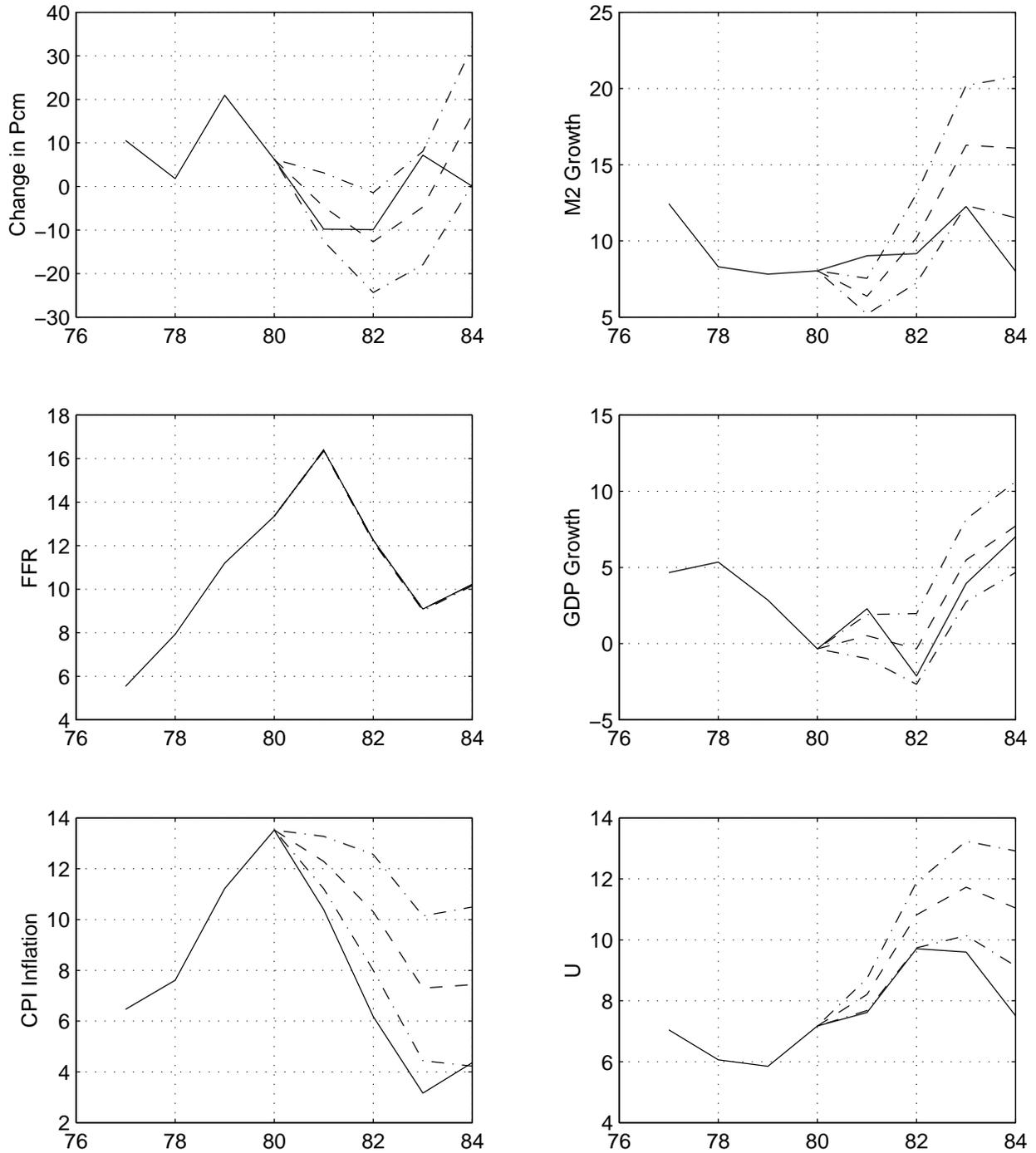

Solid line: actual; dashed line: posterior mean of forecast; dashed and dotted line: bounds of .68 probability bands.

Figure 4. Marginal p.d.f. of the 1982 GDP Forecast with Parameter Uncertainty under Flat Prior.

Solid vertical line: actual.

Figure 5. 1980:12 Conditional Forecasts with MLEs under SZ Prior.

Solid line: actual; dashed line: posterior mean of forecast; dashed and dotted line: bounds of .68 probability bands.

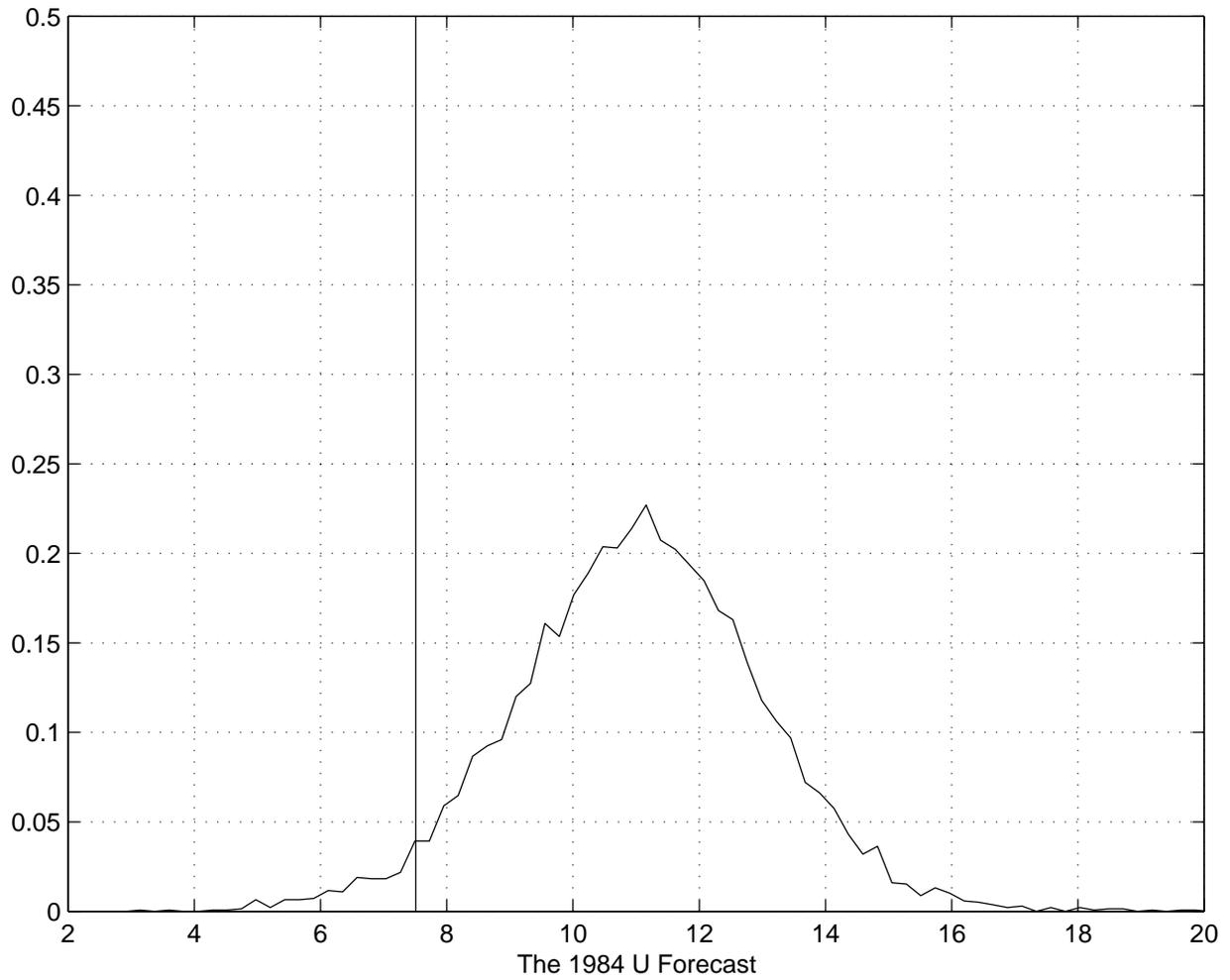**Figure 6.** Marginal p.d.f. of the 1984 U Forecast with MLES under SZ Prior.

Solid vertical line: actual.

Figure 7. 1980:12 Conditional Forecasts with Parameter Uncertainty under SZ Prior.

Solid line: actual; dashed line: posterior mean of forecast; dashed and dotted line: bounds of .68 probability bands.

Figure 8. Marginal p.d.f. of the 1984 U Forecast with Parameter Uncertainty under SZ Prior.

Solid vertical line: actual

Figure 9. Reduction in Variance of Estimate with n_2 under $p = 0$, 0.5, and 1.
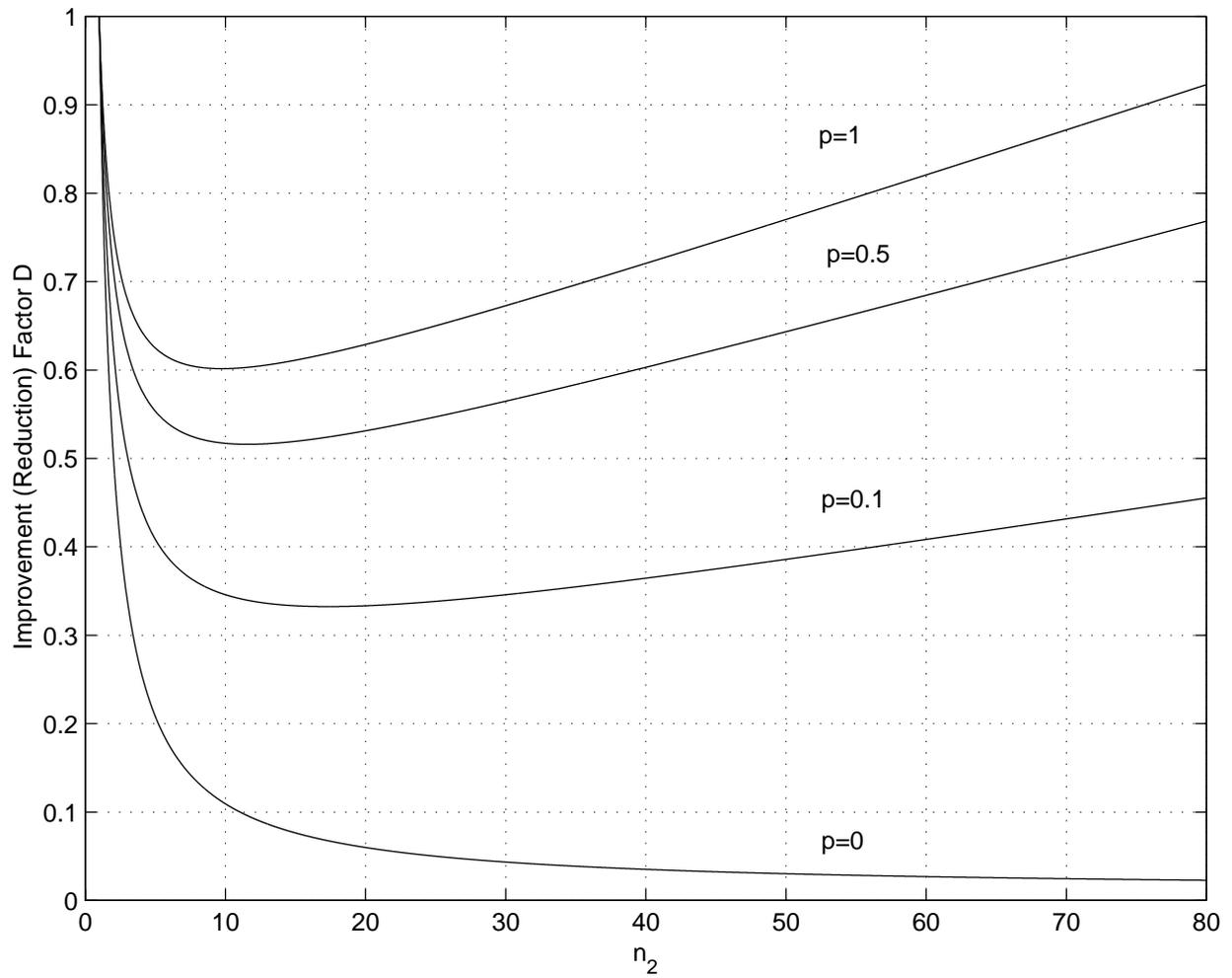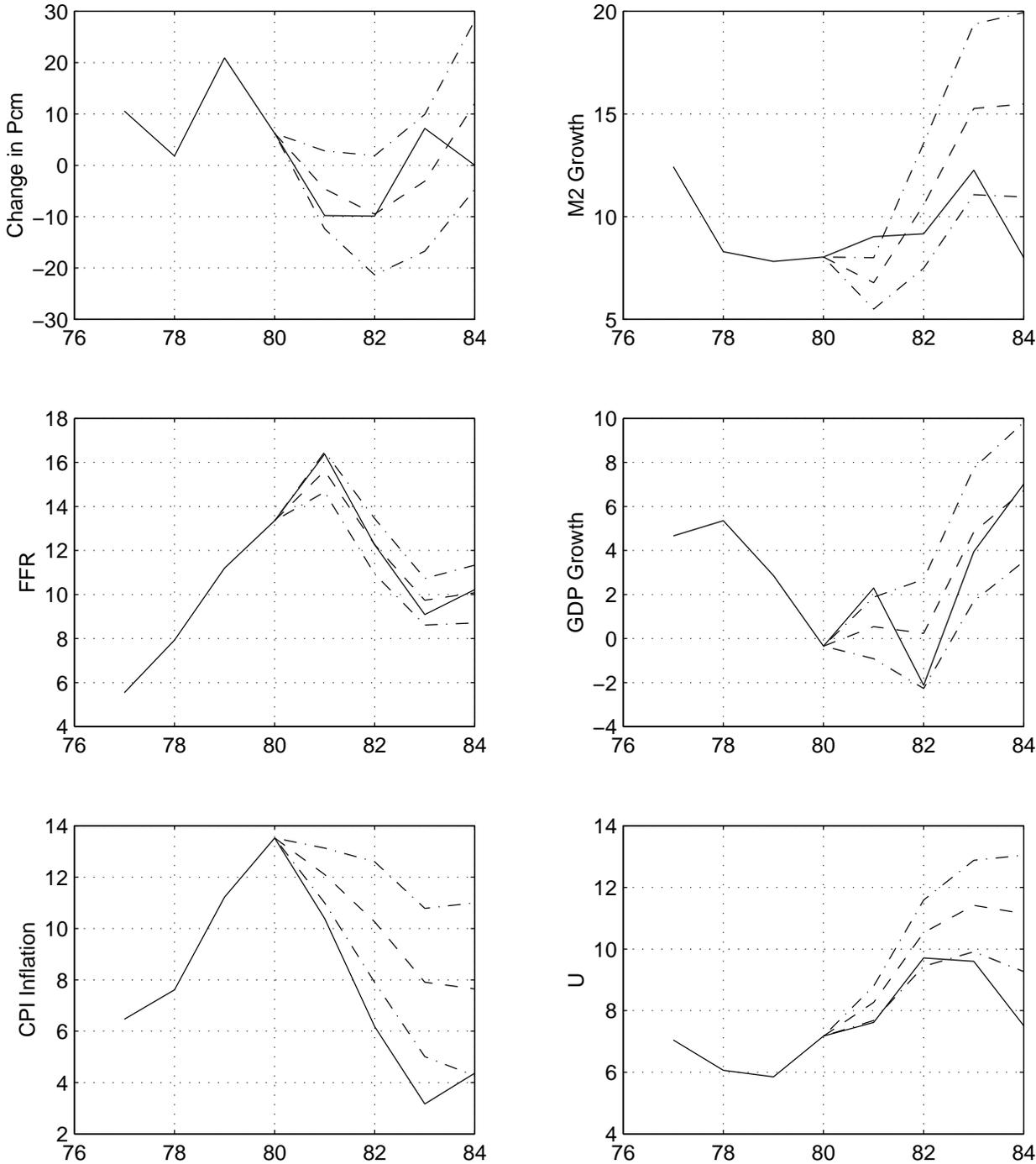
Figure 10. 1980:12 Conditional Forecasts with Soft Condition and Parameter Uncertainty under SZ Prior.

Solid line: actual; dashed line: posterior mean of forecast; dashed and dotted line: bounds of .68 probability bands.