working paper

11 28

A Medium Scale Forecasting Model for Monetary Policy

by Kenneth Beauchemin and Saeed Zaman

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 11-28 October 2011

A Medium Scale Forecasting Model for Monetary Policy

by Kenneth Beauchemin and Saeed Zaman

This paper presents a 16-variable Bayesian VAR forecasting model of the U.S. economy for use in a monetary policy setting. The variables that comprise the model are selected not only for their effectiveness in forecasting the primary variables of interest, but also for their relevance to the monetary policy process. In particular, the variables largely coincide with those of an augmented New-Keynesian DSGE model. We provide out-of sample forecast evaluations and illustrate the computation and use of predictive densities and fan charts. Although the reduced form model is the focus of the paper, we also provide an example of structural analysis to illustrate the macroeconomic response of a monetary policy shock.

Key words: Bayesian Vector-Autoregression, Forecasting, Monetary Policy.

JEL classification: C11; C32; C53; E37.

Ken Beauchemin (corresponding author) is at the Federal Reserve Bank of Minneapolis. He can be reached at ken.beauchemin@mpls.frb.org. Saeed Zaman is at the Federal Reserve Bank of Cleveland. He can be reached at Saeed.Zaman@clev.frb.org. Ken Beauchemin gratefully acknowledges that much of the research reported in this paper was conducted while he was a member of the Research Department at the Federal Reserve Bank of Cleveland. The authors also thank Todd Clark for comments and advice.

1 Introduction

Lately, there has been a resurgence of interest in interest in using Bayesian vector autoregression (VAR) models for forecasting and policy analysis. Much of this enthusiasm has been generated by studies demonstrating the strong forecasting performance of computationally convenient versions of the 'Minnesota prior' that dates back to Doan, Litterman, and Sims (1984) and Litterman (1986).

In an interesting twist that reversed years of untested folk wisdom, Bańbura, Giannone, Reichlin (2010) demonstrated that large Bayesian VAR models–ones with more than a handful of variables–outperform their smaller counterparts provided that the priors of the larger models are sufficiently tight. They also show that large Bayesian VAR models perform well against existing large data set methods, in particular, the factor-augmented vector autoregressions (FAVAR) models proposed by Bernanke and Boivin (2003). Koop (2010) extends the results of Bańbura, et. al. (2010), showing that the large Bayesian vector autoregressions also compare favorably to more complicated and computationally intensive prior specifications.

In light of this work, this paper sets forth a sixteen variable Bayesian VAR forecasting model designed to be used in a monetary policy setting. Until recently, sixteen variables would have been at the very edge of what researchers would have considered appropriate considering the problems of collinearity and overfitting. But as recent work helps to underscore, forecast accuracy is also compromised by artificially small systems that introduce errors due to misspecification. As in Bańbura, et. al. (2010) and Koop (2010), we implement a natural-conjugate version of the Minnesota prior (Kadiyala and Karlsson, 1997; Sims and Zha, 1998) that obviates the need for computationally expensive simulation methods for model estimation. This feature is particularly appealing in a real-time policy environment where forecasts are updated on a regular basis. We also make use of the "sum-of-coefficients" prior of Sims (1992).

Because our model is designed to be useful in a monetary policy context, it includes the four variables that must be forecast by all members of the Federal Open Market Committee (FOMC): real GDP, the unemployment rate, personal consumption expenditures (PCE) prices, and core PCE prices (PCE prices less food and energy prices). Our selection of the remaining variables is motivated by a combination of macroeconomic theory and out-of-sample forecast performance, with an emphasis on the former. In particular, we appeal to the “New-Keynesian” (NK) general equilibrium framework to fill out the model. Although our focus here is on the reduced-form models to be used primarily for unconditional forecasting, theory-inspired variable selection facilitates unconditional forecasting experiments that may help “inform causal hypotheses” (Doan, Litterman, and Sims; 1984) of monetary policy and the business cycle. The variable selection also provides a basis for choosing informative over-identifying restrictions for structural analysis. We provide examples of both unconditional forecasting and structural experiments. For the latter, we examine the effect of a contractionary monetary policy shock in the vein of Christiano, Eichenbaum, and Evans (1999).

Like Doan, Litterman, and Sims (1984), our forecast procedure selects hyperparameter values that control the assertiveness of the priors to maximize out-of-sample forecasting performance. In particular, we choose hyperparameter values that maximize the marginal likelihood of the data. The primary reason for this choice is that these hyperparameters can be shown to minimize one-step-ahead forecast errors (Geweke and Whiteman, 2006), and considering the persistence of macroeconomic time series, these forecasts can also be expected to perform well over longer forecast horizons. Giannone, Lenza, and Primiceri (2010) recommend this procedure on the based on out-of-sample forecasting performance that is superior to that of using noninformative (or “flat”) priors or factor methods over multiple forecast horizons. We have also experimented with optimization over various marginal distributions from the predictive density, but find this method problematic for a number of reasons to be discussed herein.

Our results indicate that the Bayesian methods generally and markedly improve forecasting performance relative to naive random walk (with drift) forecasts and those produced by an identical model with flat priors. A small exception is for forecasts of the federal funds rate in which a random walk produces forecasts of similar quality at very short forecast horizons, which we conjecture is the result of Federal Reserve procedures that maintain the federal funds rate constant for often lengthy periods of time–precisely the situations in which one could expect random walk forecasts to do well. Consistent with that observation, the Bayesian model outperforms the random walk in forecast horizons of a year or more.

With respect to structural analysis, the model produces responses to a contractionary monetary policy shock generally in line with those obtained by Christiano, et. al., with the exception of those for the price variables. The pre-established results that output prices fall in response to the shock, perhaps after a small initial increase. The tendency for prices to rise initially and has been dubbed the "price puzzle." In contrast to the consensus view, the price response generated by our model makes the puzzle even more puzzling. We find that the initial price increase that frames the price puzzle is less pronounced but very long lasting. But the magnitude of the response is small enough to be fairly characterized as non-existent. In consideration of the forecasting success of the model with optimized priors over the identical one with flat priors, we conjecture that the established view on the price response may be the result of overfitting, but leave the investigation to future work.

The rest of the paper proceeds as follows. In the following section, we present the VAR model and the Bayesian prior assumptions. In section 3 we motivate the selection of variables, and section 4 discusses the hyperparameter selection. In section 5 we evaluate the forecast performance of the model. Section 6 illustrates the computation and uses of predictive densities and fancharts as measures of forecast uncertainty, and provides two example conditional forecasting experiments. Section 7 presents the structural analysis and section 8 concludes.

2 Model and Prior Specification

To begin, let $Y_t = (y_{1,t}, y_{2,t}, \dots, y_{n,t})'$ be the data vector of n random variables. The model, or likelihood, is defined by the VAR(p) model

$$Y_t = B_c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + \varepsilon_t \tag{1}$$

for $t = 1, \cdots, T$, where $B_c = (c_1, c_2, \dots, c_n)'$ is an n-dimensional vector of constants, $B_1, \dots, B_n$ are $n \times n$ matrices of VAR coefficients, and ε_t is an n-dimensional Gaussian white noise process with $\mathrm{E}\varepsilon_t \varepsilon_t' = \Sigma$. Each equation in this n-variable system has $k = np + 1$ regressors. For our quarterly model, $n = 16$ and $p = 4$ meaning that 65 coefficients per equation must be estimated using some 200-plus observations.

Because the VAR model is richly parameterized, the limited data history means that it is susceptible to overfitting. The traditional solution to the overfitting problem is to sharply restrict the number of variables in the system, risking misspecification instead. Both overfitting and misspecification compromise forecasting accuracy. To balance these risks, Bayesian methods are often applied to shrink coefficient estimates in $B_1, \ldots, B_n$ and Σ to their prior means. We use the Minnesota-type prior introduced by Litterman (1986), but with the Normal-inverted Wishart (N-iW) form proposed by Kadiyala and Karlsson (1997) and Sims and Zha (1998). Among other advantages, the Normal-inverted Wishart prior is also the natural conjugate prior (i.e. produces a posterior that is also Normal-inverted Wishart).

The prior beliefs regarding the first and second moments of the coefficient matrices are as follows:

$$\mathrm{E}\left[B_k^{(i,j)}\right] = \begin{cases} \delta_i, & \text{if } i = j,\ k = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \mathrm{Var}\left[B_k^{(i,j)}\right] = \lambda^2 \frac{1}{\ell^2} \frac{\sigma_i^2}{\sigma_j^2}, \quad \ell = 1, \ldots, p. \tag{2}$$

In the original Minnesota prior, $\delta_i = 1$ for all variables reflecting a belief that the system is a collection of random walks correlated only through the innovations. For non-trending variables, the setting $\delta_i = 0$ should be used instead producing a system described *a priori* by a mixture of random walk and white-noise processes. These priors retain the original Minnesota prior notion that recent lags are more economically significant than more distant ones by shrinking the scale factor $\frac{1}{\ell^2}$ on the variances as the lag length increases. The hyperparameter λ controls the overall tightness of the prior so that as $\lambda \to 0$ the prior dominates and as $\lambda \to \infty$ the prior becomes increasing non-informative (or "flat") and the coefficients converge to OLS estimates.

The prior specification in (2) differs from the traditional Minnesota prior in that it is symmetric in its treatment of own lags of the dependent variables in each equation and all other lags. In contrast, the traditional specification allows the coefficients on the own lags to play a stronger role than the others by letting the prior variances of own-lag coefficients to be larger. Despite the intuitive attraction, the added flexibility turns out to add little, if anything, to forecast accuracy (Koop, 2010; Carriero, Clark, and Marcellino, 2011). Furthermore, the symmetry of the Normal-inverted Wishart prior facilitates equation-by-equation estimation with the data matrices augmented by specific sets of dummy observations. Estimated properly, the traditional Minnesota system requires computationally-expensive posterior simulation methods although equation-by-equation estimation is often (and improperly) used. Due to its properties we will alternative refer to the Normal-inverted Wishart prior as the "natural conjugate" prior or the "symmetric Minnesota" prior.

A final difference between the natural conjugate prior and the traditional Minnesota prior as implemented by Litterman (1986) involves the treatment of the error-covariance matrix of innovations Σ. Whereas the natural conjugate prior treats the elements of Σ as objects to be estimated, the traditional implementation of the Minnesota prior assumes Σ to be a fixed diagonal matrix where each diagonal element is given by the OLS standard regression error estimate of the univariate AR(p) process for the corresponding variable. Although we estimate Σ, we follow the standard practice of setting the scaling parameters σ_i^2 in (2) in accordance with the traditional implementation.

To describe the prior and posterior distributions, it is useful to express the VAR system (1) more compactly as

$$Y = XB + \varepsilon \tag{3}$$

Here, Y is a $T \times n$ matrix of dependent variable observations and X is a $T \times k$ matrix of independent variable observations with $k = np + 1$. In this formulation, the t-th row of Y is the $1 \times n$ vector of dependent variables y_t' and the t-th row of X is the $1 \times k$ vector $\left(1, y_{t-1}', \ldots, y_{t-p}'\right)$. The matrix $B = [B_c \; B_1 \; \cdots \; B_p]$ is the corresponding $k \times n$ matrix of VAR coefficients and ε is the $T \times n$ matrix of disturbance terms with the t-th row given by ε_t'.

The prior can now be written:

$$vec\,(B)\,|\Sigma \sim N\left(vec\,(B_0)\,, \Sigma \otimes \Omega_0\right), \qquad \Sigma \sim iW\left(S_0, \upsilon_0\right) \tag{4}$$

where the prior parameters B_0, Ω_0, S_0, and υ_0 are chosen to produce a VAR system (1) satisfying the prior moment conditions in (2). Multiplication by the likelihood function produces the corresponding posterior kernel

$$vec\,(B)\,|\Sigma, Y \sim N\left(vec\left(\overline{B}\right), \Sigma \otimes \overline{\Omega}_0\right), \qquad \Sigma \sim iW\left(\overline{S}, \overline{\upsilon}\right). \tag{5}$$

where

$$\overline{B} = \left(\Omega_0^{-1} + X'X\right)^{-1}\left(\Omega_0^{-1}B_0 + X'Y\right)$$

$$\overline{\Omega} = \left(\Omega_0^{-1} + X'X\right)^{-1}$$

$$\overline{v} = v_0 + T.$$

Additionally,

$$\overline{S} = \widehat{B}'X'X\widehat{B} + B_0'\Omega_0^{-1}B_0 + B_0 + \widehat{\varepsilon}'\widehat{\varepsilon} - \widehat{B}'\overline{\Omega}^{-1}\widehat{B}$$

where $\widehat{B} = (X'X)^{-1}X'Y$ is the OLS estimate of B and $\widehat{\varepsilon} = Y - X\widehat{B}$ are the OLS residuals (Zellner, 1973). To implement the natural conjugate prior we follow the mixed estimation method proposed by Litterman (1986) and augment the data matrices with dummy observations.

We also evaluate the model using the sum-of-coefficients prior (Sims,1992; Sims and Zha 1998) in addition to the symmetric Minnesota prior (Robertson and Tallman, 1999; Bańbura, *et. al.* 2010). The sum-of-coefficients prior allows for the inexact differencing of variables in the system by imposing the restriction

$$B_1 + \cdots + B_n = I_n$$

implying that the VAR coefficients on own lags sum to one. Letting μ be the hyperparameter controlling the degree of tightness applied to the sum-of-coefficients prior, as $\mu \to 0$ the system approaches one estimated in differenced form; as $\mu \to \infty$ the effect of the prior on the posterior vanishes. To implement both priors, an additional set of dummy observations is added to the data matrices previously augmented to implement the symmetric Minnesota prior. The Appendix presents expressions for the dummy observations necessary to implement both priors.

3 Variable Selection

We approach the problem of variable selection from two perspectives. First, the model is to be used primarily for unconditional forecasting for monetary policy so that we include the four variables for which each FOMC member must submit a forecast. Our secondary objective, is to choose variables which help construct a theoretically coherent forecast narrative, or at least a partial understanding of why forecasts changes over time. Story-telling ability, albeit limited, can be obtained by unconditional forecasting exercises conducted with the reduced-form model. Furthermore, Sims and Zha (1998) show that the reduced form is a special case of a larger set of models and that the natural conjugate prior works with a particular class of identifying structures. As a consequence, a careful selection of variables motivated by theoretical considerations should help generate more useful identification schemes than variables chosen solely for their forecasting ability.

Given that our model is to be used in a monetary policy setting, we choose variables that can be associated with the New-Keynesian (NK) class of dynamic stochastic general equilibrium models (DSGE)–models that have been gaining influence among policymakers at the world's major central banks.[1] Table 1 presents the sixteen model variables beginning with the four that must be forecast: real GDP, the unemployment rate, personal consumption expenditures (PCE) prices, and PCE prices excluding food and energy prices, commonly referred to as "core" PCE prices. Real GDP, as the most comprehensive measure of aggregate output, is an obvious choice for FOMC monitoring. The year-over-year percent change in the core PCE is currently the FOMC's preferred measure for gauging the inflation rate that underlies the more noisy over PCE inflation rate. Nevertheless, energy and food prices also impact consumer welfare and potentially core prices, so forecast for the overall rate are also mandatory. Finally, the unemployment rate provides indirect evidence on capacity utilization, of interest not only for welfare considerations, but for its potential to influence inflation dynamics.

To these four variables, we add the federal funds rate–the primary operating instrument of monetary policy. In NK models, a monetary policy reaction function (e.g. the Taylor rule) closes the model and determines the equilibrium federal funds rate. The addition of hourly labor compensation and labor productivity fill out the core of the model. Inflation dynamics in NK models depend largely on some notion resource utilization typically expressed as an "output gap," or the distance between actual output and its flexible-price analogue. In basic NK models, the output gap is proportional to the real marginal cost of labor. To allow for these inflation mechanics, we include labor compensation and productivity measures which together imply a measure of unit labor costs. The price indexes in the model further imply measures of

[1] See Gali, ch. 3, for an introduction to the basic model.

real unit labor costs.

To this seven-variable core we add nine more variables. Inflation dynamics are not only impacted by labor input costs, but also by the cost material inputs to firms which we proxy with the Commodity Research Bureau's index of commodity prices. Although there is no direct measure of the productivity of materials, we suspect that labor productivity provides a useful proxy measure. As with labor costs, the two PCE indexes imply approximate real commodity price measures.

In addition to real GDP, we round out the real sector of the model with real consumption, real personal disposable income, and nonfarm payroll employment. Distinguishing consumption from total output is not only proper from a DSGE modeling perspective, it implicitly provides information on the smaller but more volatile expenditure components of fixed investment, inventory investment, and net exports. Real disposable income introduces fiscal policy information on taxes and transfers and also facilitates the computation of a personal saving rate. And, aside from being one of the most heavily monitored economic indicators, the inclusion of payroll employment captures firm and household employment decisions on the extensive margin. Information on the intensive margin, or hours worked, is implied by the simultaneous consideration of real GDP (output) and labor productivity (output divided by hours).

Finally we introduce a number of financial market variables, mainly because of their demonstrated ability to predict changes in real activity. Considering that investment decisions are more tied to longer-term interest rates, we introduce yields on 10-year U.S. Treasury notes and Aaa-rated corporate bonds. These provide information on term spreads (10-year Treasury yield minus the federal funds rate) and credit spreads (Aaa corporate yield minus 10-year Treasury yield). We also include the S&P 500 index of equity prices and the S&P 500 dividend yield. Summing these two gives the return on equities which can be compared to either the federal funds rate or the 10-year Treasury yield to extract measures of the equity premium. Finally, we include a nominal trade-weighted exchange rate as a key international variable.

Table 1 also indicates how each of the variables is transformed for the Bayesian VAR. All variables enter in log-levels with the exception of the interest rates which enter in levels. Table 1 also indicates that for all variables the first own-lag coefficient is shrunk to one in the symmetric Minnesota prior, i.e. $\delta_i = 1$ for all $1 = 1, 2, \ldots 16$.

The raw data in this paper runs from 1959Q1 through 2011Q2. Variables available at the monthly frequency are transformed to quarterly by averaging over the three months of the quarter. With the exception of financial variables, all variables are seasonally adjusted.

4 Hyperparameter Selection

Considering the complexity of economies and the number of historical correlations in the VAR models that represent them, it is difficult to imagine using purely introspective methods to select hyperparameters. To bypass that obstacle, the literature suggests several empirical Bayesian approaches to hyperparameter choice. The most popular approach for forecasting applications involves choosing hyperparameters that optimize some pre-defined criteria.

We have experimented with two general optimization methods of hyperparameter selection: 1) minimizing out-of-sample forecast error metrics based on various marginal distributions of the posterior predictive density, and 2) optimizing over the marginal likelihood of the data. Choosing hyperparameter settings based on minimizing density-weighted forecast errors is problematic for a number of reasons. One must first decide how to weight the marginal predictive densities for each of the model variables that comprise the objective function. In our experiments, we chose hyperparameters that maximized forecast performance for each of the four variables reported to the FOMC singly (real GDP, unemployment rate, PCE deflator, and core PCE deflator), and an arithmetic average of the four. As a general characterization, the resulting hyperparameter settings vary markedly across the five weighting schemes. In addition, optimal hyperparameters varied not only by weighting scheme, but also by forecast horizon. In other words, settings that work well for the near-term do not necessarily perform well over longer forecast horizons. Finally, it is our experience that the hyperparameters chosen on out-of-sample forecasting ability can vary widely depending on the data sample, making it difficult to understand why forecasts change from period to period.

In light of these observations, we prefer to use hyperparameters that maximize the marginal likelihood of the data sample. As an in-sample construct, it is not susceptible to the problems enumerated above. And, because these priors maximize the probability that the model is the true model, it gives the best chance of producing a zero one-step-ahead forecast error (Geweke and Whiteman, 2006). Considering the persistence of macroeconomic time series, optimal one-step-ahead performance is likely to carry over to longer forecast horizons. Giannone, Lenza, and Primiceri (2010) provide support for these arguments by demonstrating the superiority of the marginal likelihood approach in out-of-sample forecasting exercises when compared to identical models that use flat priors or models using factor methods over multiple forecast horizons.

In principle, the optimal hyperparameters are given by

$$[\lambda^*, \mu^*] = \underset{[\lambda,\mu]\in\mathbb{R}^2_+}{\arg\max} \ln p\,(Y)\,.$$

where

$$p(Y) = \int p(Y|\theta)p(\theta)d\theta$$

is the marginal likelihood. An attractive feature of the marginal density associated with Normal-inverted Wishart prior is that it can be obtained in closed-form as

$$\begin{aligned} p(Y) \;=\; & (\frac{1}{\pi})^{\frac{nT}{2}} \times |(I + X\Omega_0 X\prime)^{-1}|^{\frac{n}{2}} \times |S_0|^{\frac{v_0}{2}} \\ & \times \frac{\Gamma_n(\frac{v_0+T}{2})}{\Gamma_n(\frac{v_0}{2})} \times \left|S_0 + (Y - XB_0)'(I + X\Omega_0 X\prime)^{-1}(Y - XB_0)\right|^{-\frac{v_0+T}{2}}. \end{aligned}$$

(See Clark, et. al. (2010) or Giannone, et. al. (2010) for derivations.) In practice, we evaluate the marginal likelihood over a discrete grid of candidate hyperparameters defined by $\lambda \in$ [0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 1] and $\mu \in$ [0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.75, 1, 2, 3, 5, 10, 15] for each forecasting exercise.

To establish the claim that the marginal likelihood approach generates stable hyperparameters, we compute the optimal set recursively from 1970Q1 to 2011Q2 (i.e. the 1970Q1 values use the 1959Q1–1969Q4 sample, those for 1970Q2 use the 1959Q1–1970Q1 sample and so on). Figure 1 display the optimal values of λ, and μ over time. As can be seen, they are roughly constant after the sample size grows beyond a certain size corresponding to a sample ending roughly in 1984 for λ and in 1983 for μ, converging approximately to 0.20 and 0.25, respectively. It is interesting to note that both remain essentially constant throughout the 2007Q4–2009Q2 recession period and subsequent recovery. In spite of the rapidly changing economic conditions of the time, this procedure assigns virtually no forecast error to changes in the model hyperparameters from period-to-period.

5 Forecast Evaluation

Our main purpose in this section is to evaluate the forecasting performance of our medium-scale model using standard accuracy metrics for point forecasts. In what follows, we restrict our attention to the four model variables *GDPR*, *UR*, *PCXFE*, *and RFF*. Note that this set is not identical to the one for which FOMC members are required to submit forecasts. From that set we have dropped the overall PCE price (PC) index because it is highly collinear to the core PCE price index and therefore unlikely to add little independent insight on the forecast performance of the model. In its place, we have added the effective federal funds rate (*RFF*) in light of its key role in monetary policy.

Given the posterior mean of the parameters, computing H-step-ahead point forecasts is straightforward. In terms of the H-step ahead predictive density, it is the forecast path generated by a draw of model parameters from the posterior density (5) corresponding to the mean $\overline{B}$ and zero disturbance errors. Using the VAR system (1), the one-step ahead forecast is obtained as

$$\widehat{Y}_{t+1} = \overline{B}_c + \overline{B}_1 Y_t + \overline{B}_2 Y_{t-1} + \ldots + \overline{B}_p Y_{t-p+1}$$

The remaining $h = 1, \ldots, H-1$ step-ahead point forecasts are computed by recursive substitution. For example, the $H = 2$ forecast is then computed

$$\widehat{Y}_{t+2} = \overline{B}_c + \overline{B}_1 \widehat{Y}_{t+1} + \overline{B}_2 Y_t + \ldots + \overline{B}_p Y_{t-p+2}$$

and so on. More generally,

$$\widehat{Y}_{t+h} = \overline{B}_c + \overline{B}_1 \widehat{Y}_{t+h-1} + \overline{B}_2 \widehat{Y}_{t+h-2} + \ldots + \overline{B}_p \widehat{Y}_{t+h-p}, \quad h = 1, \ldots, H \tag{6}$$

where $\widehat{Y}_{t+h} = Y_{t+h-p}$ for $h \leq p$.

The forecast evaluations are conducted on a recursive basis in which the sample period is lengthened by one observation for each forecast. The maximum forecast horizon is two years or $H = 8$ quarters. We will evaluate the model's ability to forecast each of the four variables of interest from the first quarter of 1970 (1970Q1) onwards. Given that our data set runs from 1959Q1 to 2011Q2 and that our model uses $p = 4$ lags, the initial estimation sample runs from 1960Q1 to 1969Q4, followed by the 1960Q1–1970Q1 period, an so on.[2] As in other recent studies, the evaluations are conducted in "pseudo real time" meaning that we do not use vintage (or "real time") data , i.e. data that was available when a forecast would have been initially performed. Forecast performance statistics are also reported for the evaluation period running from 1987Q1 to 2010Q4. This sub-period is often singled out as one characterized by a single and distinct monetary policy regime and therefore worthy of separate consideration. We provide the results to facilitate comparisons to other studies.

Point forecast evaluations are based on the mean square forecast error (MSFE) statistic. Letting T_0 denote the beginning of the evaluation period minus one period (1969Q4 or 1986Q4) and T_1 the end period (2011Q2), the mean square forecast error is given by the expression

$$MSFE_{i,h} = \frac{\sum_{t=T_0}^{T_1-h}(Y_{i,t+h}^{data} - \widehat{Y}_{t+h})^2}{T_1 - h - T_0 + 1}$$

for each forecast variable $i \in \{GDPR, UR, PCXFE, RFF\}$ and forecast horizon $h = 1, \ldots, H$. Note that variables entering the VAR as log-levels are exponentiated to return them to raw levels before computing the MSFEs. We report MSFE statistics relative to three different benchmark model forecasts. The first, is produced by a model comprised of independent random walks with drift (denoted RW), i.e. a model in which the VAR system (1) is restricted such that B_1 is the identity matrix, $B_2 = \cdots = B_p = 0$, and B_c is the vector of drift terms estimated by ordinary least sqares. The second benchmark model is the univariate AR(4) process for each variable that enters the VAR (denoted AR). Each univariate model forecast is produced using the same procedure as the Bayesian VAR forecast. Specifically, the AR(4) models are estimated

[2] We also perform forecast evaluations using a rolling sample of 10 years (40 quarters) starting from the initial sample period of 1960Q1–1969Q4, followed by 1960Q2–1970Q1, and so on. The results prove to be quite similar to those generated with the recursive scheme and are not presented here for the sake of brevity, but are available upon request from the authors.

under the Minnesota prior (in this case the symmetry property is not meaningful) and the sum-of-coefficients prior with hyperparameter setting that maximize the marginal likelihood. This comparison evaluates the forecast value of the cross-correlation information contained in Bayesian VAR model coefficients. And lastly, we evaluate the forecasting accuracy gained by using our informative priors by comparing the Bayesian VAR forecasts to those generated by the same system under flat priors (OLS). Defining $MSFE_{i,h}^{BVAR}$ as the MSFE for the Bayesian VAR with optimized hyperparameters and $MSFE_{i,h}^{m}$ where $m \in \{RW, AR, OLS\}$ as the ones corresponding to each benchmark model, the relative mean squared forecast error (RMSFE) statistic is expressed as the ratio of the former to the latter,

$$RMSFE_{i,h} = \frac{MSFE_{i,h}^{BVAR}}{MSFE_{i,h}^{m}} \ ,$$

so that values less than one imply superior forecasts from the optimized Bayesian VAR model. Since forecasts for trending variables are typically reported as growth rates, the tables below also report RMSFEs for the annualized quarterly growth rates implied by the forecasts of real GDP and the core PCE in addition to the four variables enumerated above.

Table 2 reports the relative mean square forecast errors with the model that implements only the symmetric Minnesota, or equivalently, a model in which μ is set to an arbitrarily large number so that the sum-or-coefficients prior is essentially inoperative. Each forecast run implements the hyperparameter λ that maximizes the marginal likelihood in sample. Table 3 reports the same set of statistics for the model where both hyperparameters are optimized. Comparing the two tables reveals evidence largely in favor of adding the sum-of-coefficients prior. Forecasts for real activity improve substantially with sizeable accuracy gains recorded for both real GDP (in levels and growth rates) and the unemployment rate. Note however, that the real GDP growth forecast beats the random walk forecast only in the first two forecast quarters in the full 1970Q1–2011Q2 evaluation period, and the first four quarters in the post–1986 period. Forecasts for the federal funds rate are dramatically improved, although the improved forecast is only about as accurate as the random walk for the first four quarters in the 1970Q1–2011Q2 evaluation period, and substantially worse in the post–1986 period. Bańbura, et. al. (2010) and Koop (2010) report similar results for the federal funds rate. We conjecture that this result is an artifact of the Federal Reserve operating procedure that changes the target federal funds rate on an infrequent basis which often leaves effective rate roughly unchanged for lengthy periods–precisely the conditions in which a random walk forecast could be expected to do well. But for longer forecast horizons, the auto- and cross-correlations of the Bayesian VAR help to outperform the random walk. The core PCE forecast both in levels and in growth rates are the only ones in which accuracy is not improved by the sum-of-coefficients prior, but the accuracy loss is minor on both counts.

In Table 4, we report the RMSFEs generated by the Bayesian VAR model and the Bayesian AR(4) benchmark models. The results generally mimic those reported for the random walk comparison in Table 3 with two exceptions. As can be expected, the RMSFEs are generally higher in Table 4 simply indicating that as a generalization of the random walk with drift, the AR(4) specification captures higher-order dynamics and that the small number of lags coupled with the informative priors can reasonably be expected to neutralize the disadvantages caused by overfitting. The other exception is that the Bayesian AR(4) forecasts for the core PCE and its growth rate are far superior to the Bayesian VAR analogues at all forecast horizons. This suggests to us that the symmetry property that treats the prior precision of own-lags may be overly restrictive for the case of aggregate price and inflation forecasting. If so, a case could possibly be made for the traditional Minnesota prior, or even a more flexible specification, but only by sacrificing the computational conveniences of using the Normal-inverted Wishart conjugate prior.

Before leaving this section, we assess the gains from using Bayesian prior information in the first place. Table 5 expresses the forecast accuracy of the Bayesian formulation with both the symmetric Minnesota and sum-of-coefficients prior as in Table 3, relative to forecasts produced with flat priors (λ and μ set to arbitrarily large numbers). Because approximately sixteen years of data is expended estimating the model using flat priors, results are only available for the 1987Q1–2011Q2 period. Comparing Table 4 to Table 3 reveals that the forecasts produced with the informative priors improve overwhelmingly on the flat priors, with the largest accuracy gains recorded for the federal funds rate and the smallest for core PCE prices.

6 Forecast Uncertainty and Conditional Projections

Point forecasts provide a useful focal point for policy discussions, but a complete rendering of a forecast requires its probabilistic properties. For reasons obvious to professional economists and policymakers, an intelligible decision-making process requires a probabilistic assessment of alternative outcomes. By definition (considering that the real numbers are dense), a point forecast can never be technically correct. Although the point seems trivial, it is easily lost on a public that continually chide forecasters for “always getting it wrong.” Offering a range of likely and not-so-likely outcomes in the monetary policy discussion also brings the potential of improving the monetary policy dialogue with the public. The many advantages of a probabilistic approach to policymaking (and some potential pitfalls) have been discussed at length by Sims (2007). The advantages have become more apparent as central banks move toward either explicit or de-facto inflation targeting regimes.

In this section, we introduce the posterior predictive density, the fundamental forecast object of Bayesian VAR models, and demonstrate its two basic applications: 1) to characterize the extent of forecast uncertainty, and 2) to assess the likelihood of forecasts that satisfy as set of conditions or specific sets of future disturbances.

6.1 The Posterior Predictive Density and Fan Charts

Prediction in a Bayesian framework is based on the posterior predictive density (or “predictive density” for short). The predictive density provides a complete probability assessment of future values of the model variables given current and past observations of those variables.

Let $Y^T = (y_1', \ldots, y_T')'$ represent the entire history of the data Y_t. Given history, we wish to make predictions h periods into the future. If we let $Y^{T+1,T+H} = \left(y_{T+1}', y_{T+2,}', \ldots, y_{T+H}'\right)'$ represent an arbitrary forecast path in the set of all possible future paths, then constructing the predictive density requires us to assign a probability to each path. The predictive density is thus

$$p\left(Y^{T+H} \mid Y^T\right) = \int p\left(Y^{T+1,T+h}, \theta \mid Y^T\right) d\theta \tag{7}$$

where $p\left(Y^{T+1,T+H}, \theta \mid Y^T\right)$ is the joint density of model parameters and future variable observations. Using the rules of probability, the integrand can be written in a way that highlights the two sources of forecast uncertainty:

$$p\left(Y^{T+1,T+H},\theta \mid Y^{T}\right) = p\left(Y^{T+1,T+H} \mid Y^{T},\theta\right) p\left(\theta \mid Y^{T}\right), \tag{8}$$

The first term on the right-hand side of (8) describes the uncertainty on future observables given the observed data and model parameters, or equivalently, the forecast uncertainty due to future disturbances that impact the VAR. The second term is the model posterior distribution describing parameter uncertainty. Since both distributions have analytical expressions under the Normal-inverted Wishart prior, simple Monte Carlo methods are used to produce a numerical analogue of the predictive density.[3]

To use the h-step ahead predictive density to measure the forecast uncertainty of a single forecast variable (h-steps ahead), the marginal predictive density is constructed in by integrating all other VAR variables out of (7) and setting them equal to their expected values. The fan chart representation of the forecast uses the $h = 1, \ldots, H$ marginal predictive densities to illustrate forecast uncertainty over the entire forecast horizon. Figures 2 and 3 provide fan charts constructed with nine density probability bands ranging from 10% to 90% at each point in the forecast horizon with the darker shades indicating higher probabilities. The figure 2 forecast is produced using the 1959Q1–2009Q2 sample, and figure 3 displays the one corresponding to the 1959Q1–2005Q4 subsample. A comparison of the two figures shows that there is more dispersion of forecasts surrounding the point forecast produced from the more recent sample indicating more forecast uncertainty, despite having been produced with more observations–the result of introducing an exceptionally large recession to the historical sample.[4] For example, forecasts for year-on-year real GDP growth for 2006Q4 (i.e. from the perspective of 2005Q4) ranges between 0.4 percent and 4.0 percent with 70 percent probability (figure 3) while the year-on-year growth projection for 2010Q4 (from the perspective of 2009Q4) has a 70 percent range between 1.7 percent and 6.9 percent, or 520 basis points versus 360 basis points in the earlier period. The corresponding figures for 2007Q4 and 2011Q4 are 430 and 552 basis points, respectively. Similar results hold for the unemployment rate, core PCE inflation, and the federal

[3] The details of the algorithm are available from the authors upon request .

[4] One could control for the number of observations by using a rolling sample.

funds rate.[5]

6.2 Conditional Projections

Because the predictive density assigns likelihoods to all possible forecast paths, the Monte Carlo methods used to produce the approximate numerical density are easily applied to gauge the likelihoods of specific sets of economically interesting paths. A simple application would be to compute the probability of a "recession" in the coming year by flagging all forecast paths which contain two or more consecutive quarters of negative growth in the first four quarters of the forecast horizon and subsequently expressing the number of qualifying paths relative to the total number of simulations. Conditional (marginal) predictive densities can then be constructed to evaluate the broader economic conditions that are likely to prevail in a recessionary environment. For example: What is the likely range of inflation rates along the recession paths? How likely is it that the unemployment rate will exceed ten percent in the next two years if there is a recession?

In other cases, policymakers may interested in gauging the likelihood of specific outcomes following an outsized event equivalent to a large forecast error (or errors) produced by one or more of the model's equations. In these cases, the analysis is based on the predictive densities produced by forecasts that condition on assumed constants for one or more variables in a given forecast period or in multiple periods. Although no causal interpretation of the experiment is available, the results may help "inform causal hypotheses" (Doan, Litterman, and Sims; 1984). For example, one may be interested in studying the effects of an unexpectedly large observation in commodity prices (a large forecast error) on a given baseline forecast. We close this section by considering both examples.[6]

[5] The result for the federal funds rate is somewhat artificial since the level specification of that equation allows projections for which the rate is negative. We choose the level specification to facilitate comparison to studies that generally favor using simple levels for asset yields.

[6] Another method for introducing extra-sample information proposed by Robertson, Tallman, and Whitman (2005) use a minimum entropy method to operate directly on the predictive densities to satisfy a set of zero-moment conditions. Although we do not consider it here, we consider it a promising avenue for further research.

6.2.1 Example: Recession probabilities

In this example, we recursively simulate the sequence of predictive densities for $h = 1, \ldots, 4$ steps ahead for each quarter in the range of dates beginning in 1970Q1 and ending 2011Q2. That is, the first four densities are constructed using the 1959Q1–1969Q4 sample period, the next set of four using the 1959Q1–1970Q1 period, and so forth. We then use the marginal predictive densities for real GDP computed at each recursion to assess the probability that the economy will experience two or more consecutive declines in real GDP in the ensuing four quarters. The probabilities are plotted in Figure 4 along with shaded regions depicting NBER-defined recession periods.[7] Although the Bayesian VAR model is not designed to produce recession probabilities like various incarnations of nonlinear models do, Figure 4 shows that the predictive densities are quite informative in forecasting recessions. Nearly all recessions are preceded by sharp increases in the recession probability with few discernible false positives.

6.2.2 Example: A shock to commodity prices

Here, we use the Bayesian VAR model to analyze the effects of a commodity price surge on a given forecast. Numerous rapid increases in energy and other commodity prices that have occurred since the latter part of the1990s make this example especially relevant in current policy environments. The experiment compares two forecasts: the unconditional, or "baseline" forecast, and one that is conditioned on a 25 percent increase in commodity prices (translating approximately to a 100 percent annual rate to facilitate quantitative comparison) in the initial forecast period. The size of the shock is large but not unprecedented when compared to some of the consecutive-quarter shocks witnessed in the mid-2000s. The baseline and conditional forecast densities are constructed for $h = 1, 2, \ldots, 8$ steps ahead using the 1959Q1-2011Q2 data sample.

Figure 5 displays the results of the experiment. One should first note that the commodity price response is highly persistent–although the forecasted commodity price growth path quickly falls in the subsequent periods (as indicated in Figure 5), there are no ensuing periods of negative

[7] These are the probabilities that policymakers would have faced throughout this period if they had access to the most recent data set–the one used thoughout this paper. A more complete analysis would apply the historical data actually available to policymakers at each recursion.

growth to bring the level of commodity prices back down to baseline levels. After an initial rise, growth in real activity slows as indicated by the behavior of real GDP growth, employment growth and the unemployment rate in Figure 5. The decrease in real activity is accompanied by rising inflation. Overall PCE inflation is quickly pushed higher but quickly attenuates, falling back down to the baseline level. Core PCE inflation rises less and more gradually than overall PCE inflation, but the increase is more persistent.

Although cause and effect is not identified by these results, they hue closely to a the story told by the NK-DSGE model about the effects of a truly exogenous supply shock. The supply of a key commodity (e.g. oil) contracts and its price rises. Growth in the productivity of complementary inputs slows–notably labor productivity–reducing output and employment growth. Slower labor productivity growth places upward pressure on unit labor cost which are passed in some measure to output prices over a number of periods. Although the causal chain is not identified by the reduced-form model, it useful to know that the dominant theme of the experiment is in overall agreement with theory. Note, however, that the initial rise in real activity is at odds with the theoretical narrative and indicates that many forces are at work here underscoring the danger of reading too much into conditional forecasts. That stated, the results do stimulate one's thinking regarding causal mechanisms. Is it that periods of rising commodity prices have been associated with rapid economic growth due to strong demand for natural resources that drive up commodity prices? If so, then the experiment's results could be interpreted as the NK-DSGE inflation mechanism layered with commodity price and growth dynamics with effects that linger and dominate at the outset of the experiment.

7 Structural Analysis

As stressed in section 3, variable selection in our Bayesian VAR is motivated by theory as well as forecast performance. In particular, variables were chosen to correspond to those one might find in an expanded version of a NK-DSGE model. Because our model variables are germane to the monetary policy discussion, we consider a structural analysis along the lines of Christiano, Eichenbaum, and Evans (1999) to identify the effects of a monetary policy shock. Whereas they impose their identification scheme on VAR using classical statistical techniques, the identification also works with the natural conjugate prior Bayesian VAR considered here (Sims and Zha, 1998).

Christiano, et. al. (1999) motivate their identification with a monetary policy rule for the central bank,

$$R_t = f\left(\Omega_t\right) + v_t\,, \tag{9}$$

where R_t is the monetary policy instrument (i.e. the federal funds rate), Ω_t is the central bank's information set, v_t is the monetary policy shock, and f is assumed to be linear. The key identifying assumption is that v_t is orthogonal to Ω_t. To implement the scheme, the data vector is partitioned as follows:

$$Y_t = \begin{bmatrix} X_{1t} \\ R_t \\ X_{2t} \end{bmatrix}. \tag{10}$$

Vector X_{1t} contains the variables for which their contemporaneous values appear in Ω_t; these variables are characterized as "slow moving" in that they do not respond contemporaneously to monetary policy shocks. They comprise the first block because they are ordered ahead of the the monetary policy instrument R_t in the VAR. In contrast, the variables in X_{2t} are "fast moving" because they can respond instantaneously to a monetary policy shock. Broadly speaking, real quantities and price levels are classified as slow moving and financial variables as fast moving. That is, monetary policy has an immediate effect on the financial variables, but influences price levels and real quantities with a lag. In terms of the present model variables, the partition is defined by

$$\begin{aligned} X_{1t} \;\; = \;\; & [GDPR_t \; YPDR_t \; CONSR_t \; PCXFE_t \; PC_t \; PCOMM_t \\ & PRODNF_t \; COMPNF_t \; EMPNF_t \; UR_t]'\,, \end{aligned} \tag{11}$$

$$X_{2t} = [RTCM10_t \; RAAA_t \; SP500_t \; SPYIELD_t \; EXCH_t]'\,, \tag{12}$$

and $R_t = FFED_t$. Although, (10)–(12) represents a complete recursive orthogonalization defined by the unique ordering of all variables, the dynamic response of all variables Y_t to a policy shock is unaffected by the ordering of variables within X_{1t} and X_{2t}. Since we are only interested in the effects of monetary policy, the order of variables as written in (11) and (12) is unimportant.

Formally, the identification of monetary policy shocks is achieved by the standard triangular decomposition of the variance covariance matrix: $\Sigma = A_0^{-1} D \left(A_0^{-1}\right)'$. In this expression, A_0 is a lower diagonal matrix with ones along the main diagonal and $D = \text{diag}(\Sigma)$. The structural model may be expressed generically as

$$A_0 Y_t = A_c + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t \tag{13}$$

By comparison to the reduced form (1), we see that $A_c = A_0 B_c$, $A_j = A_0 B_j$, for $j = 1, \ldots, p$, and $u_t = A_0 \varepsilon_t$ is the period-t vector of mutually uncorrelated, or "fundamental," shocks. To compute the impulse responses and the associated probability bands, we sample from the model posterior and for each draw of $(B_1, B_2, \ldots, B_p, \Sigma)$ we compute A_0, the implied coefficient matrices A_j, and the impulse responses (13). The probability bands are computed from the empirical distribution of impulse responses.

Table 6 and Figure 6 summarize the effects of a contractionary 100 basis point shock to the federal funds rate. Christiano, et. al. (1999) summarize the agreement in the literature regarding the properties of such responses. Our results regarding real activity are in broad agreement with previous results: the positive shock federal funds rate is highly persistent; measures of real activity (in our case real GDP, real disposable income, real consumption expenditures and employment) decline in hump-shaped fashion (and in unemployment rate rises with a hump-shape), and productivity declines initially and gradually returns to trend. As in other studies, commodity prices fall after an initial delay, but the decline is not as persistent and returns to trend within three years of the shock.

In another, and much starker contrast, our impulse responses for output prices (overall and core PCE indexes) are essentially flat compared to previous findings which show declining prices after a brief delay, or even and initial increase. Christiano, et. al. (1999) have referred to the paradoxical initial rise as the "price puzzle." If anything, our price responses can be characterized as a super-persistent price puzzle, in which the price responses remain slightly positive for an extended period: after five years, both the response to the overall PCE index and the core PCE index have yet to fall back to zero.

We offer a number of conjectures for the observed difference. First, Christiano, et. al. (1999) use the GDP deflator as there overall price measure which includes nonfinal goods prices which may be inherently more volatile. We prefer the PCE indexes because of their direct relevance to monetary policy. To investigate, we replaced the two PCE indexes in the model with the single GDP deflator and observed little difference in the two price responses. Second, our sample is updated to run through the second quarter of 2011. The 2000's added a number of years of continued price stability with core PCE inflation that averaged roughly two percent per annum. In monetary policy terms, enhanced central bank credibility may have served to more firmly anchor expectations so that observed inflation tolerated larger movements in the federal funds rate. But when we estimate using the data only through 2000, the output price response is little changed. Lastly, our model specification is larger and specified differently than others and uses Bayesian priors, whereas most previous results were produced with smaller and classically-estimated VAR models.

To investigate the last conjecture, we re-estimate the model with flat priors to match the classical techniques used by others. Here we find the likely explanation for the difference. Figure 7 shows that the responses of both the core PCE and the overall PCE price indexes, which go slightly positive initially (the price puzzle), but turn downward after about a year and enter negative after roughly two years, consistent with the consensus result. The contradictory responses raises the question of whether the classically-estimated responses are largely artifacts of overfitting to which VAR systems without shrinkage priors are prone. Or perhaps the likelihood approach of choosing optimal hyperparameters is inappropriate. We are inclined not to think the latter given that maximizing the marginal likelihood also rewards in-sample fit and leads to superior forecasts for the price variables. Nevertheless, these results are insufficient to form a conclusion and further investigation will be required.

It is also worth examining the response of the financial variables to the monetary contraction. As one would expect, the 10-year U.S. Treasury yield rises with the federal funds rate but by a much smaller amount (20 basis points compared to 100) flattening the yield curve and reducing its level–an overall movement qualitatively consistent with an expectations hypothesis of the yield curve.[8] These responses imply a term spread (10-year yield less the federal funds rate) that initially contracts by roughly 80 basis points and gradually returns to the pre-shock level after about 12 quarters. The AAA corporate bond yield, jumps initially less than the 10-year Treasury yield (basis 7 points), but has more of a hump-shaped response rising to roughly 20 basis points after five or six quarters. That means that after an initial 5 basis point contraction, the credit spread (AAA corporate yield minus the 10-year Treasury yield) widens before gradually returning to pre-shock levels. The dividend yield on equities mimics the response of the AAA corporate yield but its magnitude is far smaller. Equity prices fall in the first two quarters and gradually return to their pre-shock level. And finally, the response of the exchange rate is strongly positive with some delay and is highly persistent.

In addition to the impulse responses, Table 6 also reports variance decompositions, that gives the proportion of h-step ahead forecast error variance due to monetary policy shocks for $h = 1, 8, 12, 16$, and 20 quarters ahead. As in Christiano, et. al. (1999), the monetary policy shock is responsible for only a small percentage of inflation but rather large fractions of real activity. Consistent with the relatively muted impulse responses of core and overall PCE, the monetary policy variance contributions to inflation are even smaller than those reported by reported by Christiano, et. al. (1999). Alternatively, the contributions to the nominal bonds yields and the exchange rate are substantial, but not so much for equity prices and dividend yields.

8 Conclusion

Motivated by recent literature demonstrating the forecasting superiority of large Bayesian VAR systems with tight priors over smaller systems and other large dataset methods, this paper constructs a medium-scale Bayesian VAR model with sixteen variables chosen to be useful in the analysis of monetary policy. In particular, the variables are chosen to cohere with those contained in dynamic general equilibrium models of monetary policy. We examine forecast accuracy and report favorable results similar to those in the large Bayesian VAR studies of

[8] Although these are not real returns (for which no identifying assumption can be made here), recall that inflation is little changed by the monetary shock so that the responses of nominal yields are close to those for ex-post real yields.

Bańbura, et. al. (2010) and Koop (2010).

We provide predictive density characterizations of forecast uncertainty and show how they are used to construct conditional predictive densities that provide richer characterizations of uncertain economic environments. In particular, we provide two examples of conditional forecasting: recession forecasting and the forecast consequences of a large commodity price shock.

Finally, we use the model to study the response of the economy to an exogenous monetary policy shock along the line of Christiano, et. al. (1999). Our results generally accord with previously established ones with a single exception: we find that output prices are essentially unaffected by the monetary shock. Our initial investigation suggest that the discord is generated by our application of Bayesian techniques compared to the classical approaches previously adopted, suggesting that the consensus view may be an artifact of overfitting to which classically-estimated VAR models are prone. These results are inconclusive and further investigation is required.

References

[1] Bańbura, M., D. Giannone, and L. Reichlin, 2010. "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics*, 25(1), 71-92.

[2] Bernanke, B. S. and J. Boivin, 2003. "Monetary Policy in a Data-rich Environment," *Journal of Monetary Economics*, 50(3), 525-46.

[3] Canova, F., 1991. "The Sources of Financial Crisis: Pre-and Post-Fed Evidence," *International Economic Review*, 32(3), 689-713.

[4] Carriero, A., Clark, T. E., and M. Marcellino, 2011. "Bayesian VARs: Specification Choices and Forecast Accuracy," *Federal Reserve Bank of Cleveland Working Paper, 11-12*.

[5] Christiano, L. J., M. Eichenbaum, and C. L. Evans, 1999. "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, ed. J. B. Taylor and M. Woodford, vol. 1, chap. 2, pp. 65-148. Elsevier.

[6] Doan, T., R. Litterman, and C. A. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1-100.

[7] Gali, J., 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press.

[8] Geweke, J. and C. Whiteman, 2006. "Bayesian Forecasting," in Handbook of Economic Forecasting, ed. G. Elliott, C. W. J. Granger, and A. Timmerman, vol. 1, chap. 1, pp. 3-80. Elsevier.

[9] Giannone, D., Lenza, M., and Primiceri, G., 2010. "Prior Selection for Vector Autoregressions," manuscript.

[10] Gordon, D. B. and E. M. Leeper, 1994. "The Dynamic Impacts of Monetary Policy: An exercise in Tentative Identification," *Journal of Political Economy*, 102(6), 1228-47.

[11] Kadiyala, K. R. and S. Karlsson, 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12(2), 99-132.

[12] Koop, G., 2010. "Forecasting with Medium and Large Bayesian VARs," manuscript, University of Strathclyde.

[13] Litterman, R., 1986. "Forecasting with Bayesian Vector Autoregressions–Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25-38.

[14] Robertson, J. C. and E. W. Tallman, 1999. "Vector Autoregressions: Forecasting and Reality," *Federal Reserve Bank of Atlanta Economic Review*, (Q1), 4-18.

[15] Robertson, J. C., E. W. Tallman, and C. H. Whiteman, 2005. "Forecasting Using Relative Entropy," Journal of Money, Credit, and Banking, 37(3), 383-401.

[16] Sims, C. A., 1992. "Bayesian Inference for Multivariate Time Series with Trend," mimeo.

[17] Sims, C. A., 2007. "Improving Monetary Policy Models," *Journal of Economic Dynamics and Control*, 32, 2460-75.

[18] Sims, C. and T. Zha, 1992. "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 948-68.

[19] Zellner, A., 1973. *An Introduction to Bayesian Inference in Econometrics*. Wiley.

Appendix: Implementing the Priors

We follow the standard approach and implement the prior distribution using a mixed estimation approach. Specifically, we augment the original data set with dummy observations and estimate the model using equation-by-equation ordinary least squares. Implementing the symmetric Minnesota prior involves adding the following sets of dummy observations to the original data set:

$$Y_{dum1} = \begin{pmatrix} \frac{diag(\delta_1\sigma_1,\ldots,\delta_n\sigma_n)}{\lambda} \\ 0_{n(p-1)\times n} \\ ------------------ \\ diag(\sigma_1,\ldots,\sigma_n) \\ ------------------ \\ 0_{1\times n} \end{pmatrix}$$

$$X_{dum1} = \begin{pmatrix} \frac{J_p\otimes diag(\sigma_1,\ldots,\sigma_n)}{\lambda} & 0_{np\times 1} \\ -------------- & ---- \\ 0_{n\times np} & 0_{n\times 1} \\ -------------- & ---- \\ 0_{1\times np} & \nu \end{pmatrix}$$

The first block of dummies in Y_{dum1} and X_{dum1} implements the prior beliefs regarding the autoregression coefficients, the middle blocks generate the prior for the covariance matrix, and the final blocks implement the noninformative prior for the intercept. Assigning a small value to ν implies that the prior mean for the intercept is essentially zero.

The sum-of-coefficient prior is incorporated by augmenting the data set with the following dummy observations.

$$Y_{dum2} = \begin{vmatrix} \frac{diag(\delta_1\alpha_1,\ldots,\delta_n\alpha_n)}{\mu} \end{vmatrix}$$

$$X_{dum2} = \begin{vmatrix} \frac{(12\ldots p)\otimes diag(\delta_1\alpha_1,\ldots,\delta_n\alpha_n)}{\mu} & 0_{n\times 1} \end{vmatrix}$$

where the parameter μ controls the degree of shrinkage applied to the prior. The parameter α_i is the initial mean of the variable y_{it}.

Tables and Figures

Table 1. Description of the Data Set and Transformation

Code	Series (i)	Transform	δ_i
GDPR	Real GDP (chained 2005 dollars)	log-level	1
UR	Unemployment rate	level	1
PC	PCE price index	log-level	1
PCXFE	Core PCE price index	log-level	1
RFF	Effective federal funds rate	level	1
COMPNF	Nonfarm business compensation	log-level	1
PRODNF	Nonfarm business productivity	log-level	1
CONSR	Real personal consumption expenditures	log-level	1
YPDR	Real personal disposable income	log-level	1
EMPNF	Payroll employment: total nonfarm	log-level	1
PCOMM	KR-CRB spot commodity price index: all commodities	log-level	1
RTCM10	10-year Treasury note yield at constant maturity	level	1
RAAA	Moody's seasoned Aaa corporate bond yield	level	1
SP500	S&P 500 composite stock price index	log-level	1
SPYIELD	S&P 500 composite dividend yield (percent)	level	1
EXCH	Trade-weighted exchange value of US$ vs. major currencies	log-level	1

Table 2. Forecast Comparison: Symmetric Minnesota Prior v. Random Walk

	Relative Mean Squared Forecast Error					
h	$GDPR$	$GDPR$ (%ch)	$PCXFE$	$PCXFE$ (%ch)	RU	FFR
			1970Q1 - 2011Q2			
1	0.751	0.925	0.147	0.146	0.558	1.099
2	0.812	1.011	0.147	0.222	0.617	1.323
3	0.910	1.256	0.157	0.289	0.697	1.468
4	0.984	1.317	0.168	0.340	0.783	1.603
5	1.036	1.562	0.182	0.411	0.853	1.721
6	1.082	1.610	0.201	0.507	0.903	1.887
7	1.126	1.797	0.224	0.580	0.936	2.016
8	1.181	2.084	0.252	0.657	0.973	2.120
			1987Q1 - 2011Q2			
1	0.639	0.717	0.142	0.147	0.504	1.678
2	0.716	0.980	0.151	0.212	0.568	1.940
3	0.810	1.119	0.160	0.243	0.719	1.948
4	0.882	1.189	0.171	0.262	0.869	1.895
5	0.912	1.193	0.181	0.292	0.963	1.855
6	0.931	1.188	0.196	0.351	1.013	1.805
7	0.941	1.216	0.215	0.420	1.038	1.765
8	0.946	1.272	0.239	0.523	1.028	1.749

Notes for the table: The table lists the mean squared forecast error of the Bayesian VAR model relative to mean squared forecast error from the random walk with drift model. It reports the RMSFEs for the real GDP level, real GDP growth (quarterly at annual rate), core PCE level, core inflation (quarterly at annual rate), the unemployment rate, and the federal funds rate for $h = 1, 2, \ldots, 8$ step-ahead forecasts for the evaluation period 1970Q1–2011Q2. The hyperparameter λ maximizes the marginal likelihood at each iteration.

Table 3. Forecast Comparison: Minnesota and Sum-of-Coefficients Priors v. Random Walk

	Relative Mean Squared Forecast Error					
h	$GDPR$	$GDPR$ (%ch)	$PCXFE$	$PCXFE$ (%ch)	RU	FFR
			1970Q1 - 2011Q2			
1	0.659	0.806	0.161	0.157	0.463	0.980
2	0.663	0.851	0.182	0.268	0.473	1.099
3	0.710	1.081	0.213	0.374	0.500	1.017
4	0.746	1.110	0.242	0.442	0.546	0.993
5	0.792	1.294	0.270	0.537	0.599	0.958
6	0.827	1.280	0.296	0.603	0.640	0.905
7	0.868	1.401	0.321	0.673	0.685	0.881
8	0.921	1.576	0.352	0.762	0.746	0.858
			1987Q1 - 2011Q2			
1	0.580	0.635	0.153	0.150	0.383	1.580
2	0.614	0.849	0.185	0.255	0.371	1.701
3	0.667	0.929	0.214	0.312	0.453	1.479
4	0.709	0.982	0.239	0.337	0.545	1.252
5	0.734	1.033	0.254	0.340	0.618	1.086
6	0.748	0.991	0.268	0.363	0.670	0.957
7	0.769	1.029	0.277	0.374	0.712	0.845
8	0.799	1.076	0.292	0.432	0.741	0.758

Notes for the table: The table lists the mean squared forecast error of the Bayesian VAR model relative to mean squared forecast error from the random walk with drift model. It reports the RMSFEs for the real GDP level, real GDP growth (quarterly at annual rate), core PCE level, core inflation (quarterly at annual rate), the unemployment rate, and the federal funds rate for $h = 1, 2, \ldots, 8$ step-ahead forecasts for the evaluation period 1970Q1–2011Q2. The hyperparameters λ and μ maximizes the marginal likelihood at each iteration.

Table 4. Forecast Comparison: Bayesian VAR v. Bayesian AR(4)

Relative Mean Squared Forecast Error						
h	$GDPR$	$GDPR$ (%ch)	$PCXFE$	$PCXFE$ (%ch)	RU	FFR
			1970Q1 - 2011Q2			
1	0.855	0.919	1.253	0.994	0.740	0.892
2	0.852	0.900	1.517	1.059	0.682	0.980
3	0.853	1.077	1.642	1.006	0.641	0.906
4	0.862	1.099	1.626	0.909	0.640	0.906
5	0.886	1.263	1.565	0.873	0.674	0.864
6	0.911	1.243	1.493	0.801	0.722	0.844
7	0.943	1.345	1.409	0.749	0.778	0.819
8	0.992	1.520	1.333	0.764	0.849	0.794
			1987Q1 - 2011Q2			
1	0.823	0.861	1.594	1.495	0.687	1.827
2	0.841	0.991	2.087	1.785	0.734	1.694
3	0.834	0.956	2.482	1.903	0.781	1.498
4	0.845	0.998	2.680	1.887	0.860	1.279
5	0.843	1.039	2.794	1.827	0.926	1.107
6	0.846	0.998	2.814	1.744	0.995	0.982
7	0.861	1.036	2.809	1.614	1.034	0.878
8	0.889	1.083	2.663	1.563	1.065	0.799

Notes for the table: The table lists the mean squared forecast error of the Bayesian VAR model relative to mean squared forecast error from the corresponding Bayesian AR(4). It reports the RMSFEs for the real GDP level, real GDP growth (quarterly at annual rate), core PCE level, core inflation (quarterly at annual rate), the unemployment rate, and the federal funds rate for $h = 1, 2, \ldots, 8$ step-ahead forecasts for the evaluation period 1970Q1–2011Q2. The hyperparameters λ and μ maximizes the marginal likelihood at each iteration in both VAR and AR(4) models.

Table 5. Forecast Comparison: Informative v. Flat Priors

h	$GDPR$	$GDPR$ (%ch)	$PCXFE$	$PCXFE$ (%ch)	RU	FFR
Relative Mean Squared Forecast Error						
1987Q1 - 2011Q2						
1	0.435	0.355	0.624	0.574	0.374	0.234
2	0.412	0.371	0.769	0.733	0.342	0.231
3	0.421	0.425	0.785	0.666	0.380	0.256
4	0.459	0.491	0.799	0.643	0.394	0.246
5	0.481	0.559	0.762	0.524	0.414	0.221
6	0.497	0.477	0.709	0.455	0.443	0.201
7	0.510	0.439	0.657	0.429	0.465	0.187
8	0.507	0.460	0.621	0.419	0.473	0.181

Notes for the table: The table lists the mean squared forecast error of the Bayesian VAR model relative to mean squared forecast error from the VAR estimated with flat priors. It reports the RMSFEs for the real GDP level, real GDP growth (quarterly at annual rate), core PCE level, core inflation (quarterly at annual rate), the unemployment rate, and the federal funds rate for $h = 1, 2, \ldots, 8$ step-ahead forecasts for the evaluation period 1987Q1–2011Q2. The hyperparameters λ and μ maximizes the marginal likelihood at each iteration.

Table 6. Impulse Responses and Variance Decomposition: Monetary Policy Shock

Vars/Horizon	Impulse Responses						Variance Decomposition					
	0	4	8	12	16	20	1	4	8	12	16	20
GDPR	0.00	-0.31	-0.51	-0.55	-0.52	-0.46	0.0	2.5	6.4	9.2	10.4	10.8
YPDR	0.00	-0.16	-0.24	-0.25	-0.23	-0.19	0.0	1.1	2.0	2.5	2.6	2.5
CONSR	0.00	-0.31	-0.43	-0.43	-0.39	-0.34	0.0	4.4	7.1	8.1	8.2	7.9
PCXFE	0.00	0.05	0.12	0.14	0.12	0.08	0.0	0.2	0.7	0.7	0.5	0.4
PC	0.00	0.08	0.12	0.10	0.05	0.01	0.0	0.5	0.6	0.4	0.3	0.2
PCOMM	0.00	-0.12	-1.18	-1.65	-1.67	-1.54	0.0	0.0	0.7	2.0	2.9	3.3
PRODNF	0.00	-0.16	-0.10	-0.03	0.00	0.00	0.0	1.4	1.4	1.0	0.7	0.6
COMPNF	0.00	0.06	0.09	0.05	-0.02	-0.08	0.0	0.1	0.3	0.2	0.1	0.1
EMPNF	0.00	-0.18	-0.47	-0.60	-0.62	-0.59	0.0	1.9	7.2	13.1	17.8	20.9
UR	0.00	0.11	0.25	0.29	0.27	0.24	0.0	2.0	8.1	14.0	17.6	19.3
RFF	1.00	0.66	0.36	0.15	0.07	0.04	83.0	53.8	37.6	32.0	29.6	28.2
RTCM10	0.21	0.16	0.12	0.08	0.05	0.03	11.6	8.3	6.5	5.2	4.3	3.7
RAAA	0.15	0.16	0.15	0.12	0.09	0.06	10.6	9.2	8.6	7.5	6.6	5.9
SP500	-0.67	-0.48	-0.49	-0.15	0.12	0.31	0.7	0.9	0.7	0.5	0.4	0.4
SPYIELD	0.02	0.03	0.01	-0.01	-0.03	-0.03	0.8	1.9	1.3	1.0	1.0	1.1
EXCH	0.50	0.99	1.59	1.87	1.97	1.99	2.0	3.9	7.3	10.4	12.5	13.8

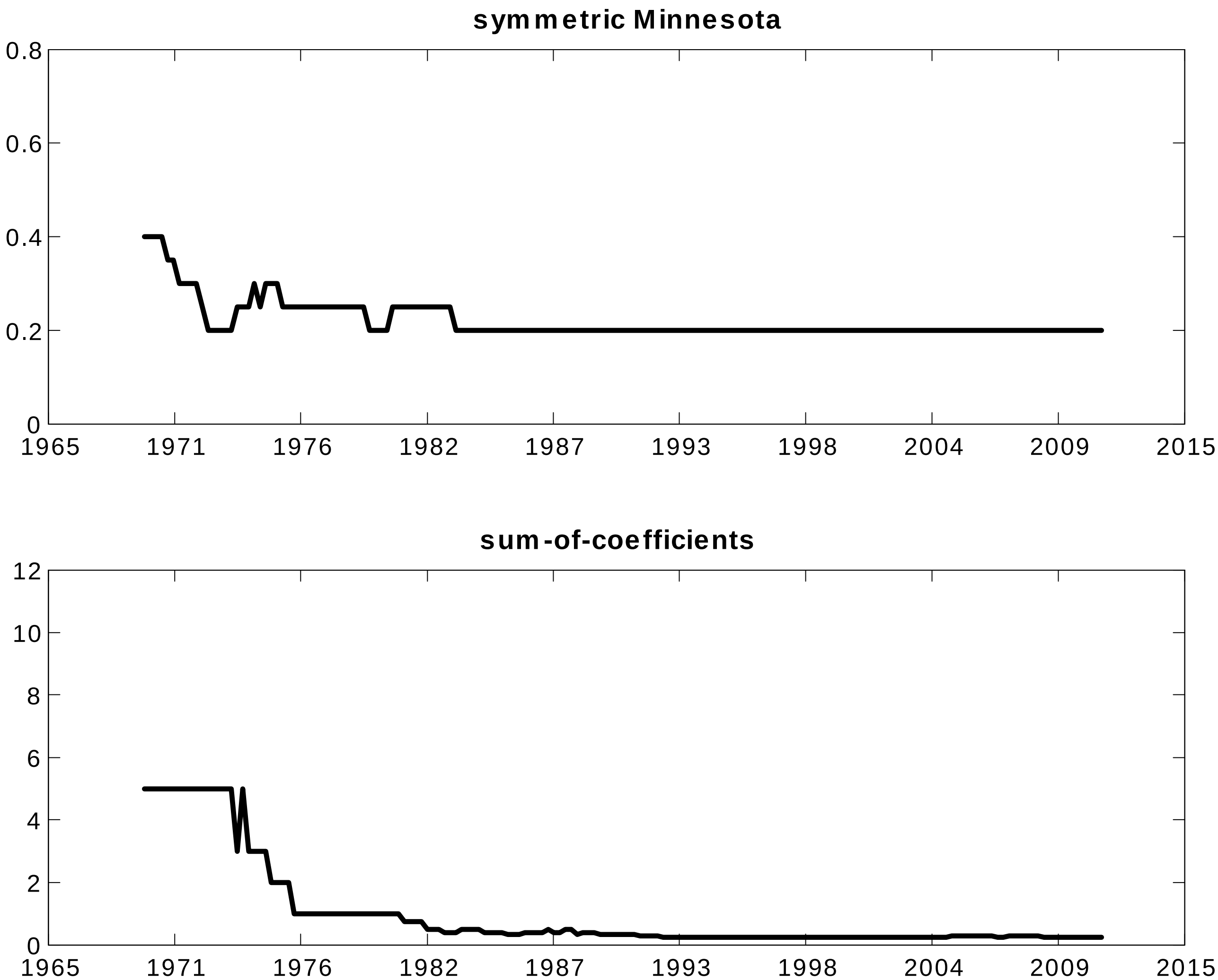

Figure 1: Marginal Likelihood Optimized Priors. Notes for the figure: The plots represent the values of λ and μ that maximize the marginal likelihood at each date from 1970Q1 to 2011Q2. The optimization was performed over a discrete grid, $\lambda \in [0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 1]$ and $\mu \in [0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.75, 1, 2, 3, 5, 10, 15]$.

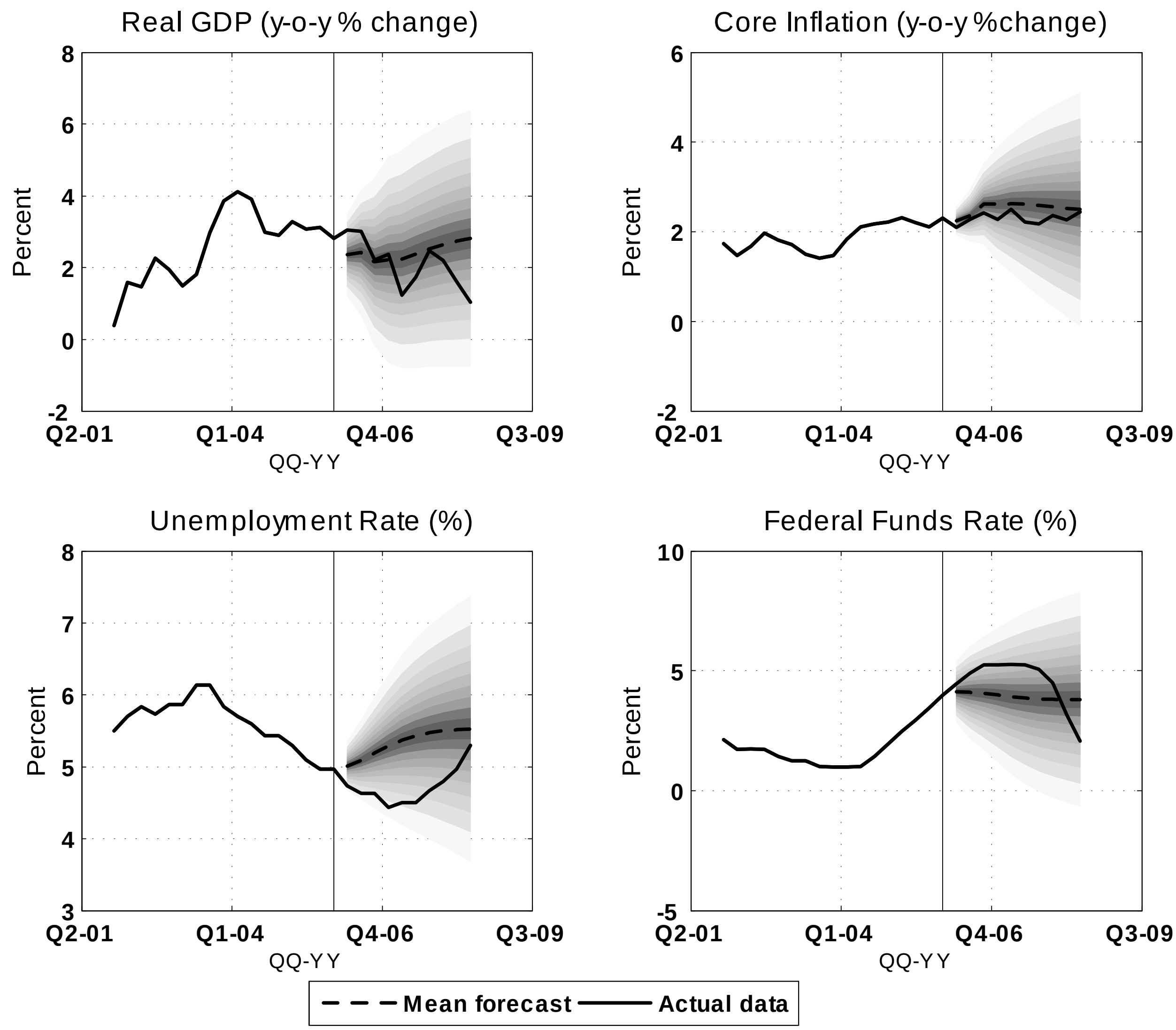

Figure 2: Fan Chart Density Forecasts: 2006Q1–2007Q4. Notes for the figure: The model was estimated till 2005Q4, and then upto eight step ahead density forecasts were computed. The shaded regions represent 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% probability intervals. Shaded intervals become lighter with interval size.

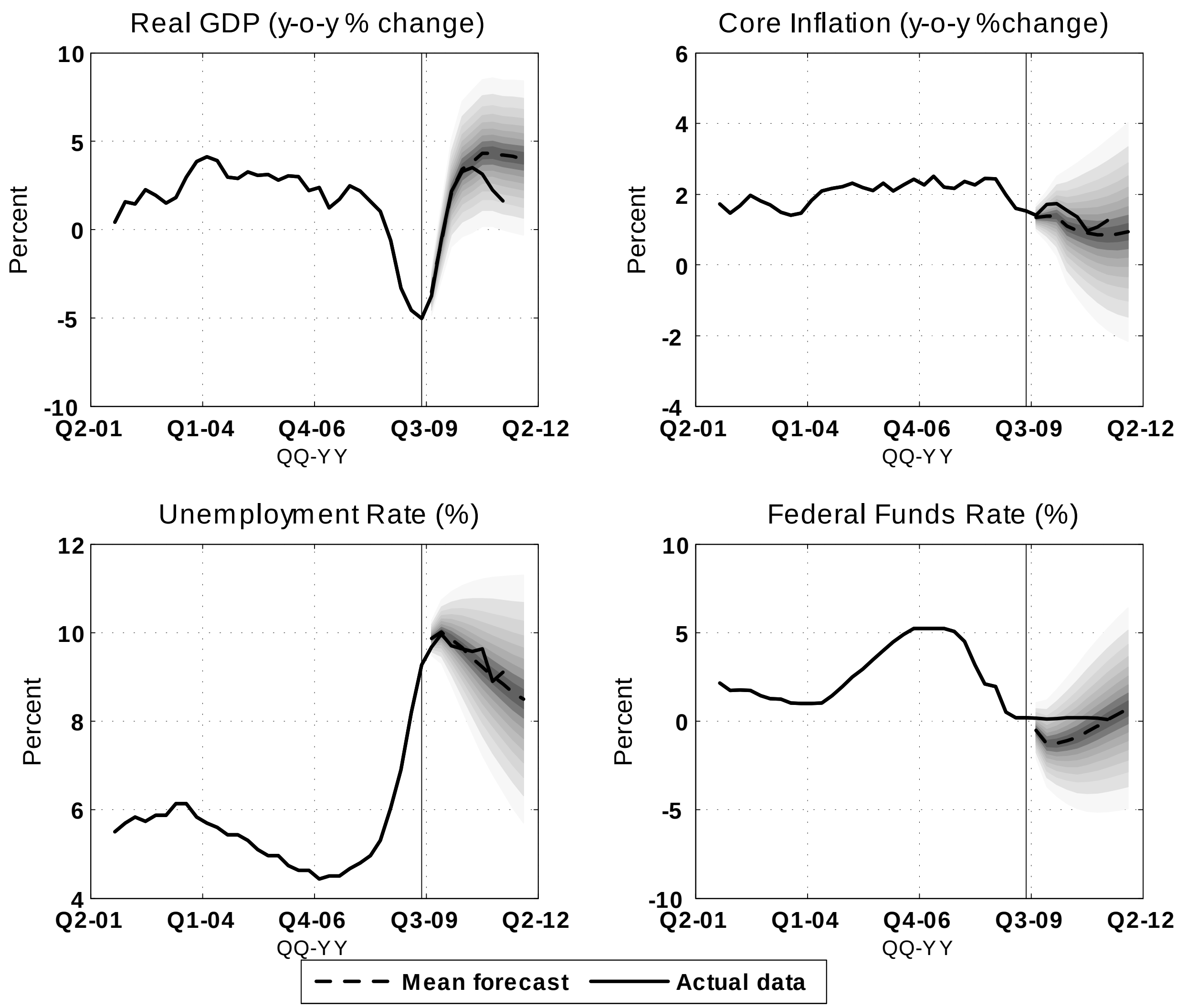

Figure 3: Fan Chart Density Forecasts: 2009Q3–2011Q4. Notes for the figure: Fancharts corresponds to the $h = 1, 2, \ldots, 8$ step ahead predictive density forecast using the model estimated over the 1959Q1–2009Q2 period. The shaded regions represent 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% probability intervals. Shaded intervals become lighter with interval size.

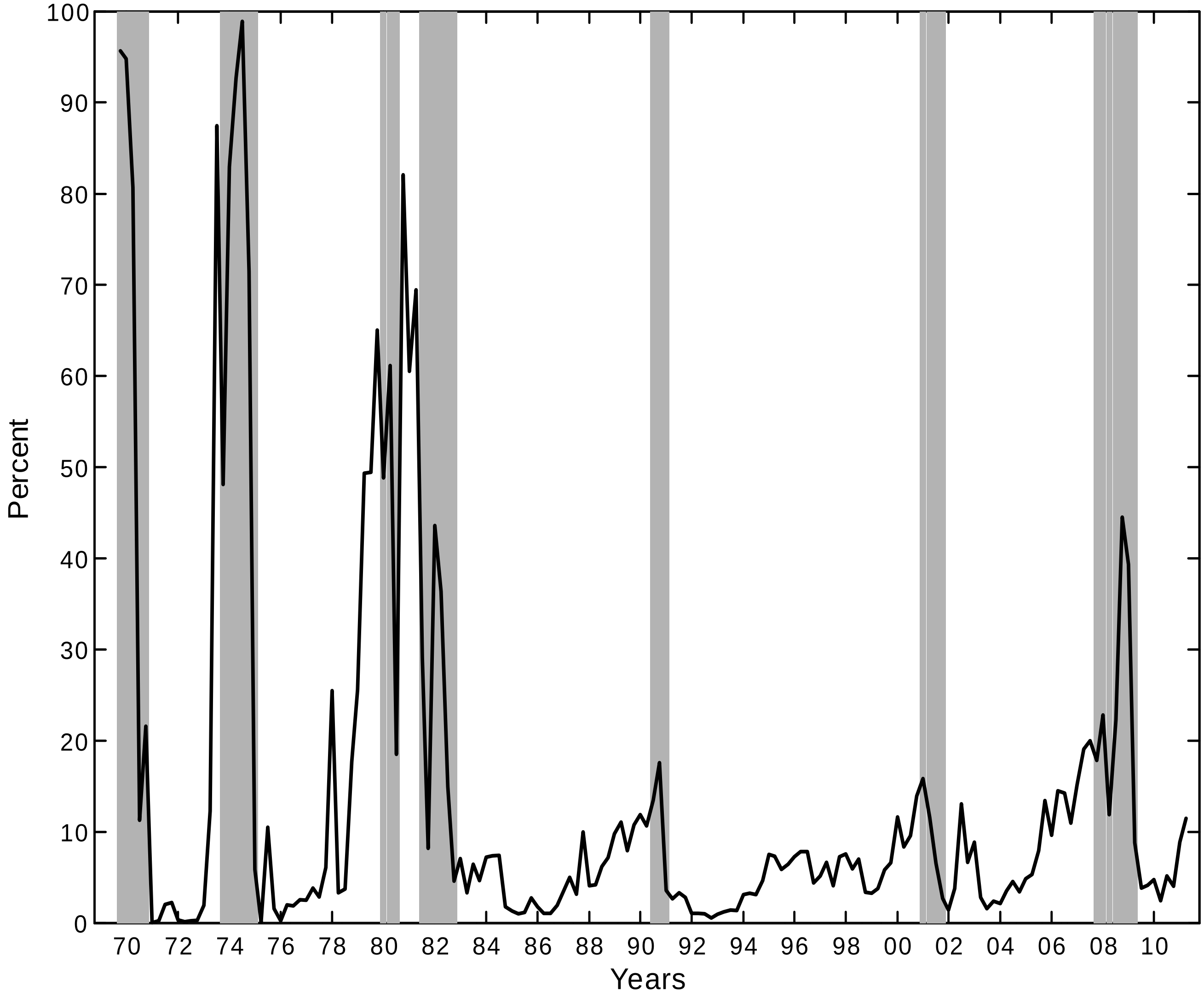

Figure 4: Recession Probabilities. Notes for the figure: Shaded regions represent NBER-defined recession periods.

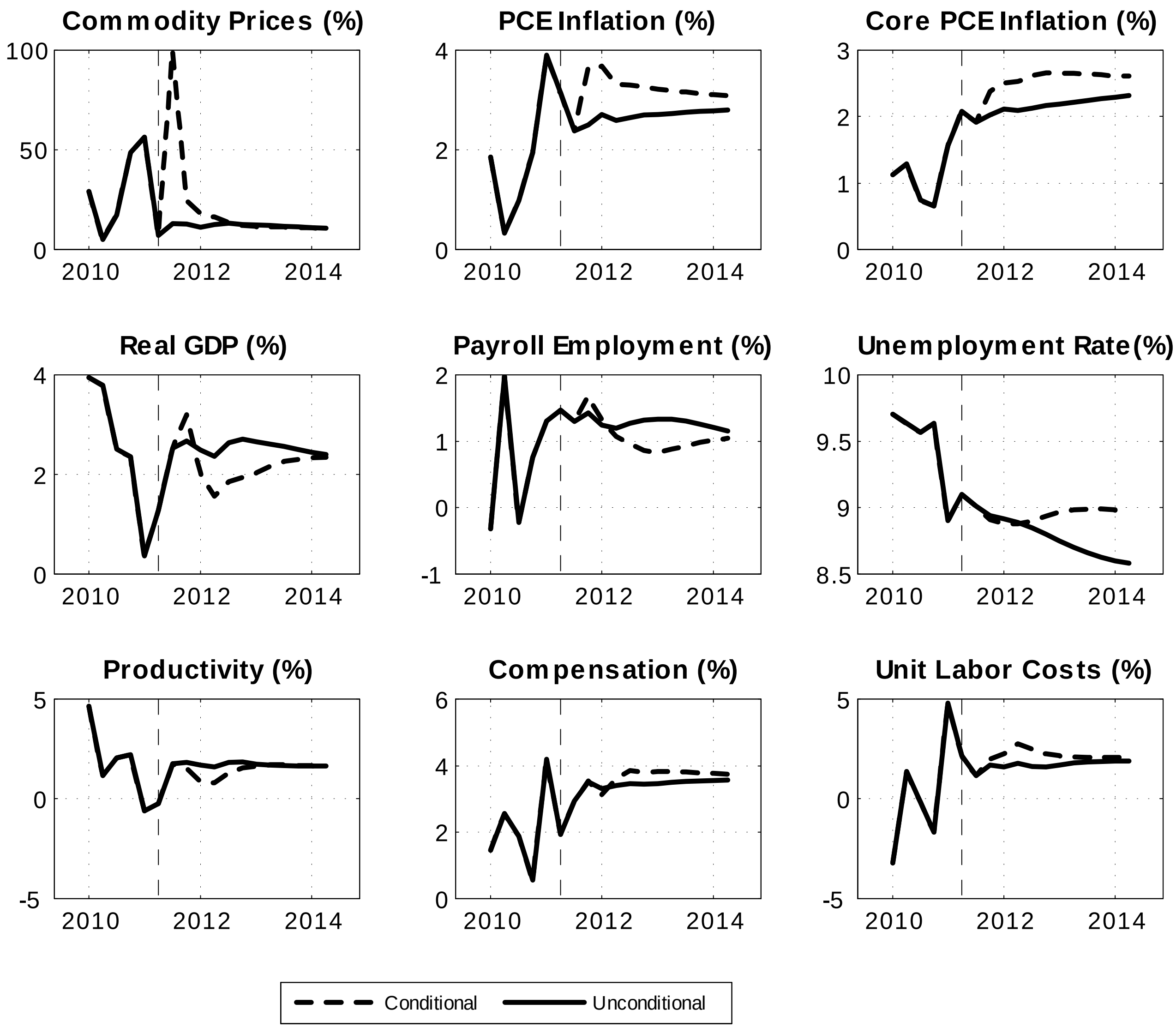

Figure 5: Unconditional and Conditional Forecasts: 2011Q3-2014Q2. Notes for the figure: Solid lines represents the unconditional forecast; dashed lines represent the forecast conditional on a value of PCOMM 25 percent higher than the unconditional forecast in the initial forecast period. All quantities reportes quarterly annualized growth rates with the exception of the unemployment rate which is in levels.

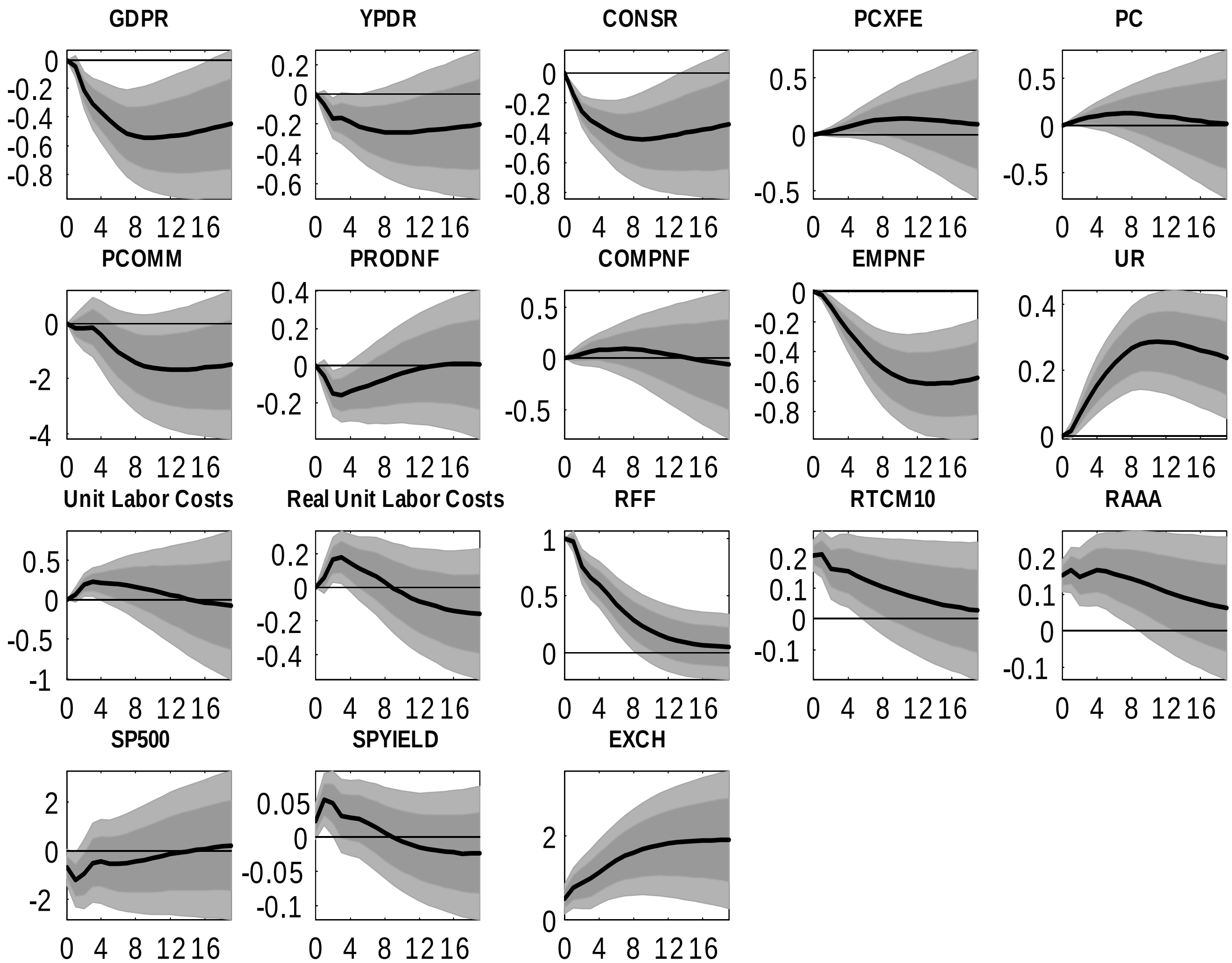

Figure 6: Impulse Responses: Monetary Policy Shock. Notes for the figure: The figure displays the impulse response functions corresponding to a contractionary monetary policy shock of 100 basis points. Solid black lines represent the median response; the shaded regions indicate the posterior coverage intervals for the 70 and 90 percent probability bands. The model is estimated from 1959Q1 to 2011Q2. Note that unit labor costs and real unit labor costs are inferred from the primitive model variables of compensation and productivity; real unit labor costs are deflated with the core PCE index.

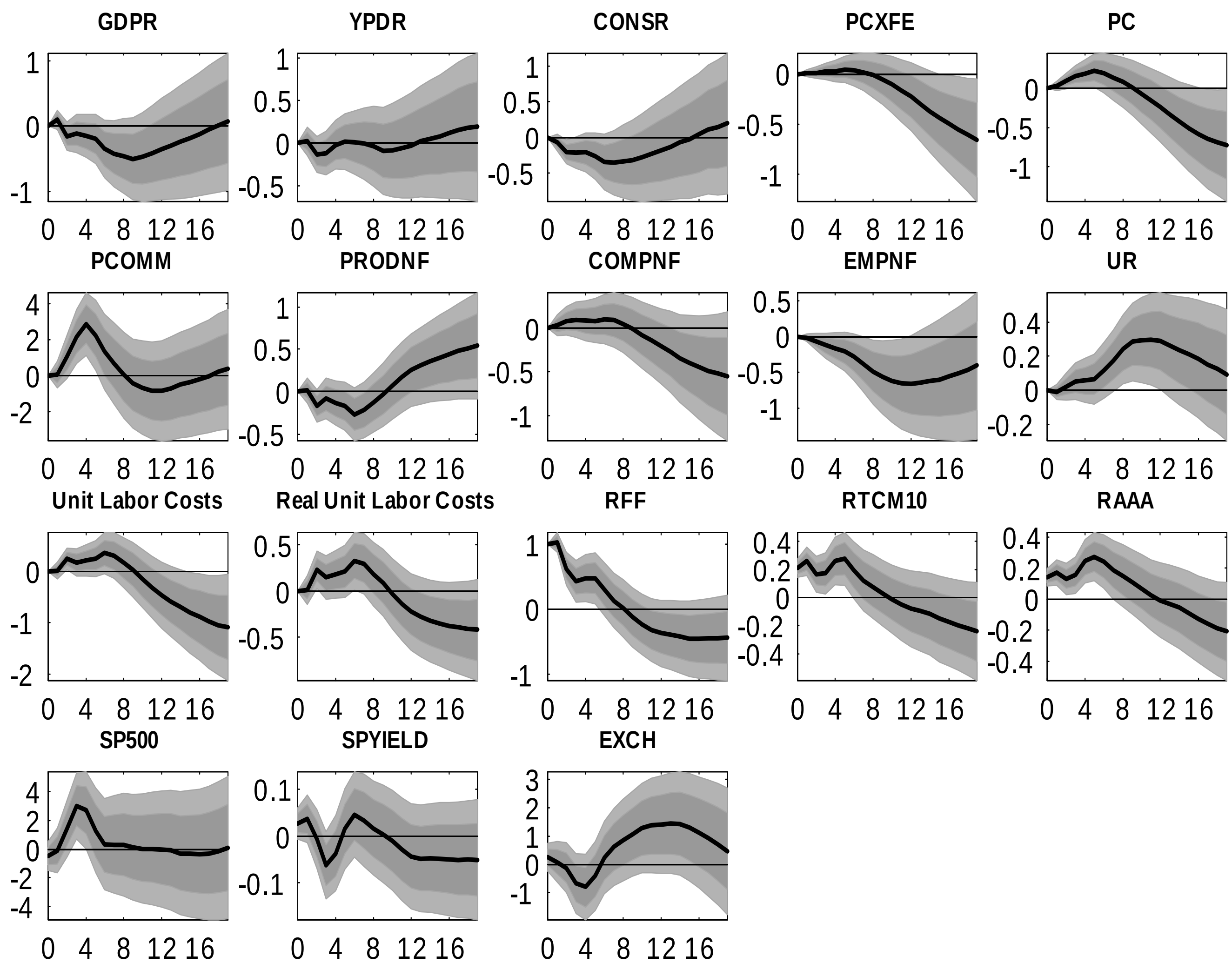

Figure 7: Impulse Responses: Monetary Policy Shock with Flat Priors. Notes for the figure: The figure displays the impulse response functions of the VAR estimated with flat priors generated by contractionary monetary policy shock of 100 basis points. Solid black lines represent the median response; the shaded regions indicate the probability intervals for the 70 and 90 percent probability bands. The model is estimated from 1959Q1 to 2011Q2. Note that unit labor costs and real unit labor costs are inferred from the primitive model variables of compensation and productivity; real unit labor costs are deflated with the core PCE index.