Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System[*]

Timothy Cogley
University of California, Davis

Sergei Morozov
Stanford University

Thomas J. Sargent
New York University and Hoover Institution

Revised: October 2003

Abstract

We estimate a Bayesian vector autoregression for the U.K. with drifting coefficients and stochastic volatilities. We use it to characterize posterior densities for several objects that are useful for designing and evaluating monetary policy, including local approximations to the mean, persistence, and volatility of inflation. We present diverse sources of uncertainty that impinge on the posterior predictive density for inflation, including model uncertainty, policy drift, structural shifts and other shocks. We use a recently developed minimum entropy method to bring outside information to bear on inflation forecasts. We compare our predictive densities with the Bank of England's fan charts.

1   Introduction

The inflation-targeting regime that has prevailed in U.K. since 1992 has caused policy makers to anticipate inflation. At regular meetings with its forecast team, the Bank of England's Monetary Policy Committee (MPC) discusses prospects for

inflation. After the meetings, the forecast team transforms the MPC's discussion into a projected distribution for future inflation in the absence of a policy change. The MPC and its forecast team devote substantial time and effort to preparing *fan charts* that express the Bank's subjective assessment of medium-term inflationary pressures in terms of a subjective measure of uncertainty. The fan charts reveal the MPC's subjective probability distribution for inflation, assuming that the Bank's policy instrument, a short-term nominal interest rate, is held constant. The MPC uses the fan chart to guide and justify its decisions to adjust the short term nominal interest rate to propel inflation toward the target.

The MPC intends the fan charts to describe several sources of uncertainty, including model uncertainty, that qualify its forecasts of GDP growth and inflation. Indeed, the Bank assesses risks by using multiple models and listening to a variety of views. The MPC begins by identifying important shocks that might affect the inflation forecast. For each perceived shock, the MPC forms a view of its size and consequences and considers how that view might be wrong. It then calibrates the degree of uncertainty and the balance of risks by examining alternative models to assess what the consequences might be if the central view is mistaken. Eventually, the MPC makes a judgment about whether the risks are skewed and by how much, and whether uncertainty is more or less than in the past.

This paper proposes another way to prepare fan charts. By applying modern Bayesian methods to a vector autoregression with a stochastic time-varying error structure, we construct and estimate a flexible statistical model that acknowledges diverse sources of uncertainty that include a form of model uncertainty. We construct fan charts and other policy informative statistics for U.K. macroeconomic data, and we compare our fan charts with those of the MPC.

Our objective is to articulate a formal and explicit probabilistic method for constructing fan charts. Our first step is to construct a posterior predictive density for a Bayesian vector autoregression like the one developed by Cogley and Sargent (2002). We can view the predictive density for inflation from our Bayesian VAR either as an end in itself or as an 'agnostic' benchmark density that is to be 'twisted' by bringing to bear additional sources of information and judgments. We apply a method developed by Robertson, Tallman, and Whiteman (2002) that transforms one predictive density into another by constraining the twisted density to satisfy some constraints that contain the outside information. This way of incorporating outside information complements our BVAR fan chart in a way that seems to us to be congenial to Central Bank policy procedures.

Our proposal has several merits relative to others currently in use for constructing fan charts. For one, our Bayesian VAR has adaptive elements that make it attractive for forecasting in an evolving monetary system such as that of the U.K. Within the last four decades, there was the Bretton Woods system, then the breakdown and float, various incarnations of fixed exchange rate rules, the decision to opt out of the fixed

2

rate system after the currency crisis of 1992, and the adoption of inflation targeting, and finally the Treasury's gift of independence to the Bank. Who knows what the future will bring after an eventual referendum on joining the European Monetary Union? Each such transition must have witnessed a period of adaptation in which the Bank learned how inflation was likely to behave in the new regime. Our Bayesian VAR allows for drift in conditional mean and conditional variance parameters and lets the predictive density adapt to a changing environment.

Our proposals also explicitly acknowledge more sources of forecast uncertainty than do other procedures currently in use. Our model has three sources of uncertainty. As in conventional models, there are shocks to the VAR that will occur in the future, as well as uncertainty about parameters describing the state of the economy at the forecast date. An additional source of uncertainty enters because some of the model's parameters will in the future drift away from their current values. This feature is important for representing uncertainty about how closely the future may resemble the present.[1] Some methods for constructing fan charts focus exclusively on VAR innovations, while others combine them with static parameter uncertainty, but most abstract from parameter drift. One contribution of this paper is to show how to construct Bayesian prediction intervals when the monetary system is evolving. We quantify the importance of this source of forecast uncertainty and compare it with the others.

We construct fan charts for U.K. inflation conditioned on data through 2002.Q4. We show the BVAR predictive density and decompose the various sources of uncertainty that contribute to it. We also illustrate how outside information alters the BVAR predictive density, and we compare our predictive densities with the Bank of England fan chart for that quarter.

2 Posterior Predictive Densities

We begin by setting down some notation. Let y_t represent a vector of observed time series at date t, and let its history be represented by

$$Y^T = [y_1', ..., y_T']'. \tag{1}$$

We want to make predictions F periods into the future. This involves characterizing probabilities associated with potential future paths, which are denoted

$$Y^{T+1,T+F} = [y_{T+1}', ..., y_{T+F}']'. \tag{2}$$

---

[1] Examples that motive the relevance of this feature include whether or when the U.K. will join the EMU, whether there is a new economy in the U.S., or how the Federal Reserve will behave after Chairman Greenspan's retirement.

The model depends on two sets of parameters, some that drift over time and others that are static. The drifting parameters are denoted Θ_t, and their history and potential future paths are represented by

$$\Theta^T = [\Theta_1', ..., \Theta_T']', \tag{3}$$
$$\Theta^{T+1,T+F} = [\Theta_{T+1}', ..., \Theta_{T+F}']',$$

respectively. Static parameters are denoted ψ.

The paper shows how to simulate two posterior densities, one corresponding to model-based forecasts from a Bayesian vector autoregression and another that transforms the first to reflect information from outside the model. The first subsection defines the BVAR predictive density, and the second reviews the method of Robertson, et. al. for including additional information.

2.1 The BVAR Predictive Density

BVAR fan charts are constructed from the posterior predictive density, which is the joint probability density over future paths of the data, conditioned on priors and the history of observables. This density can be expressed as

$$p(Y^{T+1,T+F}|Y^T) = \iint p(Y^{T+1,T+F}, \Theta^{T+F}, \psi|Y^T) d\Theta^{T+F} d\psi, \tag{4}$$

where $p(Y^{T+1,T+F}, \Theta^{T+F}, \psi|Y^T)$ is the joint posterior density for VAR parameters and future observables.

To interpret this expression, it is convenient to factor the integrand as

$$\begin{aligned} p(Y^{T+1,T+F}, \Theta^{T+F}, \psi|Y^T) &= p(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T) \\ &\times p(\Theta^T, \psi|Y^T) \\ &\times p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T). \end{aligned} \tag{5}$$

Each of the factors on the right hand side accounts for one of the sources of uncertainty listed in the introduction. The first term, $p(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T)$, represents the posterior density for future observables, treating model parameters as if they were known with certainty. This term reflects the influence of future shocks to the VAR.

The second and third terms in (5) account for parameter uncertainty. The second term, $p(\Theta^T, \psi|Y^T)$, measures parameter uncertainty within the sample. This is the joint posterior density for VAR parameters given priors and data through date T. Among other things, this includes uncertainty about Θ_T, the terminal value of the drifting parameters, as well as uncertainty about the static hyperparameters that govern the rate at which Θ_t drifts.

The third term, $p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$, reflects how future parameters may drift away from Θ_T as the economy continues to evolve. Incorporating this last source of

4

uncertainty is a novel feature of our proposal. How much this matters for the short- and medium-horizon forecasts in which the Bank is interested is a question that we address below.

2.2 Exploiting Outside Information to Sharpen BVAR Forecasts

Although a BVAR fan chart might be useful as a benchmark or starting point for policy discussions, it is unlikely to be adequate by itself. The BVAR is designed for flexibility, to fit complex patterns in the historical data and to adapt well to changing circumstances. Such a model is unlikely to be a good forecasting tool because it is too heavily parameterized. Additional information of some sort, whether from priors, observations of other variables, or forecasts from other models, is needed to augment the benchmark model. This can be brought in at a second stage using the relative entropy method of Robertson, Tallman, and Whiteman (2002).[2]

The basic idea is to take a random sample from the BVAR predictive density[3] and to transform it into a random sample from another predictive density, say $p^*(Y^{T+1,T+F}|Y^T)$, that embodies the outside information. Robertson, et. al. assume that the new information can be expressed as a vector of moment conditions,

$$E\, g(Y^{T+1,T+F}) = \bar{g}, \tag{6}$$

where the expectation is taken with respect to $p^*(\cdot)$. If it happens that

$$\int g(Y^{T+1,T+F}) p(Y^{T+1,T+F}|Y^T) dY^{T+1,T+F} = \bar{g}, \tag{7}$$

then the BVAR predictive density is fully consistent with the outside information, and no transformation is needed. But this rarely occurs. The BVAR probability weights usually have to be altered to satisfy (6). Since this involves a 'distortion' of the original density, we want to alter the weights in a minimal fashion. That is, the new density should be as close as possible to the original.

Robertson, et. al. use the Kullback-Leibler information criterion (KLIC) to measure how close the two densities are. The KLIC is defined as

$$K(p, p^*) = \int \log\left[\frac{p^*(Y^{T+1,T+F}|Y^T)}{p(Y^{T+1,T+F}|Y^T)}\right] p^*(Y^{T+1,T+F}|Y^T) dY^{T+1,T+F}. \tag{8}$$

New probability weights are chosen by minimizing $K(p, p^*)$ subject to constraints that (6) is satisfied and that the weights are probabilities. Their procedure works as follows.

---

[2] They credit Stutzer (1996) and Kitamura and Stutzer (1997) for their inspiration.

[3] Any predictive density will do, but we use the BVAR density as a starting point.

Let $(Y_i^{T+1,T+F}, \pi_i)$ represent a draw from the BVAR predictive density, where π_i is a probability weight from $p(\cdot)$. Our algorithm produces an evenly weighted sample, so $\pi_i = 1/N$, but this need not be the case. We want to calculate a new weight π_i^* so that the re-weighted sample satisfies the Monte Carlo approximation to (6),

$$\sum_{i=1}^{N} \pi_i^* g(Y_i^{T+1,T+F}) = \bar{g}. \tag{9}$$

Since the new weights are probabilities, they must also satisfy

$$\pi_i^* \geq 0, \qquad \sum_{i=1}^{N} \pi_i^* = 1. \tag{10}$$

Finally, since we want $p^*(\cdot)$ to be as close as possible to $p(\cdot)$, π_i^* is chosen to minimize the Monte Carlo approximation to $K(p, p^*)$,

$$\sum_{i=1}^{N} \pi_i^* \log\left(\frac{\pi_i^*}{\pi_i}\right). \tag{11}$$

The solution sets

$$\pi_i^* = \frac{\pi_i \exp\left[\gamma' g(Y_i^{T+1,T+F})\right]}{\sum_{i=1}^{N} \pi_i \exp\left[\gamma' g(Y_i^{T+1,T+F})\right]}, \tag{12}$$

where γ is a vector of the Lagrange multipliers associated with (6). The multipliers can be found by solving a minimization problem,

$$\min_{\gamma} \sum_{i=1}^{N} \pi_i \exp\left\{\gamma'\left[g(Y_i^{T+1,T+F}) - \bar{g}\right]\right\}. \tag{13}$$

Armed with the new weights π_i^*, one can use importance sampling techniques to resample from the original BVAR distribution. This just involves sampling $Y_i^{T+1,T+F}$ with weight π_i^* instead of weight $\pi_i = 1/N$. This can be done via the multinomial resampling algorithm of Gordon, Salmond, and Smith (1993), which we describe in an appendix.[4]

One reason why our Bayesian VAR is useful for producing benchmark forecasts is that it encompasses more sources of uncertainty than many other models. This means its tails are likely to be heavier than those of a 'better informed' proposal density. Heavy tails are attractive for a proposal density because they ensure that the tails of $p^*(\cdot)$ are well represented in the sample from $p(\cdot)$. This enhances the efficiency and numerical accuracy of importance sampling (Geweke 1989).[5]

---

[4]Their resampling scheme is a component of a more complex particle filter. Our problem involves only the 'selection' step of their algorithm.

[5]A number of participants at the conference remarked that the tails of $p(\cdot)$ might be too fat. Sylvia Kaufmann, on the other hand, warned that they might be too thin, at least in one direction.

An interesting by-product of these calculations are measures of how much the original density is altered. Notice that $K(p, p^*)$ is zero if $p^*(\cdot)$ and $p(\cdot)$ coincide. Therefore, small values of the empirical KLIC signify agreement between the BVAR and outside information, while large values signify disagreement. In the latter case, the Lagrange multipliers γ are useful for diagnosing which elements of $g(\cdot)$ are most responsible for the discrepancy.

Another symptom of disagreement is a highly uneven distribution of the new weights, π_i^*, so that the new sample is dominated by only a few points from the old sample. In effect, this means that (6) is selecting a few of the many possible sample paths from $p(\cdot)$ while ignoring the others. This can be diagnosed by plotting the histogram of the importance ratio, π_i^*/π_i, or by calculating the mean sum of squares of the m largest weights relative to the mean sum of squares of all the weights (Geweke 1989). These statistics are illustrated below in the empirical application.

The diagnostics measure how informative (6) is, and they are likely to be useful as warnings in the policy process. Outside information must alter $p(\cdot)$ in order to sharpen BVAR forecasts, but policy analysts should think twice when large discrepancies between $p(\cdot)$ and $p^*(\cdot)$ occur. If satisfaction of (6) requires severe twisting of $p(\cdot)$, then it would be prudent to review the arguments supporting (6) and to consider whether elements of (6) should be modified or relaxed. Large discrepancies are not necessarily problematic; they may result from unusually incisive judgment about the economy. But when large discrepancies occur, one should have a good story supporting the external constraints. Judgment along these lines is an essential part of the craft of central banking. Measures of the tension between $p(\cdot)$ and $p^*(\cdot)$ are intended to help inform these judgments.

3   The Benchmark VAR

Now we turn to the benchmark VAR. Our first-stage forecasts are generated from a Bayesian vector autoregression with drifting conditional mean parameters and a complex error structure. We study data on RPIX inflation, an output gap, and the nominal three-month Treasury bill rate. The data are sampled quarterly, and they span the period 1957.Q1 to 2002.Q4.

A preliminary measure of the output gap is constructed by exponentially smoothing data on real GDP. The 'trend' in GDP is computed recursively, by iterating on

$$\tau_t = \tau_{t-1} + g(x_t - \tau_{t-1}), \tag{14}$$

She points out that our model builds in a lot of symmetry because of assumptions of conditional normality. The Bank, on the other hand, emphasizes skewness arising from asymmetric risks. If $p^*(\cdot)$ is skewed and $p(\cdot)$ is symmetric, then one of the tails of $p(\cdot)$ might be too thin unless it is very over-dispersed relative to $p^*(\cdot)$. This is a legitimate concern. In practice, however, it can be diagnosed from the distribution of relative entropy weights, which would have a long right tail in this case because draws from the thin tail of $p(\cdot)$ would need to be copied many times to cover $p^*(\cdot)$.

where x_t represents the log of real GDP and the gain parameter $g = 0.075$. Then the output gap is calculated by subtracting this estimate from GDP, $x_t - \tau_t$. This is a crude way to construct an output gap, but two remarks can be offered in its defense. One is that τ_t is measured from a one-sided low-pass filter that preserves the information flow in the data. In contrast, two-sided filters look into the future when estimating τ_t. The other is that this is only a preliminary estimate. The time-varying VAR will itself imply adjustments to the level of τ_t.[6] A rough preliminary measure is all we need to get started.

We estimate a second-order vector autoregression,

$$y_t = X_t'\theta_t + \epsilon_t, \tag{15}$$

where y_t is a vector of observed endogenous variables, the matrix X_t includes constants plus two lags of y_t, and $\epsilon_t = R_t^{1/2}\xi_t$ is a vector of measurement innovations. The normalized innovations ξ_t are assumed to be standard normal, and R_t is a stochastic volatility matrix that is be described below. The variables are ordered so that the nominal interest rate is the first element in y_t, the output gap is second, and the inflation rate is third.

The VAR parameters, θ_t, are themselves a stochastic process, assumed to follow a driftless random walk with a reflective barrier that keeps them from entering regions of the parameter space where the VAR is explosive. The driftless random walk component is represented by a joint prior,

$$f(\theta^T, Q) = f(\theta^T|Q)f(Q) = f(Q)\prod_{s=0}^{T-1} f(\theta_{s+1}|\theta_s, Q). \tag{16}$$

where

$$f(\theta_{t+1}|\theta_t, Q) \sim \mathcal{N}(\theta_t, Q). \tag{17}$$

This unrestricted transition equation makes θ_{t+1} conditionally normal with an increment $v_t = \theta_t - \theta_{t-1}$ that has mean zero and constant covariance Q. Associated with this is a marginal prior $f(Q)$ that makes Q an inverse-Wishart variate.

The reflecting barrier is encoded in an indicator function,

$$I(\theta^T) = \prod_{s=1}^{T} I(\theta_s). \tag{18}$$

The function $I(\theta_s) = 0$ when the roots of associated VAR lag polynomial lie inside the unit circle, and $I(\theta_s) = 1$ otherwise. This is a stability condition for the VAR, reflecting an *a priori* belief about the implausibility of explosive representations for the variables in which we are interested. Whether it is plausible for other variables or other economies is a matter of judgment.

---

[6]The VAR allows a non-zero mean in the gap, and this varies over time. At the end of the day, one can calculate an adjusted gap as $x_t - \tau_t - \gamma_t$, where γ_t is the posterior estimate of the mean gap. The adjusted trend is therefore $\tau_t + \gamma_t$.

The reflecting barrier truncates and renormalizes the random walk prior, so that the joint prior for θ^T and Q becomes

$$p(\theta^T, Q) \propto I(\theta^T) f(\theta^T, Q). \tag{19}$$

From this, one can derive that the conditional prior for $\theta^T | Q$ is

$$p(\theta^T | Q) = \frac{I(\theta^T) f(\theta^T | Q)}{m_\theta(Q)}. \tag{20}$$

The normalizing constant $m_\theta(Q) = \int I(\theta^T) f(\theta^T | Q) d\theta^T$ is the prior probability of a non-explosive sample path for θ^T, conditional on a particular value of Q. Thus, the stability condition truncates and renormalizes $f(\theta^T | Q)$ to eliminate explosive θ's.

Similarly, the marginal prior for Q becomes

$$p(Q) = \frac{m_\theta(Q) f(Q)}{m_Q}, \tag{21}$$

where $m_Q = \int m_\theta(Q) dQ$ is the unconditional prior probability of a non-explosive trajectory for θ^T. Here the marginal prior $f(Q)$ is transformed so that values of Q which are more likely to generate explosive θ's are penalized. This shifts the prior for Q toward values implying less time variation in θ.

The reflecting barrier also alters the one-step transition density, which becomes

$$p(\theta_{t+1} | \theta_t, Q) \propto I(\theta_{t+1}) f(\theta_{t+1} | \theta_t, Q) \phi(\theta_{t+1}, Q). \tag{22}$$

The term $\phi(\theta_{t+1}, Q) = \int I(\theta^{t+2,T}) f(\theta^{t+2,T} | \theta_{t+1}, Q) d\theta^{t+2,T}$ represents the probability that random-walk paths emanating from θ_{t+1} will remain in the nonexplosive region going forward in time. This transformation censors explosive draws from $f(\theta_{t+1} | \theta_t, Q)$ and down-weights those likely to become explosive.

In addition to drifting conditional mean parameters, the model also has drifting conditional variances. The VAR innovations ϵ_t are assumed to be conditionally normal with mean zero and covariance R_t. To model drifting variances, we adapt a formulation from the literature on multivariate stochastic volatility models,[7] specifying that

$$R_t = B^{-1} H_t B^{-1\prime}. \tag{23}$$

The matrix H_t is assumed to be diagonal,

$$H_t = \begin{pmatrix} h_{1t} & 0 & 0 \\ 0 & h_{2t} & 0 \\ 0 & 0 & h_{3t} \end{pmatrix}, \tag{24}$$

---

[7] For example, see Aguilar and West (2000), Jacquier, Polson, and Rossi (1999), and Pitt and Shephard (1999).

with univariate stochastic volatilities along the main diagonal. We further assume that the volatilities evolve as driftless, geometric random walks,

$$\ln h_{it} = \ln h_{it-1} + \sigma_i \eta_{it}. \tag{25}$$

The volatility innovations η_{it} are mutually independent, standard normal variates. The variance of $\Delta \ln h_{it}$ depends on the associated free parameter σ_i^2.

The matrix B is lower triangular with 1's along the main diagonal,

$$B = \begin{pmatrix} 1 & 0 & 0 \\ b_{21} & 1 & 0 \\ b_{31} & b_{32} & 1 \end{pmatrix}. \tag{26}$$

The free elements in B, which we collect into a vector β, allow for correlation in the measurement innovations. The VAR innovations can thus be orthogonalized,

$$B\epsilon_t = \varepsilon_t \tag{27}$$

so that ε_t has variance H_t. While this is just a rotation, not an identification, specification (23) implies that y_{1t} has only one source of volatility, that y_{2t} adds another source, and so on. Thus, the i-th column of R_t consists of a linear combination of $h_{1t}, \ldots, h_{it}$. In principle, this means that the ordering of variables can matter for estimates of the VAR innovation variance.[8]

Finally, for tractability and parsimony, we assume that the random-walk parameter increments v_t are uncorrelated at all leads and lags with the standardized VAR innovations ξ_t,

$$E(\xi_t v_s') = 0, \text{ for all } t \text{ and } s. \tag{28}$$

We also assume that both v_t and ξ_t are independent of the volatility innovations, η_t.

To calibrate our priors, we assume that the hyperparameters and initial value of the drifting parameters are independent across blocks, so that the joint prior factors into a product of marginal priors,

$$f(\theta_0, H_0, Q, \beta, \sigma) = f(\theta_0) f(H_0) f(Q) f(\beta) f(\sigma). \tag{29}$$

Each of the terms in (29) is selected from a family of natural conjugate priors, and each of the marginal priors is specified so that it is proper yet weakly informative.

The unrestricted prior for the initial state is Gaussian,

$$f(\theta_0) \propto \mathcal{N}(\bar{\theta}, \bar{P}). \tag{30}$$

To calibrate (30), we estimate a time-invariant VAR using a training sample covering the period 1957.Q1 to 1961.Q4. The parameters $\bar{\theta}$ and $\bar{P}$ were set equal to the resulting point estimate and asymptotic covariance, respectively. Because the preliminary

---

[8]Cogley and Sargent (2002) reported that the ordering mattered little in U.S. data, but we have not yet checked this for U.K. data.

sample is so short, this is weakly informative for θ_0. The prior mean is just a ballpark number, and the prior variance allows a substantial range of outcomes.

Our prior for Q is inverse-Wishart,

$$f(Q) = \mathcal{IW}(\bar{Q}^{-1}, T_0), \tag{31}$$

with scale matrix $\bar{Q}$ and degree-of-freedom parameter T_0. Because Q is a key parameter governing the rate of drift in θ, we want to maximize the weight that the posterior puts on sample information. The parameter T_0 must exceed the dimension of θ_t in order for $f(Q)$ to be proper, and we set T_0 to the minimal value,

$$T_0 = \dim(\theta_t) + 1. \tag{32}$$

To set the scale matrix $\bar{Q}$, we assume

$$\bar{Q} = \gamma^2 \bar{P}, \tag{33}$$

and we calibrate γ so that $\bar{Q}$ is comparable to the value used by Cogley and Sargent (2002). This results in a conservative setting for $\bar{Q}$, as it involves only a minor perturbation from time-invariance.

Priors for the stochastic volatility parameters are also calibrated to put considerable weight on sample information. The prior for h_{i0} is log-normal,

$$f(\ln h_{i0}) = \mathcal{N}(\ln R_0^{ii}, 10), \tag{34}$$

where $\ln R_0^{ii}$ is the estimate of the log of residual variance of variable i from the preliminary sample. A variance of 10 is huge on a log scale and allows a wide range of values for h_{i0}. As is the case of θ_0, the prior mean for H_0 is no more than a ballpark number surrounded by considerable uncertainty.

Similarly, the prior for β is normal with mean zero and a large variance,

$$f(\beta) = \mathcal{N}(0, 10000 \cdot I). \tag{35}$$

Lastly, the prior for σ_i^2, the variance of the stochastic volatility innovations, is inverse-gamma

$$f(\sigma_i^2) = IG(\delta/2, 1/2), \tag{36}$$

with scale parameter $\delta = 0.0001$ and degree-of-freedom parameter equal to 1. This distribution is also proper yet extremely diffuse.

In the end, our f-prior is weakly informative with its influence being strongly diluted as time goes forward. The p-prior is more informative, as the truncation concentrates the posterior in the stable part of the parameter space.

In terms of the notation of the previous section, we have

$$\Theta^T = (\theta^T, H^T), \tag{37}$$
$$\psi = (Q, \sigma, \beta).$$

11

Our goal is to simulate the BVAR predictive density, $p(Y^{T+1,T+F}|Y^T)$. This is accomplished by sampling in parts, using equation (5) as a guide. Each of the steps is detailed in the appendix.[9]

Briefly, our simulator works as follows. The first step involves sampling from $p(\Theta^T, \psi|Y^T)$, the second term in (5). This is done via the Markov Chain Monte Carlo algorithm of Cogley and Sargent (2002), which is described in appendix A.1. We simulated a Markov chain of length 100,000 and discarded the first 50,000 to ensure convergence to the ergodic distribution. Convergence was checked by inspecting recursive mean plots and comparing results across parallel chains starting from different initial conditions. We found that convergence is rapid and reliable and that the effects of the initial draw dissipated quickly. To economize on storage and to reduce serial correlation in the chain, we saved every 10th draw. Doing so increases the variance of ensemble averages from the simulation, yet it yields a sufficient sample of 5000 draws from $p(\Theta^T, \psi|Y^T)$.

The next step involves simulating a future path for the drifting parameters, $\Theta^{T+1,T+F}$, conditional on a realization of (Θ^T, ψ) from step 1. This involves a slight extension of one of the elements of Cogley and Sargent (2001) to handle the stochastic volatilities, a feature absent from that paper, and it is described in appendix A.2. The forecast horizon F was set 20 quarters ahead,[10] and we executed 5000 replications. This delivers a sample from $p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$, the third term in (5).

The third step is to sample a path for future observables, $Y^{T+1,T+F}$, given a realization from the first two steps. This is just a simulation of a VAR with known coefficients, and it is described in appendix A.3. Again we executed 5000 replications, obtaining a sample from $p(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T)$, the first term in (5).

Taken together, these three simulations deliver a random sample from the joint density, $p(Y^{T+1,T+F}, \Theta^{T+F}, \psi|Y^T)$. We marginalize as in equation (4) by discarding the draws for Θ^{T+F} and ψ, thus eliminating nuisance parameters. The result is an evenly weighted random sample from the BVAR predictive density, $p(Y^{T+1,T+F}|Y^T)$. BVAR fan charts are constructed by reading from the percentiles of this Monte Carlo distribution.

4   Features of $p(\Theta^T, \psi|Y^T)$

This section briefly reviews some features of the retrospective posterior, $p(\Theta^T, \psi|Y^T)$. We focus on elements that are relevant for the construction of fan charts. A more

---

[9]It is worth emphasizing that although the algorithms are tedious, they are not hard. The programs are also modular, which we hope will facilitate modifications, extensions, and other applications. Indeed, Primiceri (2003) and Del Negro (2003) have already introduced important improvements and extensions to our model and methods.

[10]The Bank of England fan charts run 9 quarters ahead, but we were curious about longer horizons as well.

thorough and detailed exploration can be found in Morozov (2004). Benati (2002, 2003) estimates closely related models and studies the evolution of key variables over a much longer span of time.

4.1 Core Inflation

The first figure depicts how long-horizon forecasts of inflation have varied over the last four decades. They are estimated from local linear approximations to mean inflation evaluated at the posterior mean, $E(\theta_{t|T})$. One can express the VAR in companion form as

$$z_t = \mu_{t|T} + A_{t|T} z_{t-1} + u_t, \tag{38}$$

where z_t consists of current and lagged values of y_t, $\mu_{t|T}$ contains the intercepts in $E(\theta_{t|T})$, and $A_{t|T}$ contains the autoregressive parameters. Then mean inflation at date t can be approximated by

$$\bar{\pi}_t = s_\pi (I - A_{t|T})^{-1} \mu_{t|T}, \tag{39}$$

where s_π is a row vector that selects inflation from z_t. This is an approximation to the Beveridge-Nelson (1981) trend in inflation. Because movements in $\bar{\pi}_t$ are expected to persist, we label it 'core' inflation.

Estimates of $\bar{\pi}_t$ are shown in figure 1. Core inflation swept up from approximately 3 percent in the early 1960s to around 13 percent by the mid 1970s. It remained in double digits until 1979, but declined sharply in the first half of the 1980s. It fluctuated around 3.75 and 5.5 percent from 1983 to 1992, and then fell to around 2.5 percent after the adoption of inflation targeting in 1992. Since 1997, core inflation has again begun to drift downward, falling to approximately 2.2 percent by 2002.Q4. Thus, our estimate of core inflation is now a bit below to 2.5 percent, the center of the Bank's official targeting range.
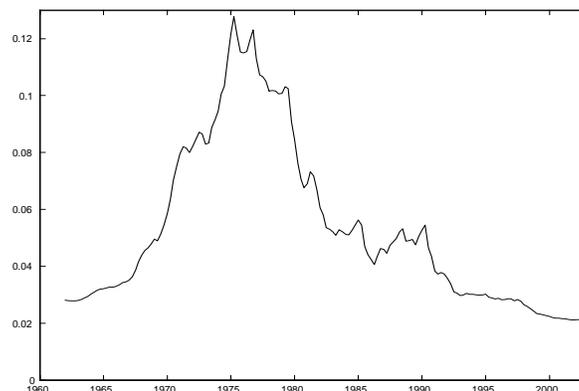

Figure 1: Long-Horizon Forecasts of Inflation

13

A substantial amount of uncertainty surrounds the estimate in any given year, but the figure describes the general historical pattern of inflation. For prediction, what matters most is uncertainty at the end of the sample. This is illustrated in figure 2, which portrays the marginal distribution of posterior estimates of core inflation as defined by (39) as of 2002.Q4. A histogram for $\bar{\pi}_T$ was constructed by calculating a value for every draw of θ_T in the posterior simulation.
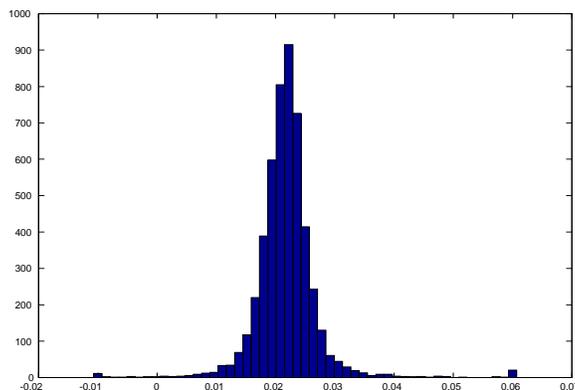

Figure 2: Marginal Distribution for Core Inflation in 2002.Q4

The mean and median of this sample are 2.18 and 2.17 percent respectively. Thus, the posterior distribution is approximately symmetric and centered roughly 30 basis points below the Bank's target. The posterior is also somewhat diffuse, but most of the mass –more than 90 percent– lies between 1.5 and 3 percent. The median absolute devation[11] is 22 basis points, and the interquartile range is 1.94 to 2.38 percent. The probability that $\bar{\pi}_T$ is greater than 3.5 percent is 0.014, and there is also a 0.44 percent chance that $\bar{\pi}_T$ is less than zero.

That core inflation is now below 2.5 percent, the center of the official targeting range, may reflect that policy makers have an asymmetric loss function. If the Bank of England behaved minimizing a discounted quadratic loss function, it would choose a policy rule that sets core inflation equal to the center of the targeting range. If its loss function penalized positive deviations more than negative ones, then it would be optimal for the Bank to choose a policy rule that is biased downward, i.e. one that resulted drove core inflation below 2.5 percent.

From a forecasting perspective, however, this matters because core inflation identifies the limit point to which long-horizon BVAR forecasts converge. As we shall see below, setting the end point equal to the target rather than the core would twist the forecast distribution in a meaningful way. The difference between core and target inflation will manifest itself as a tension between $p(\cdot)$ and $p^*(\cdot)$ in the relative entropy calculations below.

---

[11] There are outliers in the sample, which are collected in the end bins.

For constructing fan charts, it is important that the proposal density $p(\cdot)$ recognize movements in $\bar{\pi}_T$. Core inflation determines the end point to which long-horizon BVAR forecasts converge. If movements in this variable were suppressed, the end point for inflation forecasts would reflect an average of the values seen in the historical sample, which may or may not still be relevant for prediction. For example, suppose the proposal density were based on a constant-coefficient model estimated using a 30 year window of data. The end point for inflation forecasts would depend on the estimated mean from this model, which would reflect an average of the high values of the 1970s with lower values from the 1980s and 1990s. One could argue that the adoption of inflation targeting in 1992 made the high values of the 1970s obsolete. If so, the model's end point would be too high, and inflation forecasts would persistently revert to a sample average that is no longer relevant. A moving window would eventually drop those observations, but this would take time. A random coefficients model like ours adapts more quickly.[12]

4.2   Inflation Persistence

Next we turn to evidence on changes in the degree of inflation persistence. As in our earlier work, inflation persistence is measured by the normalized power spectrum,

$$g_{\pi\pi}(\omega, t) = \frac{2\pi f_{\pi\pi}(\omega, t)}{\int_{-\pi}^{\pi} f_{\pi\pi}(\omega, t) d\omega}, \tag{40}$$

where $f_{\pi\pi}(\omega, t)$ is the instantaneous power spectrum,

$$f_{\pi\pi}(\omega, t) = s_{\pi}(I - A_{t|T}e^{-i\omega})^{-1}\frac{E(R_{t|T})}{2\pi}(I - A_{t|T}e^{i\omega})^{-1\prime}s_{\pi}{}'. \tag{41}$$

The latter is constructed from the posterior mean of the companion matrix, $A_{t|T}$, and the innovation variance, $R_{t|T}$. The normalization in (40) adjusts the spectrum by dividing by the variance in each year, so that $g_{\pi\pi}(\omega, t)$ measures autocorrelation rather than autocovariance. We also multiply by 2π to convert to units that are more easily interpretable. In these units, a white noise process has $g_{\pi\pi}(\omega) = 1$ at all frequencies.[13]

The next two figures depict our estimates of $g_{\pi\pi}(\omega, t)$. Time is plotted on the x-axis, frequency on the y-axis, and power on the z-axis. The first shows how $g_{\pi\pi}(\omega, t)$ has evolved from year to year, and the second compares three estimates from the beginning, middle, and end of the sample.

---

[12]See Kozicki and Tinsley (2001a,b) or Cogley (2002) for more discussion on the significance of moving end points.

[13]This normalization differs from the one in Cogley and Sargent (2002). There a white noise process has a flat spectrum at a level $1/2\pi$.
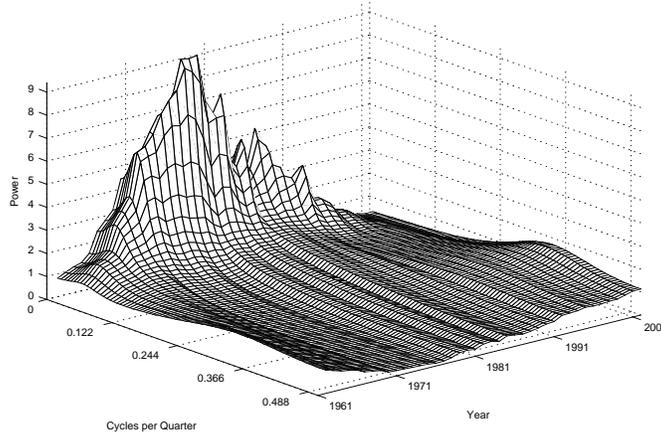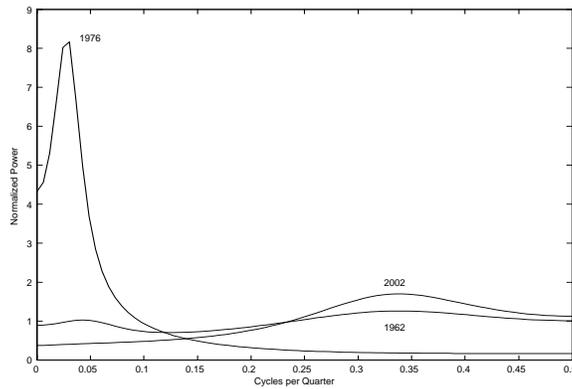
Figure 3: Normalized Spectrum for Inflation

Figure 4: Normalized Spectrum in Selected Years

In the early 1960s, the spectrum for inflation was flat, reflecting that inflation was approximately white noise at that time. By the mid-1970s, however, the spectral shape was very different. There was much more low-frequency power, and a pronounced spectral peak had developed at roughly 8 years per cycle. This pattern reflects positive autocorrelation at short horizons and negative autocorrelation over longer ones. In other words, the monetary policy rules in force were permitting substantial variation in inflation to go unchecked for years at a time, and only very gradually bringing π_t back toward $\bar{\pi}_t$. Thus, not only did core inflation rise during this period, but also deviations from core inflation became more persistent.

This process was reversed after 1980 and especially after 1992. By the end of the sample, the spectrum for inflation had a peak at high frequencies and a trough at low frequencies. This is characteristic of a series that is negatively autocorrelated at short horizons, and it implies partial mean reversion in the price level. This is a sign that monetary policy quickly offsets shocks to inflation, and indeed overshoots its

16

target on the way back. Thus, the policy rules adopted after 1980 have significantly reduced not only the average level of inflation but also its persistence. Shocks to U.K. inflation once persisted for years, but now they are quickly extinguished.

Changes in the transitory dynamics of inflation are also material for constructing fan charts because the shape of prediction intervals depends on the degree of persistence. For example, a white noise process has rectangular prediction intervals with a width proportional to the unconditional standard deviation. A stationary process with positive autocorrelation has prediction intervals that eventually converge to this width, but they are narrower at short horizons, and they fan out at a rate that depends on the degree of persistence. The limit is reached more quickly for a weakly autocorrelated process and more slowly for one that is strongly dependent. For a random walk process, the forecast error variance increases without bound, and fan charts grow ever wider as the forecast horizon increases.

For the same reason that it is important for the proposal density to recognize movements in core inflation, it is also important that it discern changes in inflation persistence. Otherwise, the half life of shocks would be overestimated, and fan charts would have the wrong shape. A constant-coefficient model for U.K. inflation would overstate inflation persistence for two reasons. One is that it would disregard movements in core inflation and measure persistence in terms of reversion toward a full-sample average rather than toward a more appropriate local average. The half-life of the former is greater than that of the latter. A constant-coefficient model would also average high local-persistence observations of the 1970s with lower local-persistence values from the 1990s, giving them equal weight. Since the high-persistence observations come from a different monetary regime, they are less relevant for forecasting and should be given less weight relative to more recent observations. As a consequence of both, prediction intervals constructed from a constant-coefficients model would have the wrong shape, fanning out too slowly from the terminal observation.

4.3   Innovation Variances

Another feature that is important for constructing prediction intervals is the size of future shocks. The following two graphs report on this feature of the data. Figure 5 illustrates the evolution of the posterior mean of the VAR innovation variance $R_{t|T}$. Standard deviations, expressed as basis points at annual rates, are shown in the left-hand column, and correlations are shown in the right. Figure 6 portrays the total prediction variance, $\log |R_{t|T}|$, a measure of the total variance hitting the system at each date.
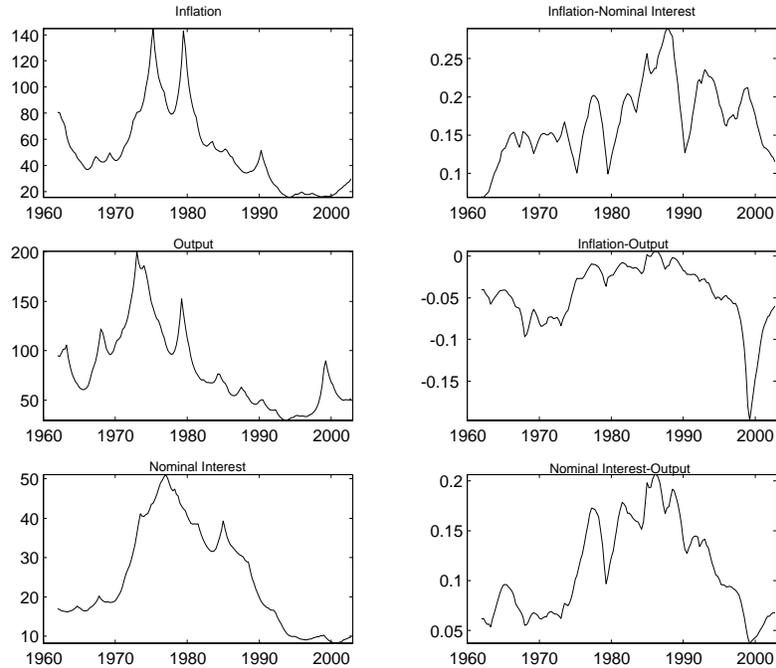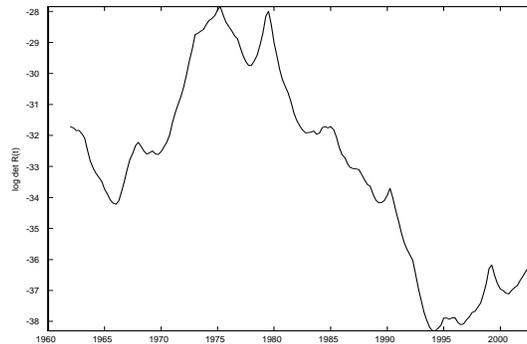
Figure 5: VAR Innovation Variance

Figure 6: Total Prediction Variance

The innovation variance for all three variables has fallen considerably – by 75 or 80 percent – since the 1970s. Although the timing of the stabilization differs across variables in figure 5, figure 6 furnishes evidence of a prolonged decline in the total prediction variance between 1979 and 1994, with a partial reversal thereafter.

The right-hand column of figure 5 suggests there are relatively weak contemporaneous correlations among the variables. There is an inverse correlation between inflation and GDP residuals, though it is small in magnitude. Innovations in nominal interest rates are positively correlated with those in inflation and output, suggesting feedback from current quarter news about inflation and GDP to nominal interest

rates. This is a sign of 'reverse causality,' that the Bank is raising the short rate in response to reports of higher inflation or output.

Within the fan chart community, there is some resistance to using historical prediction errors to measure innovation variances. To some extent, this reflects an attitude that such measures overstate the amount of short-term uncertainty that central banks currently face. A proposal density based on constant historical innovation variances would certainly be subject to this criticism. Such a model would average high variances from the volatile 1970s with lower values from the stable 1990s, thus overstating the likely magnitude of future shocks. Once again, an adaptive BVAR helps address this problem by allowing parameters to drift.

Resistance to using historical prediction errors to calibrate innovation variances also reflects central banks' knowing that they have an information advantage relative to a simple vector autoregression such as ours. Otherwise the bank could not react within quarter to news about inflation or unemployment. This information advantage should reduce conditional variances relative to those we have estimated. However, the extent to which the conditional variance for multi-step forecasts falls depends on the degree of persistence, which we estimate to be rather weak at the end of the sample.

5   Inflation Fan Charts

This section illustrates and compares a number of fan charts for inflation. All are conditioned on observations on inflation, the output gap, and nominal interest through the fourth quarter of 2002. They differ in terms of assumptions about sources of uncertainty within the BVAR and constraints on forecast paths arising outside the model. The first subsection illustrates the BVAR predictive density, $p(\cdot)$, and studies the importance of parameter uncertainty and drift. The second compares $p(\cdot)$ with the Bank of England's predictive density and shows how $p(\cdot)$ is altered by imposing elements of the Bank's forecast.

5.1   BVAR Prediction Intervals and Sources of Uncertainty

Figure 7 portrays the BVAR fan chart for inflation. The sample ends in 2002.Q4, and the one-step ahead forecast is for 2003.Q1. We compare our calculations with those reported for the same forecast dates in the February 2003 *Inflation Report* published by the Bank of England.

The figure illustrates percentiles of the marginal density at each horizon. The center line records the median forecast, the next two lines fanning out from the center illustrate the interquartile range, and the outer two curves represent a centered 90 percent prediction interval. Joint densities for forecast paths can also be constructed from our simulation output, but the Bank does not do this, and we follow their lead
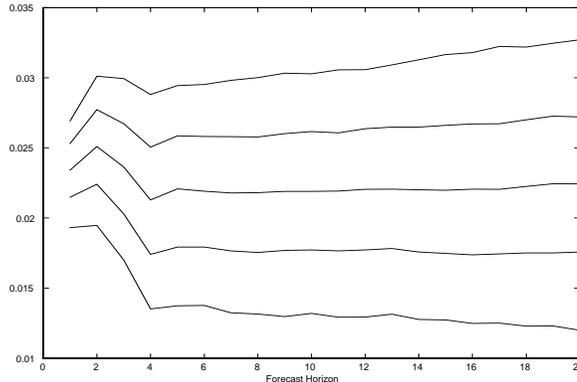
for the sake of comparison.[14]

Figure 7: BVAR Fan Chart for Inflation

In this chart and the ones that follow, inflation is measured on a year-over-year basis,

$$\pi_t^a = \sum\nolimits_{j=0}^{3} \pi_{t-j.} \qquad (42)$$

Accordingly, the prediction intervals are narrow at first, reflecting that realized inflation for the previous three quarters is known. As the forecast horizon lengthens, the prediction intervals gradually fan out as forecasts replace lagged realizations in calculating year-over-year inflation.

Two characteristics are significant. First, notice how quickly the BVAR forecast reverts to core inflation. The median forecast is approximately 2.4 percent for the first quarter of 2003, it increases to 2.5 percent in the second, and then falls back to around 2.2 percent in the third and fourth quarters. Thus, there is a transient increase in inflation which is expected to expire within a year. This quick reversion to the mean reflects the weak persistence that we estimate near the end of the sample.

Second, at the one-year horizon and beyond, when lagged realizations no longer matter, the fan chart becomes rather wide. At the one-year horizon, there is a fifty percent chance that inflation will lie somewhere between 1.75 and 2.5 percent, and a 10 percent chance that it will be lower than 1.3 percent or higher than 2.9 percent. At the 2-year horizon, the edges of the 90 percent prediction interval move out slightly to 1.25 and 3 percent, respectively. Thereafter, the fan chart continues to spread out, reflecting the influence of future parameter drift. The prediction intervals would not spread out as quickly if the current policy regime were sure to remain in force indefinitely.

To quantify the sources of uncertainty about future inflation, we conduct two alternative simulations that shut down particular features of the model. First, to isolate

---

[14]The Bank's methods do not allow for the construction of joint densities across forecast paths.

the contribution of future shocks to the VAR, we construct a fan chart that abstracts from uncertainty about the terminal estimate of the VAR parameters, (Θ_T, ψ), as well as from future parameter drift. These sources are deactivated by setting (Θ_T, ψ) equal to their posterior mean and by constraining $\Theta_{T+i} = \Theta_T$, $i = 1, ..., F$. We emphasize that this exercise still permits parameter drift within the sample, but ignores it when making out-of-sample forecasts.

The results are depicted in the next figure, which compares the median and interquartile range for two predictive densities. The dashed lines are reproduced from figure 7, and they represent results for the unrestricted BVAR. The solid lines illustrate how the predictive density is altered when parameter uncertainty and drift are shut down.
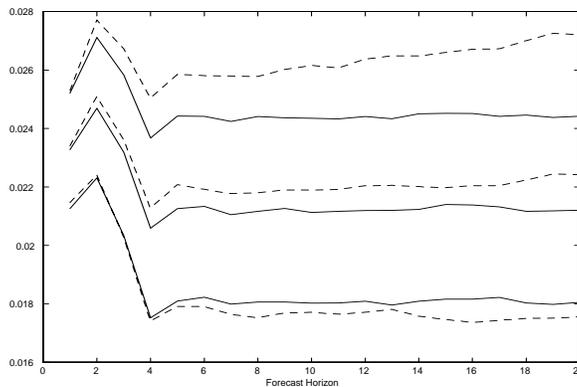

Figure 8: BVAR Fan Chart Abstracting from Parameter Uncertainty

With respect to median inflation forecasts, accounting for parameter uncertainty matters, but only slightly. Median forecasts from the restricted simulation closely track those from the full BVAR for the first year, and then they are persistently lower by about 5 or 10 basis points at longer horizons. The lower edge of the interquartile range also closely tracks that of the BVAR. The chief difference concerns the upper tail, which is quite a bit shorter when parameter uncertainty and drift are deactivated. This is especially relevant at horizons of a year or more.

One can further decompose the effects of parameter uncertainty into a part due to uncertainty about end-of-sample estimates, (Θ_T, ψ), and a part due to future parameter drift. In the next simulation, we incorporate end-of-sample parameter uncertainty by taking draws of (Θ_T, ψ) from $p(\Theta^T, \psi|Y^T)$, but we again deactivate future parameter drift by setting $\Theta_{T+i} = \Theta_T$, $i = 1, ..., F$. The results, which are shown in figure 9, are qualitatively similar to those in figure 8, though smaller in magnitude.
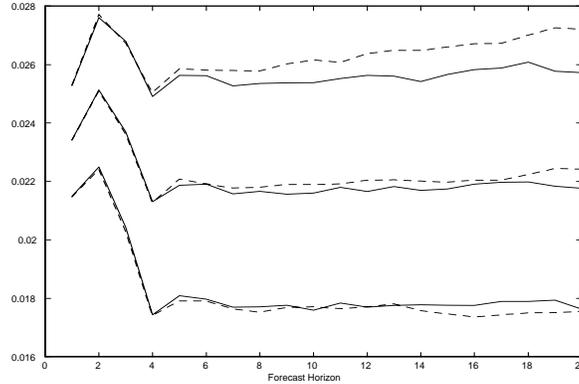
21

Figure 9: BVAR Fan Chart Abstracting from Future Parameter Drift

Now the median forecast is virtually identical to that of the full BVAR. This means that the discrepancy in figure 8 is due entirely to future parameter drift. Once again, the upper tail of the forecast distribution is too short, but only by half as much as before. For 5-year ahead forecasts, the 75th percentile is off by roughly 30 basis points in figure 8, but only by 15 basis points in figure 9. Thus, the discrepancy in the upper tail recorded in figure 8 is due in equal parts to a neglect for end-of-sample parameter uncertainty and future parameter drift.

Disregarding parameter uncertainty and drift means taking for granted that the current monetary regime will remain in force throughout the forecast period. This is a fairly reasonable assumption for the purpose of making short-term forecasts over a horizon of a year or less. But for an accurate account of long-run inflation uncertainty, one must take seriously the possibility of a change in regime. Otherwise one is likely to underestimate the risk of a recurrence of high inflation.

5.2   Relative Entropy Examples

We now provide a few simple examples to show how the BVAR fan chart is altered when additional information is introduced.[15]   As a starting point, it is helpful to compare the BVAR predictive density for inflation with that of the Bank of England for the same period. The two are summarized in the following table. We focus on forecast horizons of 1- and 2-years ahead, which correspond to the dates 2003.Q4 and 2004.Q4. The entries in the table refer to the probability that year-over-year inflation lies in the interval shown at the top of each column. The other rows refer to outcomes of relative entropy calculations, and they are discussed below.

---

[15]See Robertson, et. al. for a number of other, more ambitious examples involving U.S. data.

22

Table 1: Predictive Distributions for Inflation

2003.Q4	< 1.5%	1.5-2.0%	2.0-2.5%	2.5-3.0%	3.0-3.5%	> 3.5%
Bank of England	< 0.05	0.06	0.23	0.36	0.25	0.09
BVAR	0.147	0.263	0.338	0.178	0.055	0.019
Target	0.113	0.240	0.351	0.194	0.069	0.034
MPC Mean	0.033	0.123	0.277	0.250	0.147	0.170
MPC Mean, Variance	0.003	0.050	0.261	0.372	0.230	0.085
2004.Q4						
Bank of England	0.06	0.15	0.25	0.25	0.17	0.12
BVAR	0.152	0.228	0.331	0.189	0.070	0.031
Target	0.089	0.181	0.322	0.231	0.106	0.072
MPC Mean	0.073	0.158	0.278	0.237	0.132	0.123
MPC Mean, Variance	0.043	0.139	0.294	0.278	0.144	0.101

Note: Rows may not sum to one because of rounding errors.

The Bank's numbers are reproduced from the February 2003 *Inflation Report*, and they correspond to a scenario in which the overnight interest rate is expected to remain constant at 3.75 percent. We emphasize that we are not trying to assess the accuracy of the Bank's approximation to the BVAR density or vice versa. The two predictive densities are predicated on different information sets and subjective priors, and therefore are distinct objects. We are just curious about how they differ. We gradually put them on a more comparable footing below, when we carry out relative entropy calculations.

There are two notable differences between the BVAR fan chart and the Bank's predictive density. One is that the Bank's fan chart is shifted to the right relative to that of the BVAR. For example, the Bank's median forecast for 2003 and 2004 are 2.78 and 2.64 percent, respectively, while the BVAR median forecasts are approximately 2.15 and 2.2 percent. The second difference is that the BVAR puts more weight in the tails. In particular, while the Bank forecasts only a 7 percent chance that inflation will fall below 2 percent in 2003, the BVAR fan chart says that this will occur on roughly 40 percent of the sample paths. The BVAR compensates by reducing the probability mass in the 2.5 to 3.5 percent range. According to the Bank, there is a 61 percent chance that inflation will fall in this interval in 2003, while the BVAR puts only a 23 percent probability on this outcome.

The point is not that one is necessarily superior to the other, but that they differ. Presumably the differences reflect additional information that the Bank possesses. Next, we gradually introduce elements of the Bank's information via relative entropy calculations to see how they alter the BVAR density. A central question concerns how much twisting is required to reconcile the BVAR predictive density with the Bank's outlook. The spirit of the relative entropy calculations is to provide feedback on the

Bank's scenarios. The more twisting that is involved, the better the story one would want to have.

A natural place to start is with target inflation. One piece of information the Bank possesses which the BVAR does not is knowledge that the official target is 2.5 percent. Equality of the BVAR end point with the Bank's target can be enforced via a relative entropy transformation. To construct an altered fan chart that satisfies this condition, we resample from the BVAR predictive distribution using probability weights designed so that expected inflation equals 2.5 percent at horizons longer than 2 years. The next figure shows how this alters the BVAR fan chart.
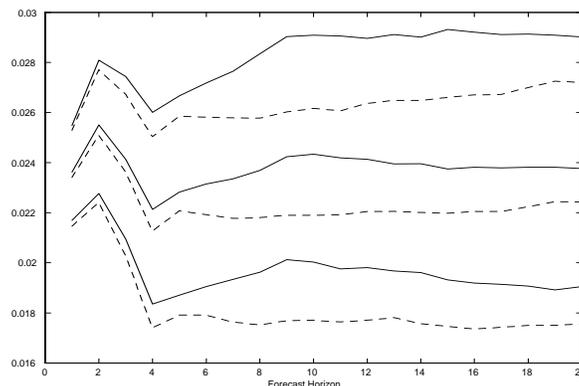

Figure 10: Relative Entropy Fan Chart for the Constraint On Target Inflation

Once again, results for the BVAR are reproduced as dashed lines, while solid lines depict the modified predictive density. Because our estimate of core inflation is below the Bank's target, this constraint shifts the forecast upward at long horizons. The new probability weights π_i^* favor sample paths involving higher inflation and downweight those with lower inflation, so the constraint also shifts the interquartile range. But notice that this matters more at longer horizons. For horizons of a year or less, the forecast distribution changes only slightly.[16]

This visual impression is reinforced by table 1, which compares the relative entropy distribution with that of the BVAR and the Bank. The rows labelled 'Target'[17] record a decrease in the probability of low inflation outcomes relative to the BVAR density and an increase in the probability of high inflation outcomes. This shift is relatively minor at the 1-year horizon, however, and more substantial for 2-year ahead forecasts. At short horizons, the relative entropy density remains closer to that of the BVAR, and it moves gradually toward that of the Bank as the forecast horizon lengthens.

Table 2 reports diagnostics that measure how much twisting is required to enforce the constraint. The $KLIC$ associated with the target constraint is not too far from

---

[16]It is not obvious that this should happen because the re-weighting applies to forecast trajectories, not to individual forecast horizons.

[17]The other rows correspond to other constraints, and they are discussed below.

zero, the maximum importance weight is relatively modest at 15.4, and the relative importance of the largest 10 weights, ω_{10}, is also modest at 63.6. For the sake of comparison, Robertson, et. al. report statistics of 0.66, 119.5, and 443.29, respectively, for an example based on constraints implied by a consumption-CAPM model. Because the consumption-CAPM is known to fit data rather poorly, their example may be taken as a gauge for a high degree of twisting. The imposition of target inflation as the forecast end point is not in the same category.

<div align="center">

Table 2: Twisting Diagnostics

	$KLIC$	$\max(\pi_i^*/\pi_i)$	ω_{10}
Target	0.124	15.4	63.6
MPC Mean	0.502	47.3	220.9
MPC Mean, Variance	0.675	9.4	26.3

</div>

How should the Bank interpret feedback such as this? If the MPC weighs forecast errors symmetrically and truly desires to set mean inflation at long horizons equal to the target, then it is a bit puzzling that the BVAR estimate of core inflation is smaller. On the other hand, the posterior for $\bar{\pi}_T$ is diffuse, and the distance between $\bar{\pi}_T$ and the target is slightly more than 1 median absolute deviation, so the evidence for undershooting is not overwhelming. Moreover, enforcing the target involves only a modest degree of twisting. With a good reason to support the target constraint and little opposition from the diagnostics, one could proceed confidently with this alteration.

The next example modifies the BVAR predictive density in a more forceful way, using more of the Bank's information. The Bank constructs fan charts by reducing all its information to 6 numbers, the mean, variance, and skewness for inflation at horizons of 1 and 2 years. These numbers are then used to calibrate a two-piece normal density for those horizons, and densities for other horizons are found by interpolation. Here we borrow two of their numbers, the mean of inflation 4 and 8 quarters ahead, and use them to form relative entropy constraints. The results are recorded in figures 11 and 12 and in the rows labelled 'MPC Mean' in tables 1 and 2.

The Bank's mean forecast evinces more concern for high inflation than the BVAR forecast. As noted above, the Bank's mean inflation forecast is about 60 basis points higher than the BVAR's for 2003 and 45 basis points higher for 2004, so this constraint lifts the BVAR forecast path and moreover lifts it at shorter horizons than the previous intervention. Indeed, over horizons of 2 years or less, the median of $p^*(\cdot)$ closely tracks the 75th percentile of $p(\cdot)$. Notice also how much the interquartile range shifts. For horizons out to 2.5 years, the 75th percentile of $p^*(\cdot)$ lies above the 90th percentile of $p(\cdot)$, and the 25th percentile for $p^*(\cdot)$ is close to the median for $p(\cdot)$.
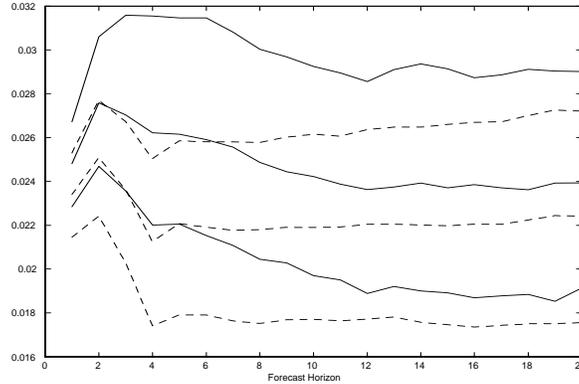
Figure 11: Relative Entropy Fan Chart for Constraints Involving MPC Mean
Inflation

This constraint involves a more energetic twisting than the one involving target
inflation. As shown in table 2, the $KLIC$ rises to 0.50, the maximum importance
weight is 47.3, and the ω_{10} becomes 220.9, values that approach those for the problem-
atic consumption-CAPM example of Robertson, et. al. Further evidence of aggressive
twisting is recorded in figure 12, which plots the histogram of importance weights,
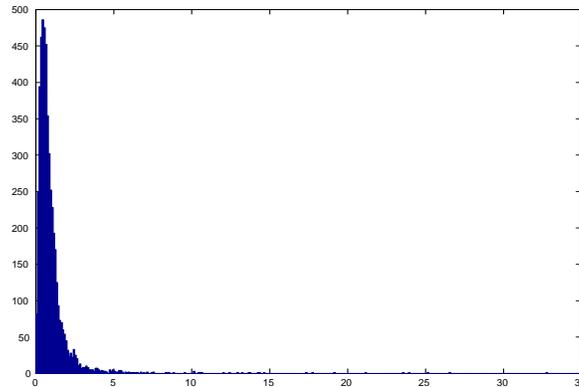π_i^*/π_i, for this example.

Figure 12: Relative Entropy Weights for the MPC Mean Constraint

Two warning signs are evident in this figure, a long upper tail and a concentration
below 1/2. The long upper tail indicates that a few sample paths are being copied
many times in going from $p(\cdot)$ to $p^*(\cdot)$. For example, sample paths from $p(\cdot)$ that have
importance weights in the top ten percent of this distribution account for 37.7 percent
of the draws in $p^*(\cdot)$, while those in the top 1 percent account for 13.6 percent. Some
further detective work reveals that these are high inflation paths. The concentration
of importance weights below 1/2 means that many sample paths (approximately 1/3)
from $p(\cdot)$ are down-weighted in the sample from $p^*(\cdot)$. Draws from $p(\cdot)$ in the bottom

26

quartile of the importance distribution account for only 6.7 percent of paths in $p^*(\cdot)$, and those in the bottom decile account for only 1.6 percent.

The message in this graph, along with the statistics in table 2, is that the MPC mean constraint is highly informative. Simulations of the benchmark BVAR produce a collection of plausible sample paths for inflation. From that collection, the relative entropy procedure emphasizes a few and disregards many others. This is not necessarily a bad thing, but in view of the energetic twisting one would want a good story to support the selection. The diagnostics are telling us that the Bank is boldly taking a particular stand vis a vis the benchmark forecast. A prudent central banker would ask his staff for convincing supporting evidence in the Bank's other intelligence. If such evidence could be assembled, one could proceed confidently with $p^*(\cdot)$. If not, one would reformulate the constraints, looking for informative adjustments to $p(\cdot)$ that could be supported.

Table 1 shows that the MPC mean constraint moves the relative entropy distribution closer to the Bank's outlook, especially at the 2-year horizon, but significant differences remain at the 1-year horizon. Here the relative entropy distribution remains more dispersed, with greater mass in both tails. For example, the Bank says there is a 7 percent chance that inflation will fall below 2 percent at the 4-quarter horizon, while the relative entropy distribution puts a 16 percent probability on this outcome. The Bank also says the probability that inflation will exceed 3.5 percent is 9 percent, but according to the relative entropy distribution this occurs on 17 percent of the sample paths.

How much twisting is needed to align the BVAR fan chart with the Bank's predictive distribution? To match the two, we impose as constraints that the inflation forecasts match both the MPC mean and variance.[18] The results are shown in figures 13 and 14 and in the rows of tables 1 and 2 labelled 'MPC Mean and Variance.'
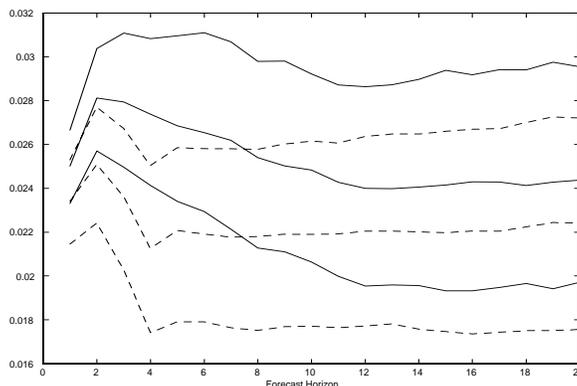

Figure 13: Relative Entropy Fan Chart for the MPC Mean and Variance Constraints

---

[18]Their skewness numbers were not needed to achieve a good match, and in any case they were close to zero in this quarter.

Table 1 shows that these constraints are sufficient to make the relative entropy density conform with the Bank's outlook. But figure 13 and table 2 again suggest that a severe twisting of the BVAR density is needed to accomplish this. Two of the diagnostics, the maximum importance ratio and the relative importance of the 10 largest weights, are actually smaller than in the previous example. The problem is the $KLIC$, which is not only greater than before but also slightly larger than the value for the problematic consumption-CAPM example of Robertson, et. al. It is the high value of the $KLIC$ that suggests a problem.

Once again, we dig deeper by examining the histogram for importance weights, which is portrayed in figure 14. The pattern shown here is similar to that in figure 12, except that the upper tail is a bit shorter and the lower tail more concentrated near zero. In this case, many sample paths from $p(\cdot)$ are assigned virtually no weight at all in the sample from $p^*(\cdot)$. Draws from $p(\cdot)$ in the bottom half of the importance distribution account for only 8.4 percent of the sample paths in $p^*(\cdot)$, while those in the bottom quartile account for only 1.2 percent.
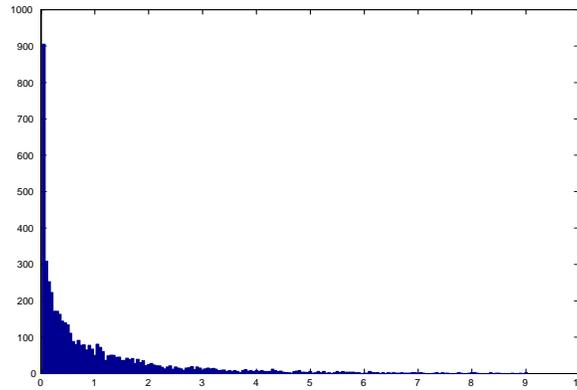
Figure 14: Relative Entropy Weights for the MPC Mean and Variance Constraints

As before, the message is that this is a highly informative selection. Unless one had good, external evidence to back up the constraints, the high value of the $KLIC$ and uneven distribution of importance weights would be a cause for reconsidering the outlook.

6   Conclusion

This paper shows how to construct Bayesian fan charts from a random-coefficients vector autoregression. The adaptive elements of our Bayesian VAR make it well suited to characterizing inflation uncertainty in an economy in which monetary institutions and policy are subject to ongoing change. Measures of core or target inflation, the degree of inflation persistence, and the variance of shocks to inflation contribute to the

28

size and shape of prediction intervals. Since our model is designed to track changes in these features, the fan charts that it produces will adapt in size and shape as the monetary system evolves.

We have illustrated our methods by constructing a variety of fan charts for U.K. inflation for 2003 to 2008. Among other things, we find that parameter uncertainty and drift matter only slightly for inflation forecasts one or two years ahead, the forecast horizons upon which central banks usually focus; but their contribution to inflation uncertainty grows as the forecast horizon lengthens, and they become quite important for forecasts 4 or 5 years out. For characterizing the uncertainty associated with long-term inflation forecasts, one should not take for granted that monetary relations will remain unchanged. There have been important changes in the past, and perhaps we should concede that important changes can happen again.

We have also reviewed how to sharpen BVAR forecasts by introducing outside information and have given some examples to illustrate how to interpret some diagnostics. Our examples concerned elements of the Bank of England's outlook for 2002.Q4. There are pronounced differences between the BVAR fan chart for that quarter and the Bank's forecasts. The BVAR intervals can be made to conform with the Bank's fan chart, but the relative entropy diagnostics point to a severe twisting of the BVAR density. In cases of substantial twisting such as this, we would recommend a careful review of the external evidence being used to twist the forecasts. The more vigorous the twisting, the more persuasion seems called for.

A   Simulating the BVAR Posterior Predictive Density, $p(Y^{T+1,T+F}|Y^T)$

As outlined in the text, the BVAR predictive density is simulated in parts according to the factorization in (5).

- $p(\Theta^T, \psi|Y^T)$ is simulated via the MCMC algorithm of Cogley and Sargent (2002).

- Given a draw of (Θ^T, ψ) from this density, $p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$ is simulated by adapting one of the components of Cogley and Sargent (2001). In fact, because of Θ_t is Markov, at this point one can marginalize with respect to the history Θ^{T-1} and simulate $p(\Theta^{T+1,T+F}|\Theta_T, \psi, Y^T)$.

- Given a draw of (Θ^{T+F}, ψ) from the first two steps, $p(Y^{T+1,H}|\Theta^{T+F}, \psi, Y^T)$ is a VAR with known parameters, and it is simulated in the usual way.

- The nuisance parameters Θ^{T+F} and ψ are then discarded, yielding a sample from $p(Y^{T+1,H}|Y^T)$.

A.1   Simulating $p(\Theta^T, \psi|Y^T)$

Cogley and Sargent (2002) show that

$$p(\Theta^T, \psi|Y^T) \propto I(\Theta^T) f(\Theta^T, \psi|Y^T),$$

where $f(\Theta^T, \psi|Y^T)$ is the posterior corresponding to the model that does not impose the stability constraint. Therefore a sample from $p(\Theta^T, \psi|Y^T)$ can be drawn by simulating $f(\Theta^T, \psi|Y^T)$ and discarding realizations that violate the stability constraint. Further, they developed a MCMC algorithm for simulating $f(\Theta^T, \psi|Y^T)$.

A.1.1   MCMC for $f(\Theta^T, \psi|Y^T)$

The basic idea behind MCMC is to specify a Markov chain whose transition kernel has a limiting distribution equal to the target posterior. The key to constructing an appropriate Markov chain is to break the joint posterior into various conditional distributions or submodels from which it is easy to sample. If implementable sampling procedures can be devised for the full set of submodels, then one can construct a Markov chain by cycling through simulations of each of them. The process of alternating between draws from conditional distributions is a special case of MCMC known as the 'Gibbs sampler' (Gelfand and Smith 1990). Remarkably, under mild and verifiable regularity conditions, the stationary distribution of the Gibbs sampler is the joint distribution of interest. Thus it is possible to sample from the joint posterior without knowledge of its form.

In some cases, Bayes's theorem delivers a convenient expression for a conditional kernel but not the conditional density. For example, sometimes the normalizing constant is too costly to compute.[19] Gibbs sampling is infeasible in such cases, because it requires the full set of conditional densities. But one can resort to hybrid MCMC method known as 'Metropolis-within-Gibbs' that involves replacing some of the Gibbs steps with Metropolis accept/reject steps. The latter typically involve the conditional kernel instead of the conditional density, but the target posterior is still the stationary distribution of the chain. To design a Metropolis-Hastings step, one chooses a 'proposal' density which is cheap to simulate and which closely mimics the shape of target and accepts or rejects with a certain probability, designed to make the proposal sample conform to a sample from the target.

We compose a Metropolis-within-Gibbs sampler for the unrestricted Bayesian VAR. The Markov chain cycles through 5 steps:

1. sample θ^T from $f\left(\theta^T|Y^T, H^T, Q, \sigma, \beta\right)$,

2. sample Q from $f\left(Q|Y^T, \theta^T, H^T, \sigma, \beta\right)$,

---

[19]In our model this is true of the stochastic volatilities.

3. sample β from $f\left(\beta|Y^T, \theta^T, H^T, Q, \sigma\right)$,

4. sample σ from $f\left(\sigma|Y^T, \theta^T, H^T, Q, \beta\right)$,

5. for each i, draw $\{h_{it}\}_{t=1}^T$ by cycling through T Metropolis chains that target $f\left(h_{it}|h_{-it}, Y^T, \theta^T, \sigma_i\right)$, where h_{-it} denotes the rest of the h_i vector at dates other than t.

VAR Coefficients, θ^T: Knowledge of the histories of volatility innovations h_i^T and their contemporaneous correlation B implies knowledge of the measurement innovations variances R^T. The joint density of the history of VAR innovations depends on h_i^T and B only through R^T. While sampling from $f(\theta^T|Y^T, R^T, Q)$ might seem difficult because it is a high-dimensional distribution, we can use the efficient block-sampling algorithm of Carter and Kohn (1994) and Fruhwerth-Schnatter (1994), which is known as *forward-filtering, backward-sampling*.

This algorithm is an application of the Kalman filter. By repeated application of Bayes's theorem, the unrestricted density $f\left(\theta^T|Y^T, R^T, Q\right)$ can be factored as

$$f\left(\theta^T|Y^T, R^T, Q\right) = f\left(\theta_T|Y^T, R^T, Q\right) \prod_{t=1}^{T-1} f\left(\theta_t|\theta_{t+1}, Y^t, R^T, Q\right)$$
$$= \mathcal{N}\left(\theta_{T|T}, P_{T|T}\right) \prod_{t=1}^{T-1} \mathcal{N}\left(\theta_{t|t+1}, P_{t|t+1}\right),$$
(43)

where

$$\begin{aligned}
\theta_{t|t} &= E(\theta_t|Y^t, R^T, Q), \\
\theta_{t|t+1} &= E(\theta_t|\theta_{t+1}, Y^t, R^T, Q), \\
P_{t|t} &= \text{Var}(\theta_t|Y^t, R^T, Q), \\
P_{t|t-1} &= \text{Var}(\theta_t|Y^{t-1}, R^T, Q), \\
P_{t|t+1} &= \text{Var}(\theta_t|\theta_{t+1}, Y^t, R^T, Q),
\end{aligned}$$
(44)

represent conditional means and variances of the unrestricted conditional transition densities. The second line in (43) follows since the unrestricted state and measurement transitions are conditionally normal and are fully characterized by their means and variances defined in (44). These can be computed recursively starting from some initial values by using the Kalman filter. We initialize the Kalman filter with the prior mean and variance for θ_0,

$$\begin{aligned}
\theta_{0|0} &= \bar{\theta}, \\
P_{0|0} &= \bar{P}.
\end{aligned}$$
(45)

The forward Kalman filter then iterates on

$$
\begin{aligned}
P_{t|t-1} &= P_{t-1|t-1} + Q, \\
K_t &= P_{t|t-1}X_t \left(X_t'P_{t|t-1}X_t + R_t\right)^{-1}, \\
\theta_{t|t} &= \theta_{t-1|t-1} + K_t \left(y_t - X_t'\theta_{t-1|t-1}\right), \\
P_{t|t} &= P_{t|t-1} - K_tX_t'P_{t|t-1},
\end{aligned}
\tag{46}
$$

where K_t is the Kalman gain. As t approaches T, the forward recursion delivers $\theta_{T|T}$ and $P_{T|T}$, pinning down the second term in (43).

To derive $\theta_{t|t+1}$ and $P_{t|t+1}$, we use a backward recursion that updates conditional means and variances to reflect the additional information about θ_t contained in θ_{t+1}. Because θ_t is conditionally normal, this updating follows

$$
\begin{aligned}
\theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}\left(\theta_{t+1} - \theta_{t|t}\right), \\
P_{t|t+1} &= P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t}.
\end{aligned}
\tag{47}
$$

A random trajectory for θ^T is generated alongside the backward recursion. θ_T is drawn from $\mathcal{N}(\theta_{T|T}, P_{T|T})$. Then, conditional on its realization, (47) is used to compute $\theta_{T-1|T}$ and $P_{T-1|T}$ and θ_{T-1} can be drawn from $\mathcal{N}(\theta_{T-1|T}, P_{T-1|T})$, and so forth back to the beginning of the sample.

Innovation Variance, Q: Conditional on the history of VAR coefficients θ^T and observed data Y^T, the measurement innovations ϵ_t are observable. The other conditioning variables are redundant in reconstructing the conditional distribution of Q, because the random walk increment v_t is uncorrelated with ϵ_s and $\eta_{i\tau}$. The prior distribution of the covariance matrix Q is inverse-Wishart, with scale matrix $\bar{Q}$ and degrees of freedom T_0:

$$
f(Q) = \mathcal{IW}(\bar{Q}^{-1}, T_0).
\tag{48}
$$

This is convenient because the combination of an inverse-Wishart prior with a normal likelihood yields an inverse-Wishart posterior (Kotz, Balakrishnan, and Johnson 2000),

$$
f\left(Q|Y^T, \theta^T\right) = \mathcal{IW}(Q_1^{-1}, T_1),
\tag{49}
$$

with the scale matrix Q_1 and degrees of freedom T_1:

$$
\begin{aligned}
Q_1 &= \bar{Q} + \sum_{t=1}^{T} v_t v_t', \\
T_1 &= T_0 + T.
\end{aligned}
\tag{50}
$$

To draw a random matrix from $\mathcal{IW}(Q_1^{-1}, T_1)$, draw $N(1+NL)T_1$ random numbers from the standard normal distribution, arrange it into a $N(1+NL) \times T_1$ matrix u, and set

$$
Q = \left(Q_1^{-1/2}uu'\left(Q_1^{-1/2}\right)'\right)^{-1}.
\tag{51}
$$

Standard Deviation of the Volatility Innovations, σ: Here we are looking for a way to draw from the full conditional distribution for $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$,

$$f\left(\sigma \big| Y^T, \theta^T, H^T, Q, \beta\right) = \prod_{i=1}^{N} f\left(\sigma_i | h_i^T\right). \tag{52}$$

The simplification afforded in (52) follows because volatility innovations are mutually independent, and scaled volatility innovations, $\sigma_i \eta_{it}$, can be directly computed using only the h_i^T history via equation (25). All other conditioning variables either contain information already encapsulated in H^T,[20] or they are irrelevant as they only contain information about random variables that are uncorrelated with volatility innovations.

The scaled volatility innovations are *iid* normal with mean zero and variance σ_i^2. Our prior is that σ_i^2 are independent inverse-gamma random variables,

$$f\left(\sigma_i^2\right) = \mathcal{IG}(\frac{\mathcal{G}_0}{2}, \frac{\delta_0}{2}), \tag{53}$$

where $\mathcal{G}_0$ and δ_0 are the degrees of freedom and scale parameters, respectively. Combining normally distributed $\sigma_i \eta_{it}$ data with an inverse-gamma prior results in inverse-gamma posterior (Johnson, Kotz, and Balakrishnan 1994),

$$f\left(\sigma_i^2 | h_i^T\right) = \mathcal{IG}(\frac{\mathcal{G}_1}{2}, \frac{\delta_1}{2}), \tag{54}$$

where

$$\begin{aligned}
\mathcal{G}_1 &= \mathcal{G}_0 + T, \\
\delta_1 &= \delta_0 + \sum_{t=1}^{T}(\sigma_i \eta_{it})^2 = \delta_0 + \sum_{t=1}^{T}(\Delta \ln h_{it})^2.
\end{aligned} \tag{55}$$

This is a convenient distribution from which to sample.

Covariance of Volatility Innovations, β: Next, we consider the distribution of β conditional on the data and other parameters

$$f\left(\beta | Y^T, \theta^T, H^T, Q, \sigma\right). \tag{56}$$

This distribution can be obtained from the relation (27), which can be viewed as a system of unrelated regressions. Bayesian updating of regression coefficients is a well established procedure. Denote β_j as the nontrivial regression coefficients in j-th

---

[20] For example, B orthogonalizes R_t and therefore carries information about H_t, but this is redundant given direct observations on H_t.

equation in (27), $j \geq 2$. Then, (27) can be rewritten as renormalized regressions with independent standard normal residuals,

$$\frac{\epsilon_{jt}}{\sqrt{h_{jt}}} = -\left(\frac{\epsilon_{1:j-1,t}}{\sqrt{h_{jt}}}\right)' \beta_j + \frac{\varepsilon_{jt}}{\sqrt{h_{jt}}}, \quad j \geq 2. \tag{57}$$

It follows that

$$f\left(\beta | Y^T, \theta^T, H^T, Q, \sigma\right) = f\left(\beta | Y^T, \theta^T, H^T\right) = \prod_{j=2}^{N} f\left(\beta_j | \theta^T, Y^T, h_j^T\right). \tag{58}$$

Many of the conditioning variables in (58) drop out because we only need θ^T and Y^T to recover ϵ_t and h_j^T to normalize the residuals in (57). Independence of the elements of ε_t implies that h_i^T, $i \neq j$ are irrelevant for β_j, and this justifies the second equality in (58).

Assuming an independent normal prior for β_j,

$$f\left(\beta_j\right) = \mathcal{N}(\beta_{j0}, V_{j0}), \tag{59}$$

the posterior is also normal (Judge, Lee, and Hill 1988),

$$f\left(\beta_j | \theta^T, Y^T, h_j^T\right) = \mathcal{N}(\beta_{j1}, V_{j1}), \quad j \geq 2, \tag{60}$$

where

$$V_{j1} = \left(V_{j0}^{-1} + \epsilon_{1:j-1,1:T}\left(\frac{1}{\sqrt{h_{j,1:T}}}\right)\left(\frac{1}{\sqrt{h_{j,1:T}}}\right)' \epsilon_{1:j-1,1:T}'\right)^{-1}$$

$$\beta_{j1} = V_{j1}\left(V_{j0}^{-1}\beta_{i0} + \epsilon_{1:j-1,1:T}\left(\frac{1}{\sqrt{h_{j,1:T}}}\right)\left(\frac{1}{\sqrt{h_{j,1:T}}}\right)' \epsilon_{j,1:T}'\right). \tag{61}$$

Stochastic Volatilities, H^T: Since the stochastic volatilities h_{it} and $h_{j\tau}$ are independent for all $j \neq i$, t and τ, we can proceed on the univariate basis:

$$f\left(h_1^T, h_2^T, \ldots, h_N^T | Y^T, \theta^T, Q, \sigma, \beta\right) = \prod_{i=1}^{N} f\left(h_i^T | Y^T, \theta^T, Q, \sigma, \beta\right). \tag{62}$$

To sample from $f\left(h_i^T | Y^T, \theta^T, \beta\right)$, we exploit the univariate algorithm of Jacquier, Polson and Rossi (1994), with some modifications.

Jacquier, et. al. adopted a date-by-date blocking scheme and developed the conditional kernel for

$$f\left(h_{it} | h_{-it}, Y^T, \theta^T, \psi\right) = f\left(h_{it} | h_{it-1}, h_{it+1}, \varepsilon_{it}, \sigma_i\right). \tag{63}$$

Equation (63) is based on the fact that h_{it-1}, h_{it+1}, ε_{it} and σ_i are sufficient statistics for h_{it}. Indeed, conditional on Y^T, θ^T and β, we can infer the history of the orthogonalized VAR residuals ε^T. Knowledge of Q is redundant given θ^T. The variables σ_j and h_j^T for $j \neq i$ are irrelevant because the stochastic volatilities are independent by assumption. Since $\{h_{it}\}_{t=1}^T$ is Markov, knowledge of h_{it-1}, h_{it+1} and ε_{it} is sufficient. The lognormal form of the volatility equation and the normal form for the conditional sampling density imply (via application of Bayes theorem) the somewhat unusual form for the univariate conditional density (63):

$$f\left(h_{it}|h_{it-1}, h_{it+1}, \varepsilon_{it}, \sigma_i\right) \propto f\left(\varepsilon_{it}|h_{it}\right) f\left(h_{it}|h_{it-1}, \sigma_i\right) f\left(h_{it+1}|h_{it}, \sigma_i\right)$$

$$\propto h_{it}^{-3/2} \exp\left\{-\frac{\varepsilon_{it}^2}{2h_{it}} - \frac{(\ln h_{it} - \mu_{it})^2}{2\sigma_i^2}\right\}. \tag{64}$$

The variable μ_{it} is the conditional mean of $\ln h_{it}$ implied by (25) and knowledge of h_{it-1} and h_{it+1}:

$$\mu_{it} = \frac{\ln h_{it-1} + \ln h_{it+1}}{2}. \tag{65}$$

Here we combine both of the lognormal terms and complete the square on $\ln h_{it}$.

At the beginning and end of the sample, the formula (64) has to be modified because only one of the adjacent values for h_{it} is available, and also there is no observed value of ε_{i0}. Suppose that the prior for $\ln h_{i0}$ is $\mathcal{N}(\bar{\mu}_0, \bar{\sigma}_0^2)$. Then the posterior conditional kernel becomes

$$f\left(h_{i0}|h_{i1}, \sigma_i\right) \propto h_{i0}^{-1} \exp\left\{-\frac{(\ln h_{i0} - \mu_{i0})^2}{2\sigma_{ic}^2}\right\}, \tag{66}$$

where

$$\sigma_{ic}^2 = \frac{\bar{\sigma}_0^2 \sigma_i^2}{\bar{\sigma}_0^2 + \sigma_i^2},$$

$$\mu_{i0} = \sigma_{ic}^2 \left(\frac{\bar{\mu}_0}{\bar{\sigma}_0^2} + \frac{h_{i1}}{\sigma_i^2}\right). \tag{67}$$

Similarly, at the end of the sample the conditional kernel becomes

$$f\left(h_{iT}|h_{iT-1}, \varepsilon_T, \sigma_i\right) \propto h_{iT}^{-3/2} \exp\left\{-\frac{\varepsilon_{iT}^2}{2h_{iT}} - \frac{(\ln h_{iT} - \ln h_{iT-1})^2}{2\sigma_i^2}\right\}. \tag{68}$$

Direct sampling from (64), (66) and (68) is difficult because of the nonstandard functional form. Notice also that right hand sides of (64), (66) and (68) are conditional kernels, not conditional densities. There is no analytical expressions for the normalizing constant, which itself is a function of conditioning arguments and varies as the sampler proceeds. Therefore, a pure accept/reject sampling method is difficult to apply. Instead, we adopt a Metropolis accept/reject sampler.

Let f stand for the target conditional density, one of the (64), (66) or (68). A Metropolis algorithm uses a *proposal density* q which is cheap to simulate and which closely mimics the shape of f. We choose q to be the log-normal density implied by the volatility equation,

$$q(h_{it}) \propto h_{it}^{-1} \exp\left\{-\frac{(\ln h_{it} - \mu_{it})^2}{2\sigma_i^2}\right\},\tag{69}$$

with obvious modifications for the beginning and end of sample.[21] The acceptance probability for the m-th draw can be computed as

$$\begin{aligned}\alpha_m &= \frac{f\left(\varepsilon_{it}|h_{it}^m\right)q(h_{it}^m)}{q(h_{it}^m)}\frac{q(h_{it}^{m-1})}{f\left(\varepsilon_{it}|h_{it}^{m-1}\right)q(h_{it}^{m-1})}\\[2mm] &= \frac{(h_{it}^m)^{-1/2}\exp\left\{-\frac{\varepsilon_{it}^2}{2h_{it}^m}\right\}}{\left(h_{it}^{m-1}\right)^{-1/2}\exp\left\{-\frac{\varepsilon_{it}^2}{2h_{it}^{m-1}}\right\}}.\end{aligned}\tag{70}$$

Therefore sampling of h_{it}^m proceeds as follows. Draw a trial value h_{it} from (69), and draw ξ from uniform distribution on $[0,1]$. If $\xi \le \alpha_m$, accept the trial value, $h_{it}^m = h_{it}$. Otherwise set $h_{it}^m = h_{it}^{m-1}$.

A.1.2 Rejection Sampling

The algorithm outlined thus far delivers a sample from the posterior for the unrestricted model. To enforce the stability constraint, we rely on rejection sampling. At the end of each cycle of the Metropolis-within-Gibbs sampler, we compute the VAR roots associated with draws of $\theta_{t|T}$, and we discard the entire draw of the cycle if there are explosive VAR roots at any date.

A.2 Simulating $p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$

Having processed data through date T, the next step is to simulate the future parameters. Conditional on the data, hyperparameters ψ, and terminal states θ_T, H_T,[22] the posterior density for future states is easy to compute. First, as for the in-sample density, the restricted density can be expressed as

$$p(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T) \propto I(\Theta^{T+1,T+F})f(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T).\tag{71}$$

---

[21] This is simpler than proposal of Jacquier, et. al. who use inverse-gamma proposal density. Both proposal densities blanket the tails of true conditional density (64) which ensures geometric convergence of MCMC algorithm, see Johannes and Polson (2002). On the other hand, Jacquier, et. al. say the inverse-gamma proposal density is more efficient with regard to the rejection rate.

[22] Since VAR coefficients and stochastic volatilities are Markov, it is sufficient to know their current values only, not the entire past history.

We can therefore sample from $f(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$ and reject explosive realizations.

Second, because innovations to θ_t and h_t are independent, $f(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T)$ can be factored as

$$f(\Theta^{T+1,T+F}|\Theta^T, \psi, Y^T) = f\left(\theta^{T+1,T+F}|\theta_T, H_T, \psi, Y^T\right) f\left(H^{T+1,T+F}|\theta_T, H_T, \psi, Y^T\right),$$

$$(72)$$

$$= \prod_{i=1}^{F} f\left(\theta_{T+i}|\theta_{T+i-1}, Q\right) \prod_{i=1}^{F} f\left(H_{T+i}|H_{T+i-1}, \sigma\right).$$

Apart from the restriction on explosive autoregressive roots, θ_{T+1} is conditionally normal with mean θ_T and variance Q. Similarly, conditional on θ_{T+1} and Q, θ_{T+2} is normal with mean θ_{T+1} and variance Q. Therefore, to sample future θ_{T+i}, we simply draw v_{T+i} from a $N(0, Q)$ and iterate on

$$\theta_{T+i} = \theta_{T+i-1} + v_{T+i}, \tag{73}$$

starting with θ_T. The stability restriction is then enforced by discarding explosive draws.

In the same way, we draw future stochastic volatilities by starting with h_{iT} and iterating on the volatility equation,

$$\ln h_{jT+i} = \ln h_{jT+i-1} + \sigma_i \eta_{jT+i}, \ \ j = 1, \dots, N, \tag{74}$$

with η_{jT+i} being drawn from a univariate standard normal distribution.

A.3 Simulating $p(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T)$

All that remains is to simulate future data from $p\left(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T\right)$, which is the predictive density for a VAR with known coefficients. This can be factored as

$$p\left(Y^{T+1,T+F}|\Theta^{T+F}, \psi, Y^T\right) = \prod_{i=1}^{F} p\left(y_{T+i}|Y^{T+i-1}, \Theta_{T+i}, \psi\right), \tag{75}$$

by using the Markovian property and the observation equation. Conditional on H_{T+i} and ψ, the measurement innovation ϵ_{T+i} is distributed normal with mean zero and covariance $R_{T+i} = B^{-1}H_{T+i}B^{-1\prime}$. Therefore, to sample from (75), we take random draws of ϵ_{T+i} from $\mathcal{N}(0, R_{T+i})$ for $i = 1, \dots, F$ and iterate on the measurement equation

$$y_{T+i} = X'_{T+i}\theta_{T+i} + \epsilon_{T+i}, \ \ i = 1, \dots, F, \tag{76}$$

using lags of y_{T+i} to compute X_{T+i}.

B   A Multinomial Resampling Algorithm

Suppose we have a sample $Y_i^{T+1,T+F}$, $i = 1, ..., N_p$, from $p(\cdot)$ as well as the associated relative entropy weights π_i^*. To obtain a sample from $p^*(\cdot)$, we associate with every draw $Y_i^{T+1,T+F}$ a certain number of 'children.' Sample paths with large π_i^* have more offspring than those with small π_i^*. Let NC represent the $N_p x 1$ vector that lists the number of children associated with each draw from $p(\cdot)$. The total number of offspring is restricted so that $\sum_{i=1}^{N_p} NC_i = N_{p^*}$, the predetermined number of elements in the $p^*(\cdot)$ sample. It helps to set $N_{p^*} > N_p$. The vector NC is a multinomial random variable,

$$NC \sim MN(N_{p^*}; \pi_1^*, \pi_2^*, ..., \pi_{N_p}^*), \tag{77}$$

and a realization for NC is obtained by sampling from this distribution. Given this realization, a sample from $p^*(\cdot)$ is constructed by making NC_i copies of $Y_i^{T+1,T+F}$, $i = 1, ..., N_p$.

References

Aguilar, Omar and Mike West, 2000, "Bayesian Dynamic Factor Models and Portfolio Allocation,' *Journal of Business and Economic Statistics*, pp. 338-357.

Benati, Luca, 2002, "Investigating Inflation Persistence Across Monetary Regimes," unpublished manuscript, Bank of England.

Benati, Luca, 2003, "Evolving Post World War II U.K. Economic Preformance," unpublished manuscript, Bank of England.

Carter, C.K. and R. Kohn, 1994, "On Gibbs Sampling for State-Space Models," Biometrika 81(3), pp. 541-553.

Cogley, Timothy, 2002, "A Simple Adaptive Measure of Core Inflation," *Journal of Money, Credit, and Banking* 34, pp. 94-113.

Cogley, Timothy and Thomas J. Sargent, 2001, "Evolving Post World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual* 16, pp. 331-373.

Cogley, Timothy and Thomas J. Sargent, 2002, "Drifts and Volatilities: Monetary Policies and Outcomes in the Post War U.S.," unpublished manuscript, Department of Economics, New York University.

Del Negro, Marco, 2003, "Comment on 'Drifts and Volatilities' by Cogley and Sargent," unpublished manuscript, Research Department, Federal Reserve Bank of Atlanta.

Fruhwerth-Schnatter, S., 1994, "Data Augmentation and Dynamic Linear Models," Journal of Time Series Analysis 15, pp. 183-202.

Gelfand, A.E. and Adrian F.M. Smith, 1990, "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85, pp. 398-409.

Geweke, John, 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* 57, pp. 1317-1339.

Gordon, Neil J., D.J. Salmond, and A.F.M. Smith, 1993, "A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Proceedings-F* 140, pp. 107-113.

Inflation Report, Bank of England, February 2003

Jacquier, Éric, Nicholas G. Polson, and Peter Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics* 12, pp. 371-418.

——, ——, and, ——, 1999, "Stochastic Volatility: Univariate and Multivariate Extensions," unpublished manuscript, Finance Department, Boston College and Graduate School of Business, University of Chicago.

Johannes, Michael and Nicholas G. Polson, 2002, "MCMC Methods for Financial Econometrics," unpublished manuscript, Graduate School of Business, Columbia University.

Johnson, N.L. S. Kotz, and N. Balakrishnan, 1994, *Continuous Univariate Distributions* (Wiley: New York).

Judge, G.G., T.C. Lee, and R.C. Hill, 1988, *Introduction to the Theory and Practice of Econometrics* (Wiley: New York).

Kitamura, Yuchi and Michael J. Stutzer, 1997, "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica* 65, pp. 861-874.

Kotz, S., N. Balakrishnan, and N.L. Johnson, 2000, *Continuous Multivariate Distributions* (Wiley: New York).

Kozicki, Sharon and Peter A. Tinsley, 2001a, "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, pp. 613-652.

—— and ——, 2001b, "Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations," *Journal of Economic Dynamics and Control* 25, pp. 149-184.

Morozov, Sergei, 2004, Ph.D. Dissertation, Stanford University.

Parkin, Michael, 1993, "Inflation in North America," in *Price Stabilization in the 1990s*, edited by Kumiharo Shigehara.

Pitt, Mark and Neil Shepard, 1999, "Time-Varying Covariances: A Factor Stochastic Volatility Approach," in *Bayesian Statistics* 6, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., (Oxford University Press: Oxford).

Primiceri, Giorgio E., 2003, "Time Varying Structural Vector Autoregressions and Monetary Policy," unpublished manuscript, Princeton University.

Robertson, John, Ellis Tallman, and Charles Whiteman, 2002, "Forecasting Using Relative Entropy," unpublished manuscript, Research Department, Federal Reserve Bank of Atlanta.

Sargent, Thomas J., 1999, *The Conquest of American Inflation* (Princeton University Press: Princeton, New Jersey).

Stutzer, Michael J., 1996, "A Simple Nonparametric Approach to Derivative Security Valuation," *Journal of Finance* 51, pp. 1633-1652.