Federal Reserve Bank of New York
Staff Reports

Real-Time Inflation Forecasting in a Changing World

Jan J. J. Groen
Richard Paap
Francesco Ravazzolo

Staff Report no. 388
August 2009
Revised May 2012

Real-Time Inflation Forecasting in a Changing World
Jan J. J. Groen, Richard Paap, and Francesco Ravazzolo
Federal Reserve Bank of New York Staff Reports, no. 388
August 2009; revised May 2012
JEL classification: C11, C22, C53, E31

Abstract

This paper revisits the accuracy of inflation forecasting using activity and expectations variables. We apply Bayesian-model averaging across different regression specifications selected from a set of potential predictors that includes lagged values of inflation, a host of real activity data, term structure data, nominal data, and surveys. In this model average, we can entertain different channels of structural instability by incorporating stochastic breaks in the regression parameters of each individual specification within this average, allowing for breaks in the error variance of the overall model average, or both. Thus, our framework simultaneously addresses structural change and model uncertainty that would unavoidably affect any inflation forecast model. The different versions of our framework are used to model U.S. PCE deflator and GDP deflator inflation rates for the 1960-2011 period. A real-time inflation forecast evaluation shows that averaging over many predictors in a model that at least allows for structural breaks in the error variance results in very accurate point and density forecasts, especially for the post-1984 period. Our framework is especially useful when forecasting, in real-time, the likelihood of lower-than-usual inflation rates over the medium term.

Key words: inflation forecasting, Phillips correlations, real-time data, structural breaks, model uncertainty, Bayesian model averaging

---

Groen: Federal Reserve Bank of New York (e-mail: jan.groen@ny.frb.org). Paap: Erasmus University Rotterdam (e-mail: paap@ese.eur.nl). Ravazzolo: Norges Bank (francesco.ravazzolo@norges-bank.no). This paper has benefited from helpful comments by Todd Clark and seminar participants at Erasmus University Rotterdam, the Federal Reserve Bank of Kansas City, the European Meeting of the Econometric Society in Milan, the Riksbank workshop “Modeling and Forecasting Economic and Financial Time Series with State Space Models” in Stockholm, the Centre for International Macroeconomics and Finance workshop “Forecasting under Model Instability” at Cambridge University, the Australian Meeting of the Econometric Society in Canberra, and the “Forecasting and Methods in Macro and Finance Workshop” at the NBER Summer Institute in Boston.

1 Introduction

Control of inflation is at the core of monetary policymaking and, consequently, central bankers have a great interest in reliable inflation forecasts to help them achieving this aim. For other agents in the economy accurate inflation forecasts are likewise of importance, either to be able to assess how policymakers will act in the future or to help them in forming their inflation expectations when negotiating about wages, price contracts and so on. And in the academic literature inflation predictability is assessed to get a gauge on the characteristics of inflation dynamics in general.

The time series properties of inflation measures, however, have changed substantially over time, as shown by Cogley and Sargent (2002, 2005) for the United States, by Benati (2004) for the United Kingdom and by Levin and Piger (2004) for twelve main OECD economies, all of which document significant time-variation in the mean and persistence of inflation. Related to that, Cogley and Sargent (2002) and Haldane and Quah (1999) document substantial shifts in the traditional U.S. and U.K. Phillips curve correlations between inflation and unemployment over the post-WWII period. There is also evidence that macroeconomic time series have experienced variance breaks over the post-WWII period that were unrelated to shifts in the mean. Cogley and Sargent (2005) and Cogley *et al.* (2010), for example, use a time-varying VAR model for U.S. inflation, unemployment and the interest rate with a stochastic volatility specification for the corresponding disturbance covariance matrix. Sensier and van Dijk (2004) find that for 80% of 214 U.S. macroeconomic time series over the 1959-1999 period most of the observed reduction in volatility is due to breaks in conditional volatility rather than breaks in the conditional mean. In addition, Sims and Zha (2006) claim using structural Bayesian VAR models with Markov switching time-variation that the observed time-variation in U.S. macroeconomic dynamics are entirely due to breaks in the variance of shocks and not in regression parameters.

This suggest that adding structural change to time series models may help to improve forecasting inflation. Stock and Watson (2007, 2008) show that U.S. inflation is well described by a univariate unobserved component model with a stochastic volatility specification for the disturbances. The out-of-sample performance of this particular model appears to be hard to beat by a range of alternative models. More generally, Koop and Potter (2007), through change-point models, and Pesaran *et al.* (2006), through a hierarchical hidden Markov chain model, show that forecast models that incorporate structural breaks exhibit good out-of-sample forecasting performance for a range of macroeconomic series.

Another issue for inflation forecasting is how to choose the predictor variables for future inflation. For example, a reduced form version of the Phillips curve relationship implies for forecasting a model where inflation depends on its lags, a measure of real activity (which approximates the degree of 'economic slack' or excess demand in the economy) and, possibly, a measure of inflation expectations. Typically unemployment is used as the 'slack measure' in such an inflation forecasting model. There is, however, a lot of uncertainty about the 'appropriate' measure of real activity that can be used in such a forecasting model. In fact, Stock and Watson (1999) show that unemployment-based Phillips curve models are frequently outperformed by models using alternative real activity measures. They consider two approaches. One is based on a forecast combination of the different, possible choices of Phillips curve forecasting models. Next, they also consider a single Phillips curve-based model that uses a principal component extracted from all possible 'economic slack' variables as the real activity measure.

Stock and Watson (1999) show that the out-of-sample performance of these approaches are favorable compared to autoregressive (AR) specifications, in particular in case of the factor-based approach. Wright (2009) applies a form of Bayesian model averaging across 93 potential specifications, each using one alternative activity measure, to forecast different quarterly U.S. inflation measures out-of-sample, and he is able to beat AR inflation forecasts out-of-sample. Atkeson and Ohanian (2001), on the other hand, apply the Stock and Watson (1999) exercise on a longer U.S. sample. They use Phillips curve specifications based on unemployment as well as the Chicago Fed Economic Activity index, which summarizes information across a range of activity variables, across a large gird of lag orders of the predictor variables. In the Atkeson and Ohanian (2001) case none of the inflation forecasting models are able to outperform random walk forecasts for inflation.

Like Stock and Watson (1999), we use in this paper a forecasting model that is essentially is an autoregressive model for inflation with added exogenous regressors (an ARX model). But unlike the papers surveyed above, we use a framework that allows for both instability in the relationship between inflation and predictor variables as well as uncertainty regarding the inclusion of potential predictors in the inflation forecasting regression. Bayesian model averaging [BMA] is used to deal with the latter model uncertainty, where we average over the range of regression models that incorporate all the possible combinations of predictor variables for inflation. To deal with instability, we allow for structural breaks of random magnitude in either the regression parameters, the error variance of the resulting model average, or both. Hence, our forecasting procedure simultaneously incorporates the two major sources of uncertainty, which the literature has shown to be relevant for forecasting and modeling inflation.

Our framework, described above, as well as other more regularly used approaches are used to model different definitions of U.S. inflation on a quarterly sample starting in 1960 and ending in 2011. A range of predictor variables are considered in the modeling exercise, from real variables to nominal and financial variables as well as lags of inflation. The full sample results show that our methodology has time-varying properties similar to existing studies and it identifies inflation predictor variable combinations of different composition and size for these U.S. inflation rates.

The different BMA-based specifications are then used to forecast the different inflation measures, both in the current quarter and for a one-year ahead forecasting horizon. Where necessary, we use in the out-of-sample forecasting experiments real-time data for inflation and the predictor variables, i.e. the original vintage of data that was available at the time of the forecast. The literature traditionally focuses on point forecast evaluation, see, e.g., Stock and Watson (2007, 2008, 2010) and Faust and Wright (2011), but given the interest amongst practitioners in predicting the likelihood of exceptional inflation movements, such as deflation or accelerating inflation, an assessment of the relative density forecast performance amongst different inflation forecast models seems in order. Therefore, our real-time inflation forecasting analysis will focus both on evaluating point forecasts as well as density forecasts. Relative to the latter, Jore *et al.* (2010) and Clark (2011) are closest to our exercise in that they assess density inflation forecasts from, respectively, forecast combinations and four-variable VAR models based on real-time data that allow for a limited degree of instability through rolling windows and the use of stochastic volatility specifications. However, relative to these studies we push to envelope further, as *(i)* our BMA-based specifications more explicitly takes into account model uncertainty and differing degrees of structural instability, and *(ii)* we consider a larger pool of alternative inflation forecast models.

The remainder of this paper is organized as follows. In Section 2 we introduce the framework for our BMA-based inflation models, which includes a detailed discussion of the underlying Bayesian estimation methodology. In Section 3 we operationalize our models by determining the prior values and assess whether our BMA-based models are able to replicate the aforementioned time-varying characteristics of U.S. inflation dynamics in the post-WWII era. Next, we evaluate their real-time forecasting performance in Section 4 by comparing them to other frequently used inflation forecasting models. We find that allowing for model uncertainty in combination with some form of structural instability results in superior point and density forecasts *vis-à-vis* other inflation forecasting approaches. Finally, in Section 5 we conclude.

2 A Bayesian Model Averaging Framework for Inflation

In this paper, we will use the following version of the Stock and Watson (1999) generalized Phillips curve specification as the starting point for our inflation forecasting model:

$$\begin{aligned} y_{t+h} =& \beta_0 + \sum_{j=1}^{k_1} \beta_j^a a_{jt} + \sum_{j=0}^{k_2} \beta_j^y y_{t-j} + \sigma\varepsilon_t \\ =& \beta_0 + \sum_{j=1}^{k} \beta x_{jt} + \sigma\varepsilon_t; \quad t = 1, \ldots, T-h, \end{aligned} \tag{1}$$

where T is the total number of time series observations in the sample. Variable y_t in (1) is the inflation measure, defined as $y_t = 100\Delta\ln(P_t) = 100(\ln(P_t) - \ln(P_{t-1}))$ where P_t is a particular price index and $h > 0$ is the forecast horizon with $y_{t+h} = 100\Delta\ln(P_{t+h}) = 100(\ln(P_{t+h}) - \ln(P_{t+h-1}))$.[1] The a_{jt}'s are the k_1 real activity, costs indicator, and inflation expectation proxy variables and the model also contains k_2 lagged values of y_t. The disturbance term ε_t in (1) is assumed to have $\varepsilon_t \sim \text{NID}(0,1)$ and $\sigma > 0$. For the ease of notation, we define $(x_{1,t} \cdots x_{k,t})' = (a_{1,t} \cdots a_{k_1,t}\ y_t \cdots y_{t-k_2})'$ and thus $k = k_1 + k_2$.

Clearly, the number of potential predictor variables k in (1) can be large. It can encompass traditional 'cost-push' drivers of inflation such as wage and production costs (the latter amongst others related to energy and imports), as well as proxies for excess demand in the economy that can push up inflation. A large number for k renders the model inestimable and we therefore have to make a choice about which combination of predictors to include under what circumstances. Hence, we have to adapt (1) such that it incorporates this model uncertainty.

Also, it is not realistic to assume that the relationship between inflation and its potential predictors in (1) has remained stable over time. Cogley and Sbordone (2008) and Groen and Mumtaz (2008) show that an empirical New Keynesian Phillips curve model that allows for shifts in the equilibrium inflation rates yields a time-varying reduced form inflation-activity relationship, given unchanged 'deep parameters', for a number of G7 economies. Stock and Watson (2008, 2010), on the other hand, suggest that there are non-linearities in the relationship between inflation and the amount of 'slack' in the economy, which could result in time-variation in an inflation forecasting relationship such as in (1).

2.1 Our Framework

The previous discussion implies that for forecasting inflation we need to adapt the basic regression model (1) such that it incorporates *model uncertainty* and *structural breaks* as

[1] Thus, we model the quarterly inflation rate in quarter $t+h$.

both inflation itself and the relationship between inflation and predictor variables likely have changed over time.

In our context, model uncertainty implies the uncertainty about which combination of predictor variables most accurately summarizes the impact of real activity, real costs and expectations on inflation dynamics. To allow for model uncertainty we introduce in (1) k variables $\delta_j \in \{0,1\}$ which describe the inclusion of variable x_{jt} in the regression model for $j = 1, \ldots, k$. Structural breaks in the regression parameters and the variance are incorporated by introducing time-varying regression parameters β_{jt} and σ_t in (1). This results in

$$y_{t+h} = \beta_{0t} + \sum_{j=1}^{k} \delta_j \beta_{jt} x_{jt} + \sigma_t \varepsilon_t; \quad t = 1, \ldots, T-h, \tag{2}$$

where $\varepsilon_t \sim \text{NID}(0,1)$, and where x_{jt} may include lagged vales of inflation known at time t. The vector $D = (\delta_1, \ldots, \delta_k)'$ describes which regressors are included in the regression model. It can take 2^k different values, resulting in 2^k different regression models. Model selection is therefore defined in terms of variable selection, see George and McCulloch (1993) and Kuo and Mallick (1998). Note that the intercept parameter β_{0t} is always included in the model.

The time varying parameters β_{jt} and σ_t are described by

$$\begin{aligned} \beta_{jt} &= \beta_{j,t-1} + \kappa_{jt}\eta_{jt}; \qquad j = 0, \ldots, k, \\ \ln \sigma_t^2 &= \ln \sigma_{t-1}^2 + \kappa_{k+1,t}\eta_{k+1,t}, \end{aligned} \tag{3}$$

where $\eta_t = (\eta_{0t}, \ldots, \eta_{k+1,t})' \sim \text{NID}(0, Q)$ with $Q = \text{diag}(q_1^2, \ldots, q_{k+1}^2)$. The size of the structural breaks is described by independent random shock η_{jt} with mean zero and variance q_j^2 for $j = 0, \ldots, k+1$, see, for example, Koop and Potter (2007) and Giordani *et al.* (2007) for a similar approach. To allow for occasional structural breaks we assume that the $k+2$ variables κ_{jt} are binary random variables which equal 1 in case of a structural break in the jth parameter at time t and 0 otherwise for $j = 0, \ldots, k+1$ and $t = 1, \ldots, T$. We assume that the vector $\kappa_t = (\kappa_{0t}, \ldots, \kappa_{kt}, \kappa_{k+1,t})'$ is a sequence of uncorrelated 0/1 processes with

$$\Pr[\kappa_{jt} = 1] = \pi_j; \quad j = 0, \ldots, k+1. \tag{4}$$

This specification implies that a regression parameter β_{jt} in (3) remains the same as its previous value $\beta_{j,t-1}$ unless $\kappa_{jt} = 1$ in which case it changes with η_{jt}, see, for example, Koop and Potter (2007) and Giordani *et al.* (2007) for a similar approach. Stochastic structural breaks in the variance parameter $\ln \sigma_t^2$ comply to a similar structure as the β_{jt} parameters. The flexibility of the specification in (3) stems from the fact that the parameters β_{jt} and σ_t^2 are allowed to change every time period, but they are not imposed to change at every point in time. Another attractive property of (3) is that the changes

Table 1: BMA inflation model specifications for (2).

	β_{jt}	$\ln\sigma_t^2$	κ_{jt}	$\kappa_{k+1,t}$
BMA-SBB-SBV	as in (3)	as in (3)	as in (4)	as in (4)
BMA-SBV	$\beta_{jt} = \beta_j \forall t$	as in (3)	–	as in (4)
BMA-SBB	as in (3)	$\ln\sigma_t^2 = \ln\sigma^2 \forall t$	as in (4)	-
BMA	$\beta_{jt} = \beta_j \forall t$	$\ln\sigma_t^2 = \ln\sigma^2 \forall t$	–	–
BMA-RWB-RWV	as in (3)	as in (3)	$\kappa_{jt} = 1 \forall j, t$	$\kappa_{k+1,t} = 1 \forall t$

in the individual parameters are not restricted to coincide but are allowed to occur at different points in time.

The model presented in (2)–(4) is the most general specification we consider in this paper. We will call this the BMA-SBB-SBV specification indicating that we consider Bayesian model averaging [BMA], structural breaks in the regression parameters [SBB] and structural breaks in the volatility [SBV]. We will also consider specifications where $\kappa_{jt} = 1$ for all j and t in (4) in which case the parameters follow a random walk and structural breaks occurs at every point in time. We will denote these specifications by RWB and RWV for the regression parameters and volatility, respectively. The random walk specification is useful to describe many small breaks where the discrete specification can be used to describe occasional large breaks, see also Giordani and Villani (2010) for a discussion. Table 1 provides an overview of the alternative specifications we entertain in our framework, where different aspects of parameter and variance instability in the model are switched off or altered.

2.2 Prior Specification

For parameter inference in (2)–(4), we opt for a Bayesian approach. Such an approach allows us to incorporate parameter uncertainty when forecasting inflation in a natural way. Also, Bayesian inference on $D = (\delta_1, \ldots, \delta_k)$ leads to posterior probabilities for the 2^k possible model specifications. We will use these posterior probabilities for Bayesian model averaging to incorporate model uncertainty into a single inflation forecast. Finally, the approach provides us with the posterior distribution of the unobserved κ_t processes for $t = 1, \ldots, T-h$, which can be used to incorporate uncertainty regarding the timing of structural breaks. By definition, κ_t in (2) does not depend on D which implies that the value of κ_t can be different across different values of D. Hence, structural breaks can occur in different parameters at different time periods across different models, and we average over the latter to obtain our final inflation forecasting relationship.

The parameters in the model (2)–(4) are the inclusion vector $D = (\delta_1, \ldots, \delta_k)'$, the structural break probabilities $\pi = (\pi_0, \ldots, \pi_{k+1})'$ and the vector of variances of the size of the breaks $q^2 = (q_0^2, \dots, q_{k+1}^2)'$. We collect the model parameters in a $(3k+4)$-dimensional vector $\theta = (D', \pi', q^{2'})'$.

For our Bayesian approach we need to specify the prior distributions for the model parameters. For the variable inclusion parameters we take a Bernoulli distribution with

$$\Pr[\delta_j = 1] = \lambda_j \qquad \text{for } j = 1, \dots, k. \tag{5}$$

Hence, the parameter λ_j reflect our prior belief about the inclusion of the jth explanatory variable, see George and McCulloch (1993) and Kuo and Mallick (1998). For the structural break probability parameters we take Beta distributions

$$\pi_j \sim \text{Beta}(a_j, b_j) \qquad \text{for } j = 0, \dots, k+1. \tag{6}$$

The parameters a_j and b_j can be set according to our prior belief about the occurrence of structural breaks. The expected prior probability of a break is $a_j/(a_j + b_j)$. Finally, for the variance parameters which reflect our prior beliefs about the size of the structural breaks we take the inverted Gamma-2 prior with scale parameter ω_j and degrees of freedom parameter ν_j, that is,

$$q_j^2 \sim \text{IG-2}(\omega_j, \nu_j) \qquad \text{for } j = 0, \dots, k+1. \tag{7}$$

The expected prior variance of the break size is therefore $\omega_j/(\nu_j - 2)$ for $\nu_j > 2$.

The density of the joint prior $p(\theta)$ is given by the product of the densities of the prior specifications in (5)–(7).

2.3 Posterior Simulator

Posterior results are obtained using the Gibbs sampler of Geman and Geman (1984) combined with the technique of data augmentation of Tanner and Wong (1987). The latent variables $B = \{\beta_t\}_{t=1}^{T-h}$, with $\beta_t = (\beta_{0t}, \beta_{1t}, \dots, \beta_{kt})'$, $S = \{\sigma_t^2\}_{t=1}^{T-h}$, and $K = \{\kappa_{0t}, \dots, \kappa_{k+1,t}\}_{t=1}^{T-h}$ are simulated alongside the model parameters θ. To apply the Gibbs sampler we need the complete data likelihood function, that is, the joint density of the data and the latent variables

$$p(y, B, S, K | x, \theta) = \prod_{t=1}^{T-h} \Big(p(y_{t+h} | D, x_t, \beta_t, \sigma_t^2) p(\beta_t | \beta_{t-1}, \kappa_t, Q) \\ p(\ln \sigma_t^2 | \ln \sigma_{t-1}^2, \kappa_{k+1,t}, q_{k+1}^2) \prod_{j=0}^{k+1} \pi_j^{\kappa_{jt}} (1 - \pi_j)^{1-\kappa_{jt}} \Big), \tag{8}$$

where $y = (y_{1+h}, \ldots, y_T)$ and $x = (x'_1, \ldots, x'_{T-h})'$. The density functions are given by

$$\begin{aligned}
p(y_{t+h}|D, x_t, \beta_t, \sigma_t^2) &= \frac{1}{\sigma_t\sqrt{2\pi}} \exp\Big(- \frac{(y_{t+h} - \beta_{0t} - \sum_{j=1}^k \delta_j \beta_{jt} x_{jt})^2}{2\sigma_t^2}\Big) \\
p(\beta_t|\beta_{t-1}, \kappa_t, Q) &= \prod_{j=1}^k \left(\frac{1}{q_j\sqrt{2\pi}} \exp\Big(- \frac{(\beta_t - \beta_{t-1})^2}{2q_j^2}\Big) \right)^{\kappa_{jt}} (\boldsymbol{\delta}(\beta_{t-1}))^{1-\kappa_{jt}} \\
p(\ln\sigma_t^2|\ln\sigma_{t-1}^2, \kappa_{k+1,t}, q_{k+1}^2) &= \left(\frac{1}{q_{k+1}\sqrt{2\pi}} \exp\Big(- \frac{(\ln\sigma_t^2 - \ln\sigma_{t-1}^2)^2}{2q_{k+1}^2}\Big), \right)^{\kappa_{k+1,t}} (\boldsymbol{\delta}(\ln\sigma_{t-1}^2))^{1-\kappa_{k+1,t}}
\end{aligned} \tag{9}$$

where $\boldsymbol{\delta}(\cdot)$ is a dirac delta function.

If we combine (8) together with the prior density $p(\theta)$, we obtain the posterior density function

$$p(\theta, B, S, K|y, x) \propto p(\theta)p(y, D, S, K|x, \theta). \tag{10}$$

Our Gibbs sampler is a combination of the Kuo and Mallick (1998) algorithm for variable selection and the efficient sampling algorithm of Gerlach *et al.* (2000) to handle the structural breaks. If we define $K = (K_\beta \ K_\sigma)$ with $K_\beta = \{\kappa_{0t}, \ldots, \kappa_{kt}\}_{t=1}^{T-h}$ and $K_\sigma = \{\kappa_{k+1,t}\}_{t=1}^{T-h}$, then in each iteration of the sampler we sequentially cycle through the following steps:

1. Draw D conditional on B, S, K, π, q^2, y and x.

2. Draw K_β conditional on S, K_σ, θ, y and x.

3. Draw B conditional on S, K, θ, y and x.

4. Draw K_σ conditional on B, K_β, θ, y and x.

5. Draw S conditional on B, K, θ, y and x.

6. Draw π conditional on D, B, S, K, q^2, y and x.

7. Draw q^2 conditional on D, B, S, K, π, y and x.

Note that we sample K_β and B from their joint full conditional posterior distribution. The same holds for and K_σ and S. A more detailed description of this Gibbs sampling algorithm is provided in Appendix A.

2.4 Posterior Predictive Density

To forecast inflation, we use the posterior predictive density of model (2)–(4). This density accounts for uncertainty in the inclusion of the predictors x_{jt}, uncertainty about the presence and size of structural breaks, and parameter uncertainty. The predictive density of y_{T+1+h} made at time $T+1$ conditional on past data y and x is given by

$$p(y_{T+h+1}|y,x) = \int \cdots \int \sum_D \sum_K \sum_{\kappa_{T+1}} p(y_{T+h+1}|D, x_{T+1}, \beta_{T+1}, \sigma^2_{T+1})$$
$$p(\beta_{T+1}|\beta_T, \kappa_{T+1}, Q) p(\ln \sigma^2_{T+1}| \ln \sigma^2_T, \kappa_{T+1}, q^2_{k+1}) \prod_{j=0}^{k+1} \pi_j^{\kappa_{j,T+1}} (1-\pi_j)^{1-\kappa_{j,T+1}}$$
$$p(\theta, B, S, K|y,x) d\beta_{T+1} d\ln \sigma^2_{T+1} dBdSd\pi dq^2, \quad (11)$$

where the posterior density $p(\theta, B, S, K|y, x)$ is given in (10) and $p(y_{T+h+1}|D, x_{T+1}, \beta_{T+1}, \sigma^2_{T+1})$ and $p(\beta_{T+1}|\beta_T, \kappa_{T+1}, Q)$, $p(\ln \sigma^2_{T+1}| \ln \sigma^2_T, \kappa_{T+1}, q^2_{k+1})$ are given in (9). The predictive density (11) consists of a weighted average over all possible model specifications in (2) with weights equal to the posterior model probabilities. Uncertainty regarding the timing of structural breaks is reflected in (11) by the posterior distribution of the in-sample breaks K. Computation of such a predictive distribution is straightforward using the aforementioned Gibbs draws. We simulate in each Gibbs step y_{T+h+1} using (2)–(4) as data generating process, where we replace the parameters and the in-sample latent variables by the draw from the posterior distribution. As point forecast we use the posterior mean or median of the predictive distribution depending on the chosen loss function.

3 Data, Prior Choices, and U.S. Inflation Dynamics

In this section we outline how we operationalize the different variants of our BMA framework (2)–(4) on our dataset with an aim to describe the post-WWII behavior of two U.S. inflation measures. In Section 3.1 we discuss this data, whereas in Section 3.2 we report on the prior values we chose in order to be able to take our framework to the data in a satisfactorily manner. Some in-sample posterior characteristics of the specifications from Table 1 are discussed Section 3.3.

3.1 Data

We will consider two measures of U.S. inflation observed at a quarterly frequency ranging from 1960Q1 to 2011Q2; these are the quarterly log changes in the Personal Consumption Expenditures (PCE) deflator and the Gross Domestic Product (GDP) deflator. Although both these inflation series tend to move together, there are substantial differences in their respective compositions. It is therefore likely that the two inflation measures will exhibit different short-to-medium run behavior and we thus may expect the optimal predictors for the two measures to differ as well. Potentially there is wide array of predictors for inflation that can be useful in the variant from our model (2). Stock and Watson (1999), for example, consider up to 132 potential indicator variables using dynamic factors and forecast combination techniques. However, it is our aim to predict inflation in *real-time*. As both our inflation measures of interest as well as many potential predictor variables are revised over time, we need to restrict our pool of possible predictor variables to those for which we have the original vintages of real time data. This limits the number of predictors to about fifteen series.

The inflation measures and the set of potential predictors come either directly or are constructed from five data sources. These are the Real-Time Data Set for Macroeconomists (RTDSM) at the Federal Reserve Bank of Philadelphia, the Haver Analytics database, the ALFRED® real-time database at the Federal Reserve Bank of St. Louis, the CRSP database from Wharton Research Data Services, and the Reuters/University of Michigan Survey of Consumers. We refer the reader to Appendix B for more details on the data sources and data construction.

Our range of predictor variables can be divided into three groups. For real activity and cost indicators we have real GDP in volume terms (ROUTP), real durable PCE in volume terms (RCONS), real residential investment in volume terms (RINVR), the import deflator (PIMP), the unemployment ratio (UNEMP), non-farm payrolls data on employment (NFPR), housing starts (HSTS), the real spot price of oil (OIL), the real food commodities price index (FOOD) and the real raw material commodities price index (RAW). These variables provide information about either the degree of excess demand in the economy or about the real costs that firms face.

We also include a number of nominal variables that are informative about the current and future state of the economy. One of these variables is the M2 monetary aggregate, which can reflect information on the current stance of monetary policy and liquidity in the economy as well as spending in households and firms (increased M2 growth might reflect increased spending by households and firms). In addition, we also use data on the term structure of interest rates, as this contains market-based forward-looking information about the business cycle, the stance of monetary policy and inflation expectations. It is well known that the cross-section of the term structure can be approximated very well by means of 2 to 3 factors, see, for example, Ang *et al.* (2006) and Diebold *et al.* (2006). Following the literature we therefore approximate the term structure through three factors constructed from data on Treasury Bill rates and zero-coupon bond yields: the level factor (YL), the slope factor (TS) and curvature factor (CS).

Finally, we proxy inflation expectations through the one-year ahead inflation expectations that come from the Reuters/Michigan Survey of Consumers (MS). Surveys can give potentially a very good steer about agents' expectations and indeed Ang *et al.* (2007)

claim that in an out-of-sample context inflation expectation surveys are the most accurate predictors for future U.S. inflation.

For most of these variables, we use the percentage change of the original series[2] to remove possible stochastic and deterministic trends from the series. Exceptions are housing starts, for which we use the logarithm of the respective levels, as well as the unemployment ratio, the three term structure factors and the inflation expectations survey for which we use the 'raw' levels of the series. An alternative way of removing trends from our predictor variables would be to express them as 'gap measures', i.e., log deviations from some proxy of a variable's trend (like a quadratic deterministic trend, exponential smoothing, statistical filters and so on). Orphanides and van Norden (2005), however, find that, in real time, activity growth rates are more useful predictors for U.S. inflation than activity gap measures. We standardize each predictor variable by dividing it by its standard deviation in each vintage. The standardization is very useful to set priors because all the variables can be treated *a priori* similarly in terms of breaks and break sizes in equations (2)–(3).[3]

3.2 Prior Choices and Posterior Convergence

For the empirical analysis in this paper, we use the 15 predictors as discussed in the previous subsection and we also include four inflation lags $(y_t, \ldots, y_{t-3})$ as potential explanatory variables in the model. In order to be able to use the BMA-based framework from Section 2, we need to take a stand on a proper calibration of the prior distributions outlined in Section 2.2 and Table 2 provides an overview of these calibrations.

The parameter λ_j in (5) is set to 50% which means that every variable (including the four inflation lags) has a 50% prior probability of being selected. The expected prior model size is therefore 9.5.

We set $a_0 = 0.50$ and $b_0 = 5$ in (6), and $\omega_0 = 1$ and $\nu_0 = 10$ in (7) for the intercept parameter, which suggests that we can expect, on average, an intercept break to occur every 11 quarters of size 1/8. For the regression parameters we impose the view that they are subject to less and smaller breaks than the intercept, and we therefore choose $a_j = 0.5$ and $b_j = 30$ (less breaks) and $\omega_j = 0.01$ and $\nu_j = 20$ (smaller magnitude) for $j = 1, ..., k$. This difference in prior settings reflects our interpretation of U.S. post-WWII history where large shifts in the level of inflation coincided with changes in the monetary policy regimes, such as, e.g., the shift in emphasis by the Federal Reserve System from monetary accommodation in the 1970s to a strong preference for disinflation by the late 1970s early 1980s. It is our view that the intercept in (2) should reflect these particular type of shifts

[2] That is, 100 times the quarterly change of the logarithm of the original series.

[3] When we use non-standardized predictors in conjunction with priors that are proportional to the predictor variable variances, we obtain similar results to the ones reported further on in this paper.

Table 2: Prior parameters for the priors (5)–(7) of the BMA inflation models

Parameter	Values	Parameter	Values
λ_j	0.50 for $j = 1, \ldots, k$	ν_0	10
a_j	0.50 for $j = 0, \ldots, k$	ν_j	20 for $j = 1, \ldots, k$
b_0	5	ω_0	1
b_j	30 for $j = 1, \ldots, k$	ω_j	0.01 for $j = 1, \ldots, k$
a_{k+1}	0.80	ν_{k+1}	10
b_{k+1}	2	ω_{k+1}	0.50

in average inflation, where time-variation in the remaining regression parameters then can pick-up changing correlations between inflation and its potential predictors across different phases of the business cycle. To account for the larger value of ω_0 than ω_j we choose a smaller value of ν_0 than ν_j for $j = 1, \ldots, k$.

The prior beliefs on the parameters concerning the variance equation reflect a larger number of breaks than for the intercept, but with a smaller magnitude ($a_{k+1} = 0.8$, $b_{k+1} = 2$, $\omega_{k+1} = 0.5$ and $\nu_{k+1} = 10$). These parameters correspond to beliefs in other time-varying inflation model specifications such as in, for example, Cogley and Sargent (2005) and Primiceri (2005). Finally, the priors for the initial values of β_{jt} and $\ln \sigma_t$ are normal, see Appendix A for details.

The posterior simulator described in Section 2.3 in combination with the prior values from Table 2 can easily be adapted to estimate variants of our BMA-based framework other than the BMA-SBB-SBV specification, see Table 1, by simply switching off one or all of the channels of structural instability. This is less trivial when we consider the BMA-RWB-RWV version of our framework where κ_{jt} and $\kappa_{k+1,t}$ are equal to 1 for all j and t. In this case π_j does not play any role and the information in q_j^2 using the prior values from Table 2 is possible very large since it is always valid independently of π_j. Therefore, when estimating the BMA-RWB-RWV specification we decrease the prior value of ν_j, $j = 0, ..., k+1$ to 3, which is comparable to the prior choices made in other studies that utilize a standard time-varying parameter specification, such as Cogley and Sargent (2005). The priors for the initial values for the time-varying regression parameters in the BMA-RWB-RWV specification are similar to those of the other BMA-based variants.

The posterior results for our BMA-based framework are not very sensitive to the prior settings for the λ parameter that governs the prior distribution on the variable inclusion parameters δ_j in (2). A different prior setting for the probability and size of the breaks, however, can impact the different posterior results substantially as is also shown by a simple prior sensitivity analysis in Appendix C on a simulated data set. The analysis in Appendix C shows that the posterior inclusion probabilities are hardly affected by the prior setting for the probability and size of breaks unless the prior on the expected break size on the regression parameters is very large. Furthermore, the prior setting for the probability and expected size of a break influences the posterior results of the timing and size of breaks. A prior which does not match the true breaking process in either the intercept, regression parameters or the variance influences posterior inference on the breaks both in the mean equation as well as in the variance equation.

For our empirical application, unreported posterior results show that, if we take a smaller prior value for ω_0 (i.e., the break size in the intercept), the posterior estimates of the intercept exhibit more frequent but smaller breaks than for the prior setting in Table 2. As a consequence, this alternative prior setting results in a posterior mean of the intercept parameter β_{0t} that follows the inflation series more closely and this results in a sharp decrease in the marginal posterior variable inclusion probabilities δ_j, as this pattern in the intercept erroneously reduces the contribution of the predictor variables in explaining inflation. If we set smaller prior values for b_j and higher values for ω_j $(j = 1, ..., k)$ for the regression parameters of the predictor variables than those outlined in Table 2, implying a higher prior probability of exhibiting breaks of a higher prior expected size, than the corresponding posterior parameter estimates will exhibit many small breaks and the marginal posterior variable inclusion probabilities are low. The prior setting in Table 2 turns out to be robust to different types of instability in the data and the data are still informative about the model parameters. We discuss this in more detail in the next subsection, but we observe, for example, that the posterior means of the probability of a break in the intercept π_0 are smaller than the prior means for both inflation rates at $h = 1$ and $h = 5$, and for the other regression parameters the posterior means of π_j are always higher than the prior mean.

Posterior densities for the different variants of our BMA approach, see Table 1, are computed using the prior calibrations from Table 2 and are based on 24000 iterations of the MCMC sampler outlined in Section 2.3. Of these, we take 4000 iterations as the burn-in period for the sampler and then select every 2nd draw of the remaining posterior draws. However, in case of the BMA-RWB-RWV specification we use 44000 posterior draws with a burn-in sample of 4000 and retaining every 4th draw of the remaining iterations. These choices for the number of retained draws to be used for posterior inference are based on the outcome of a range of convergence tests; see Appendix D for more details.

3.3 Some Implied Inflation Characteristics

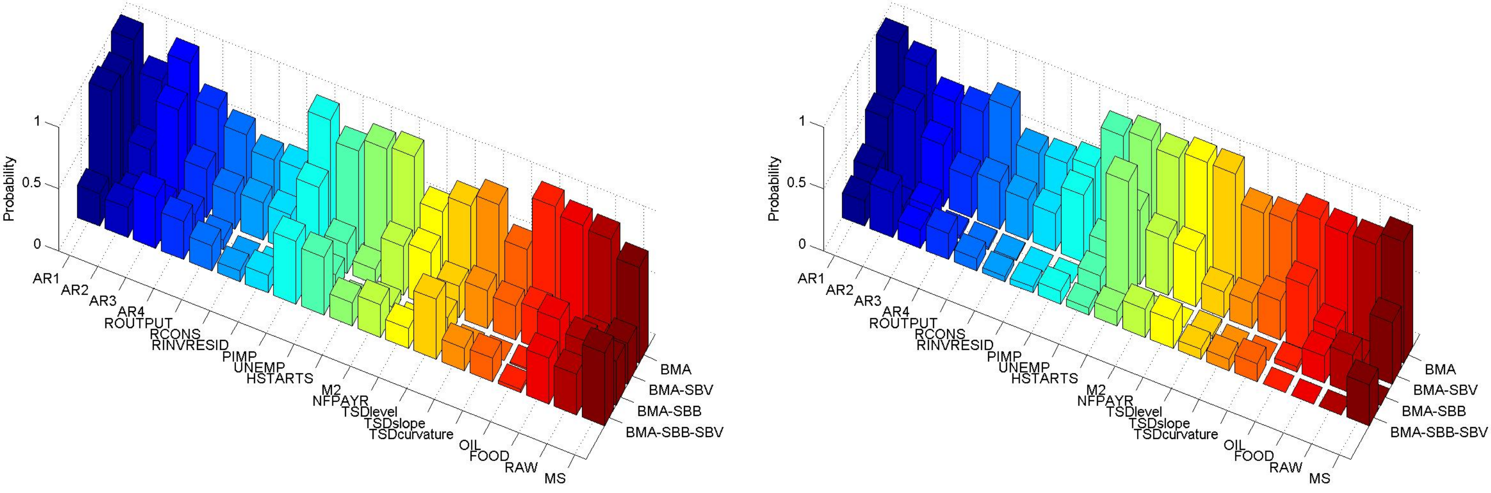
In this subsection we report on some inflation characteristics as implied by estimates of different specifications of our general forecasting model (2)–(3) over our full sample, 1960Q1–2011Q2, for both the PCE deflator and GDP deflator inflation measures. The purpose of this full-sample estimation is to investigate *ex-post* the properties of the different BMA-based specifications, such as an analysis of the relevance of the different predictor variables for inflation and whether our specifications can replicate some of the stylized facts of post-WWII U.S. inflation dynamics that have been uncovered by the vast literature we surveyed in the introduction. For the full sample estimation, we use the 2011Q3 data vintage for all variables, which contains data up to in 2011Q2. We focus here, as well as in the forecasting Section 4, on the most frequently used prediction horizons in this literature, that is, the current quarter or nowcasting horizon ($h = 1$) and the one-year horizon ($h = 5$).[4]

In Figure 1 we display the marginal posterior inclusion probabilities for each of our potential predictor variables (excluding the intercept), that is, $\Pr[\delta_j = 1|y, x]$ for $j = 1, \ldots, k$, for a number of variants of our framework outlined in Section 2.1. We notice some interesting contrasts across these BMA-based specifications, which highlight the importance of conditioning predictor variable selection and model averaging on structural breaks. When structural breaks are ignored in the variable selection (this is the BMA variant), the inclusion probabilities are generally higher than the prior value, with the average inclusion probability hovering around 80% across inflation rates and horizons which corresponds to an average model size of about 15 to 16 predictors. The other extreme is where we combine BMA with structural breaks in the regression parameters and a constant error variance specification (BMA-SBB): inclusion probabilities are now much lower than the prior value and the average inclusion probabilities range from 10% for PCE inflation at $h = 1$ to 17% for GDP inflation at $h = 1$, suggesting an expected number of predictors of about 2 to 3.

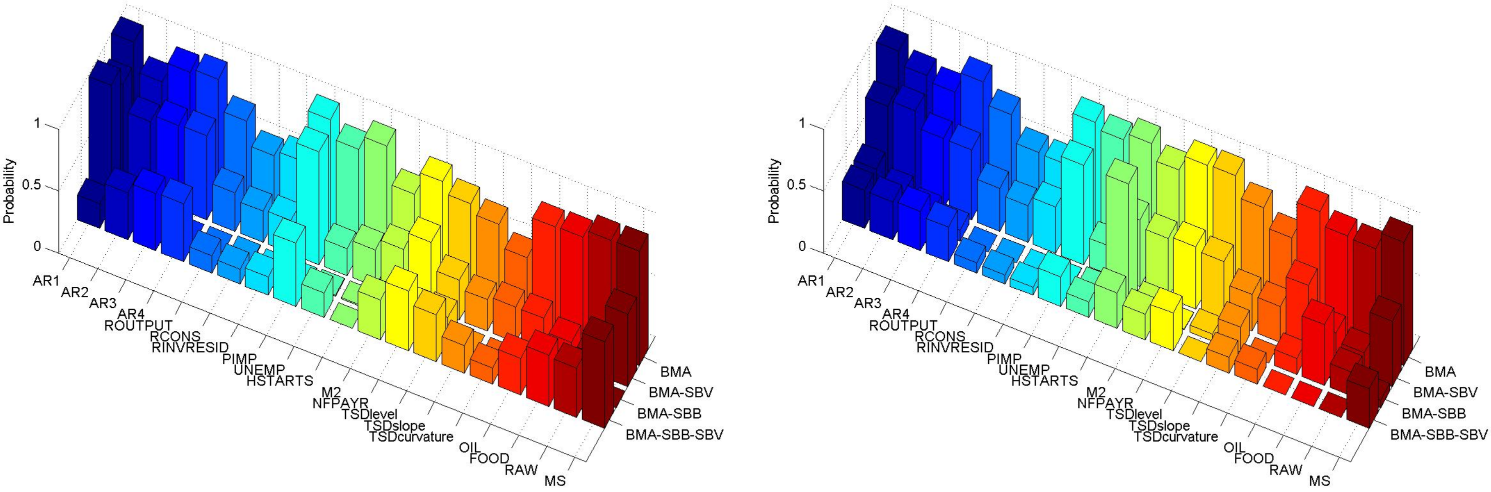
The BMA-SBV and, more general, BMA-SBB-SBV specifications represent intermediate cases. For the BMA-SBV variant the average marginal inclusion probability in Figure 1 is approximately 39% for $h = 1$ and 42% for $h = 5$. In case of the BMA-SBB-SBV model these average marginal inclusion probabilities are about 30% for $h = 1$ and about 15% for $h = 5$, which suggests that the average model size for the BMA-SBB-SBV varies with the forecasting horizon. For the BMA-SBB-SBV model, the expected number of explanatory variables is in the 5 to 6 range for the current quarter horizon across inflation measures, whereas for $h = 5$ this number fluctuates between 2 and 3. One can conclude from Figure 1 that allowing for structural breaks results in more parsimonious models. In particular when we allow for breaks in the mean and variance, as in the BMA-SBB-SBV variant, the model size seems to adapt more to the forecast horizon, with more parsimony at longer horizons.

[4]More specifically, this means modeling the quarterly percentage change of the relevant price deflator in the current quarter as well as in the quarter one year beyond the current quarter, respectively.

Figure 1: Posterior variable inclusion probabilities: 1960Q1-2011Q2



(a) PCE deflator inflation, $h = 1$ – PCE deflator inflation, $h = 5$



(b) GDP deflator inflation, $h = 1$ – GDP deflator inflation, $h = 5$

Notes: The graphs depict the variable selection probabilities for specification BMA, BMA-SBB, BMA-SBV, BMA-SBB-SBV of (2), see also Table 1.

To shed more light on what combinations of explanatory variables dominate the Bayesian model averaging, we can consider the top 10 models with the highest posterior model probability. In the interest of brevity we report these only for the BMA-SBB-SBV variant of (2), see Table 3. In general, the conclusions drawn from the results in Figure 1 are confirmed, that is, the most selected variables do show up most frequently amongst those top 10 models. For example, for PCE and GDP deflator inflation the inflation expectations of the Michigan Consumer survey is often part of the 10 best models, especially for $h = 1$. And the real spot oil price as well as the real food and raw material commodities price indices are not included in the best 10 models for $h = 5$ but are sometimes included in the models for $h = 1$. Furthermore, we see again that the number of included predictors in the models with $h = 1$ is in general larger than for the models with $h = 5$. The second best model for GDP deflator inflation for $h = 5$, for example, only contains an intercept and the Michigan survey. The role of the lagged values of inflation is more important in models which describe quarterly inflation one year ahead than in the models which describe the current quarter inflation rate.

Unreported results[5] show that for both inflation series the model size for all selected models within the BMA-SBB-SBV specification for the current quarter horizon fluctuates between 2 and 14 with a posterior mode of 6 selected predictors. For the quarterly inflation rate one-year ahead, the model size fluctuates for GDP deflator inflation between 1 and 9 (1 and 8 for PCE deflator inflation) and the posterior mode is 4 (3 for PCE deflator inflation). Finally, the number of models with a posterior model probability larger than 0.1% is about 250 for both inflation series and both horizons. The variances in the values of the posterior model probabilities are therefore larger for the models with the larger forecast horizon.

The previous discussion might give some *a priori* insight in the potential success of our BMA-based specifications in forecasting inflation out-of-sample, despite the fact that they, in principle, are heavily parameterized models (see also the predictive density (11)). In general, it is well known that Bayesian model selection prevents overfitting of the data and leads to parsimonious models. For example, the Bayesian information criterion [BIC] contains a penalty function for the number of parameters and the BIC values for different models can be used to approximate their respective Bayes factors. It is also well known that marginal likelihoods and one-step ahead predictive likelihoods are related, see, e.g., Geweke (2005, p. 67), and hence Bayesian model selection also takes into account the forecasting performance of a model. Therefore, our finding that BMA-based approaches combined with structural breaks will average over quite parsimonious regression specifi-

[5] These are available upon request from the authors.

Table 3: BMA-SBB-SBV posterior model probabilities (top 10 best models): 1960Q1-2011Q2

GDP Deflator inflation $h = 1$	
y_{t-2} ROUTP RCONS RINVR PIMP UNEMPL OIL MS	0.69
y_{t-2} y_{t-3} ROUTP RCONS PIMP UNEMPL NFPR MS	0.66
y_t y_{t-3} ROUTP RCONS PIMP HSTS MS	0.50
ROUTP RCONS HSTS M2 FOOD MS	0.44
y_{t-3} RCONS HSTS FOOD MS	0.43
y_{t-2} RINVR HSTS NFPR MS	0.40
y_t y_{t-1} y_{t-2} RINVR HSTS NFPR YL MS	0.39
y_t y_{t-1} y_{t-2} RINVR HSTS NFPR YL MS	0.39
y_{t-2} RCONS RINVR PIMP UNEMPL YL RAW	0.38
y_t y_{t-1} y_{t-2} NFPR	0.38
GDP Deflator inflation $h = 5$	
y_t y_{t-2} y_{t-3} RCONS PIMP UNEMPL HSTS	4.98
MS	3.09
y_t y_{t-1} y_{t-2} y_{t-3} ROUTP PIMP UNEMPL HSTS	2.11
y_{t-1} RCONS PIMP UNEMPL HSTS	1.86
y_{t-1} y_{t-2} CS MS	1.52
y_{t-1}	1.31
y_t y_{t-2} y_{t-3} ROUTP RCONS RINVR HSTS M2 NFPR	1.07
y_t y_{t-1} y_{t-2} y_{t-3} RCONS PIMP UNEMPL HSTS	1.03
y_t PIMP HSTS M2 NFPR	0.96
y_{t-1} y_{t-3} RCONS RINVR M2 NFPR YL TS MS	0.94
PCE inflation $h = 1$	
y_{t-3} ROUTP RINVR PIMP YL	0.56
y_t ROUTP RCONS RINVR M2 CS	0.54
y_t y_{t-1} y_{t-3} ROUTP RCONS RINVR PIMP HSTS NFPR YL FOOD MS	0.54
y_t ROUTP RCONS RINVR M2 CS	0.53
y_t y_{t-1} y_{t-3} ROUTP RCONS RINVR PIMP HSTS NFPR YL FOOD MS	0.53
y_{t-1} RCONS PIMP NFPR YL MS	0.51
y_{t-3} ROUTP UNEMPL HSTS	0.47
y_{t-1} ROUTP PIMP UNEMPL HSTS	0.45
y_t y_{t-1} y_{t-2} RCONS PIMP UNEMPL NFPR CS MS	0.41
y_{t-1} ROUTP RCONS RINVR PIMP M2 YL 0.41	
PCE inflation $h = 5$	
y_t y_{t-3} ROUTP PIMP UNEMPL HSTS	4.75
y_{t-2}	4.54
y_{t-1}	2.91
y_t y_{t-2} ROUTP PIMP UNEMPL HSTS	2.82
y_{t-1} y_{t-3} RCONS UNEMPL NFPR TS MS	2.46
y_t	2.35
y_t y_{t-1} y_{t-2} y_{t-3}	2.30
y_t y_{t-2} y_{t-3} RCONS UNEMPL HSTS M2	1.88
y_t y_{t-1} ROUTP RINVR YL TS MS	1.69
y_{t-1} y_{t-3} RCONS UNEMPL HSTS M2	1.62

Notes: See Section 3.1 for variable mnemonics.

Table 4: BMA-SBB-SBV marginal posterior break probabilities π_j: 1960Q1-2011Q2

	PCE Deflator Inflation		*GDP Deflator Inflation*	
	$h = 1$	$h = 5$	$h = 1$	$h = 5$
Intercept	0.01	0.01	0.00	0.02
y_t	0.18	0.26	0.19	0.23
y_{t-1}	0.18	0.24	0.15	0.21
y_{t-2}	0.12	0.25	0.13	0.19
y_{t-3}	0.12	0.19	0.09	0.15
ROUTP	0.14	0.23	0.15	0.20
RCONS	0.14	0.37	0.14	0.23
RINVR	0.12	0.22	0.10	0.14
PIMP	0.10	0.21	0.09	0.15
UNEMPL	0.08	0.21	0.18	0.19
HSTS	0.09	0.46	0.37	0.22
NFPR	0.13	0.15	0.09	0.17
M2	0.14	0.16	0.11	0.22
YL	0.10	0.13	0.09	0.22
TS	0.17	0.21	0.17	0.17
CS	0.15	0.20	0.13	0.14
OIL	0.26	0.16	0.06	0.17
FOOD	0.15	0.27	0.08	0.21
RAW	0.10	0.29	0.04	0.04
MS	0.08	0.11	0.04	0.10
σ_t	0.29	0.29	0.29	0.29

Notes: See Section 3.1 for variable mnemonics. The values in the table are posterior means of π_j for each $j = 0, \ldots, k + 1$.

cations, especially in case of BMA-SBB-SBV at $h = 5$, might lead us to suspect that they could do well in an out-of-sample context. Of course, in the end this is an empirical question and we will revisit this in Section 4.3.

An analysis of the regression parameters and latent breaks for all 19 predictor variables would be based on their marginal posterior distributions, which are averaged over all possible model specifications, each of which may contain different predictor variables. Therefore, in each of the individual regression specifications the regression parameter β_{jt} of a predictor j as well as the corresponding latent break variable κ_{jt} (which determines the timing of a break in β_{jt}) may be different. This complicates their interpretation unless we condition on each individual regression specification, which is an arduous task given the number of specifications we consider in this paper.[6] Nonetheless, there is some merit in analyzing the posterior means of the individual break probabilities π_j, which are averaged across all entertained model specifications, as it allows us to assess how informative the data has been in driving the posterior time variation in our BMA-based specifications. We report these in Table 4.[7]

What immediately becomes clear from Table 4 is that at 1% the posterior means of the break probability of the intercept π_0 are well below the prior mean of 9% implied by the prior settings in Table 2. In contrast, for our predictor variables the posterior means of the corresponding π_j parameters is on average equal to 13% for both inflation series at $h = 1$, and for $h = 5$ they average to about 23% (18%) in case of PCE deflator inflation (GDP deflator inflation). It is clear that this implies far more time-variation, on average, in the regression parameters of our predictor variables than what is implied by our prior settings, as these suggest that the corresponding prior mean break probabilities are about 2%. Only for the error variance the posterior mean break probability remains fairly close to its prior counterpart. The patterns in Table 4 suggest that parameter time-variation for the predictor variables increases with the forecast horizon, in particular for the real activity series, and that there is quite a bit of heterogeneity amongst these variables. For example, for PCE deflator inflation at $h = 1$, the posterior results suggests that the regression parameter of the real oil price will break on average, over time and across all possible regression specifications, about every 4 quarters, whereas for the survey-based inflation expectations this is about every 13 quarters.

For the other BMA specifications with discrete breaks, BMA-SBB and BMA-SBV, unreported results indicate qualitatively similar conclusions as for the BMA-SBB-SBV case in Table 4. In case of BMA-SBB, the posterior mean break probability for the intercept is essentially zero for the intercept and the average across predictor variables is about 17% for both inflation series at $h = 1$, 28% (20%) for PCE deflator inflation (GDP deflator inflation) at $h = 5$, with a similar degree of heterogeneity across predictors as for BMA-SBB-SBV. The error standard deviation σ_t for the BMA-SBV specification has a posterior mean break probability of 15% across both inflation rates and horizons, which is below the corresponding prior mean of 28% based on the settings in Table 2. Overall, we can conclude that the data appears to be very informative within our BMA-based approach for posterior analysis on instability in all possible model specifications for inflation forecasting.

[6]For example, our earlier analysis about variable selection makes it clear that in case of the BMA-SBB-SBV specification we would have to do this for at least 250 models.

[7]We should note that if a variable is not selected, the posterior distribution of π_j is equal to the prior distribution.

The existing literature has focused a lot on documenting the time-varying properties of inflation. Based on this literature, which we surveyed in the Introduction, we can summarize these properties as follows:

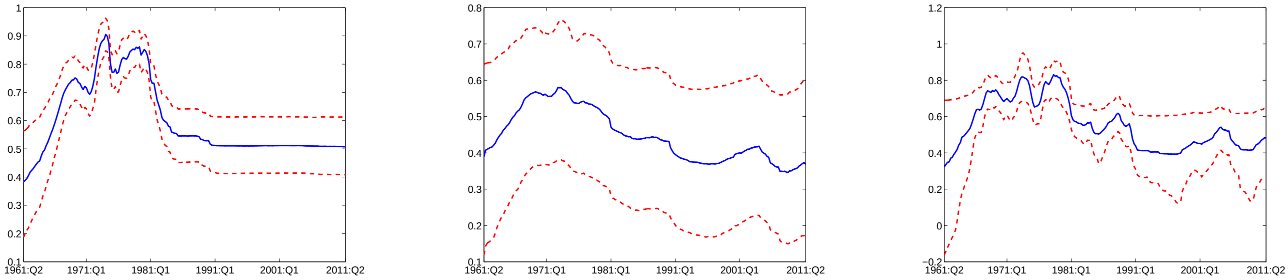
- The persistence of U.S. inflation increased during the 1970s, reaching peaks around 1974-1975 and around 1980, and subsided after 1982–1983; see, e.g., Cogley and Sargent (2005), Cogley *et al.* (2010).
- A majority of studies suggest that the downward shift in U.S. inflation variability form the late 1980s, early 1990s onwards has been due to an exogenous variance break unrelated to breaks in the mean and/or persistence; some studies claim this also happened during the 1970s; see, e.g., Sensier and van Dijk (2004), Sims and Zha (2006).

It is now our aim to investigate whether our BMA-based framework from Section 2.1 indeed implies similar inflation properties. The main purpose of this is to reassure us that our framework does not provide counterintuitive results, which would influence the interpretation of our real-time forecasting exercise in Section 4.

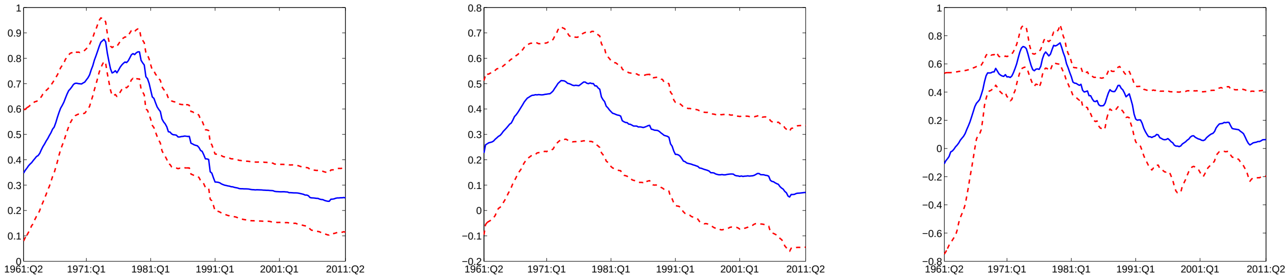
Figure 2 shows posterior estimates of inflation persistence for $h = 1$ given by the BMA-SBB and BMA-SBB-SBV specifications of our framework, both of which allow for time variation (in the form of breaks) in this persistence.[8] In this figure we also display similar posterior results for a version of BMA-SBB-SBV where x_t contains only $(y_t, y_{t-1}, y_{t-2}, y_{t-3})$, which we will call the AR-BMA-SBB-SBV specification. The time-varying average persistence and error variance terms produced by this AR-BMA-SBB-SBV model can be seen as representative of those produced by existing studies, where one usually allows for structural change but does not condition on a large set (of combinations) of additional explanatory variables. Additionally, Figure 3 displays posterior estimates for the innovation standard deviations σ_t for $h = 1$ for model specifications BMA-SBV, BMA-SBB-SBV and AR-BMA-SBB-SBV, respectively.

[8] In order to save space, we focus only $h = 1$, as this is comparable to the representations that are typically used to assess the time-varying properties of inflation. The results for $h = 5$ are available from the authors.

Figure 2: Posterior estimates of time-varying inflation persistence



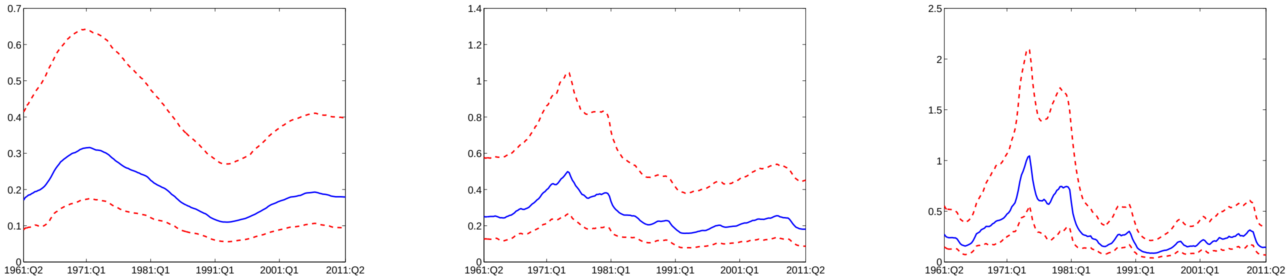
GDP deflator, $h = 1$



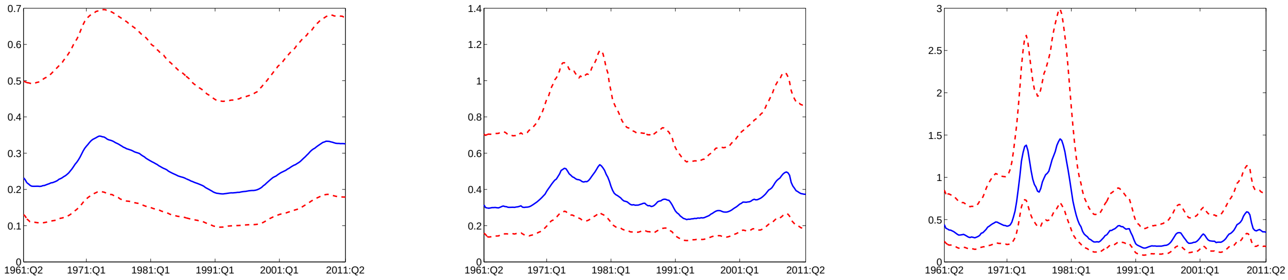
PCE deflator, $h = 1$

Note: The solid lines represent the posterior medians of the persistence parameters for the BMA-SBB (first column), BMA-SBB-SBV (second column) and AR-BMA-SBB-SBV (third column). Persistence is computed by averaging the sum of the included autoregressive parameters across all model specifications using the posterior model probabilities. The dashed lines denote the 25th and 75th percentiles of the posterior distributions.

Figure 3: Posterior estimates of the time-varying innovation standard deviation



GDP deflator, $h = 1$



PCE deflator, $h = 1$

Note: The solid lines represent the posterior medians of the innovation standard deviation parameters for the BMA-SBV (first column), BMA-SBB-SBV (second column) and AR-BMA-SBB-SBV (third column). The dashed lines denote the 25th and 75th percentiles of the posterior distributions.

Figure 2 suggests a similar pattern for inflation persistence across models: relatively low in the 1960's, a substantial increase in the 1970's, and reducing drastically in the second part of the 1980's. In case of the shock variance (Figure 3), all models exhibit a downward shift from the late 1980s, early 1990s onwards. This variance moves up, of course, towards the end of the sample as the 2007-2009 Great Recession starts to impact the data. Both figures make it clear that members of our BMA family of models with appropriate channels of time-variation are able to reproduce in a satisfactorily manner the stylized facts on inflation persistence and volatility for the post-WWII U.S. period.

4 Real-Time Prediction of U.S. Inflation Rates

In this section we focus on the out-of-sample forecasting performance of our BMA family of inflation models relative to other, often very parsimonious, models that are frequently used for inflation forecasting. Section 4.1 provides an overview of these alternative models, whereas the evaluation methodology is discussed in detail in Section 4.2. A discussion of the out-of-sample forecasting results follows in Section 4.3.

4.1 Forecasting Models

The starting point of our forecasting exercise will be the different variants of the BMA framework outlined in Section 2.1: BMA across all possible models from x_t that allows for structural breaks in regression parameters and error variance (BMA-SBB-SBV), BMA across all possible models from x_t allowing for structural breaks in error variance only (BMA-SBV), BMA across all possible models from x_t allowing for structural breaks in the regression parameters only (BMA-SBB), BMA across all possible models from x_t allowing for random walk-based time-variation in the regression parameters and error variance (BMA-RWB-RWV), and, finally, BMA across all possible models from x_t without allowing for any structural change (BMA). We refer to Table 1 for an overview.

To assess how our different BMA models perform in real-time, we need to compare them with viable alternative inflation forecasting models. First amongst these models is the Atkeson and Ohanian (2001) random walk model (RW), which is traditionally seen as a hard-to-beat model when it comes to out-of-sample inflation prediction. More specifically this specification assumes that h-quarter ahead inflation is most optimally predicted by the average inflation rate over the last 4 quarters in a given vintage of inflation data, i.e.,

$$y_{t+h} = \frac{1}{4}\sum_{i=0}^{3} y_{t-i} + \sigma\varepsilon_{t+h}; \quad \varepsilon_{t+h} \sim \text{NID}(0,1). \tag{12}$$

Also, time-invariant autoregressive specifications for inflation, using lag orders between 1 and 4, are considered as parsimonious alternatives:

$$y_{t+h} = \beta_0 + \sum_{j=1}^{p*} \beta_j y_{t-j+1} + \sigma\varepsilon_t; \quad \varepsilon_{t+h} \sim \text{NID}(0,1), \tag{13}$$

with $p*$ is the optimal lag order as determined by the Bayesian-Schwarz information criterion (BIC) across lag orders up till 4 and thus we will refer to (13) as BIC-AR. Finally, we take as a simple representation of the traditional Phillips curve model a version of (13) augmented with the current unemployment rate UNEMPL_t (BIC-AR-UR) in a given data vintage, which gives

$$y_{t+h} = \beta_0 + \sum_{j=1}^{p*} \beta_j y_{t-j+1} + \beta_{p*+1}\text{UNEMPL}_t + \sigma\varepsilon_{t+h}; \quad \varepsilon_{t+h} \sim \text{NID}(0,1), \tag{14}$$

where again $p*$ is the optimal lag order according to BIC.

The aforementioned approaches are naive in the sense that they do not allow explicitly for structural instability in the forecasting relationship, and also do not incorporate information from many, additional, predictors. A simple way to achieve the latter, which in practise has been found to have good forecasting properties, is the ridge regression approach. It encompasses a simple linear regression of the quarter-to-quarter inflation rate

h quarters ahead on all 15 predictor variables described in the previous section plus four inflation lags, that is,

$$y_{t+h} = \beta_0 + X_t'\boldsymbol{\beta_j} + \sigma\varepsilon_{t+h}, \tag{15}$$

where the predictor variables are assembled in the $k \times 1$ vector $X_t = (x_{1t}, \ldots, x_{kt})'$ and $\varepsilon_{t+h} \sim \text{NID}(0, 1)$. The $k \times 1$ parameter vector $\boldsymbol{\beta_j}$ in (15) could now be estimated using a ridge regression (shrinkage) estimator:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \left(\sum_{t=1}^{T-h} \tilde{X}_t\tilde{X}_t' + \lambda I_k\right)^{-1} \left(\sum_{t=1}^{T-h} \tilde{X}_t\tilde{y}_{t+h}'\right) \tag{16}$$

where $\tilde{y}_{t+h},\ \tilde{X}_t$ are demeaned inflation and predictor variables, respectively, to account for the effect of the intercept term, and λ is the scalar shrinkage parameter.

De Mol *et al.* (2008) shows that a ridge regression like (15)–(16) boils down to a regression of, in our case, h-quarter ahead quarterly inflation on a weighted average of all the principal components from X_t, which can be used to model the impact of the unobserved common factors in X_t on y_{t+h}. We choose $\lambda = 10$ in (16), as De Mol *et al.* (2008) show that the degree of shrinkage should be proportional to the number of regressors to achieve the best forecasting performance in a data-rich context. A Gibbs sampler based on an independent Normal-Gamma prior is utilized to estimate (15)–(16), with non-informative priors on the intercept and σ^2, and a shrinkage normal prior on the remaining regression parameters $\text{N}(\mathbf{0}, \lambda^{-1}I_k)$.

We also want to assess our range of BMA inflation forecasting models to alternatives that explicitly model structural instability in the data. One such alternative is simply a variation on our BMA-SBB-SBV model where we average solely over the inflation lags and hence $x_t = (y_t, y_{t-1}, y_{t-2}, y_{t-3})$ as discussed before in Section 3.3:

$$y_{t+h} = \beta_{0t} + \sum_{j=1}^{4} \delta_j\beta_{jt}y_{t-j+1} + \sigma_t\varepsilon_{t+h} \text{ with } \varepsilon_{t+1} \sim \text{NID}(0, 1), \tag{17}$$

together with (3)–(4). Hence, in this AR-BMA-SBB-SBV specification we use our stochastic variable selection algorithm to determine the posterior probabilities of all possible inflation lag combinations and use these posterior probabilities for BMA.

Alternatively, we can use the resulting posterior model probabilities that correspond with (17) to select the 'best' lag order for a given vintage of inflation data at forecast horizon h and use the resulting model to forecast y_{t+h}; we denote this as the AR-BVS-SBB-SBV model (with BVS standing for Bayesian variable selection). We can apply this approach also to our full set of 15 predictor variables and 4 quarterly inflation lags; that is using the model posterior probabilities that result from the BMA-SBB-SBV MCMC algorithm to select the best combination of predictor variables and inflation lags for the

Table 5: Alternative models for forecasting inflation

name	description	specification
RW	Random walk	(12)
BIC-AR	Autoregressive model with BIC lag order selection	(13)
BIC-AR-UR	Simple Phillips curve model with AR and BIC lag order selection	(14)
Ridge	Ridge regression with $\lambda = 10$	(15)–(16)
AR-BMA-SBB-SBV	BMA-SBB-SBV specification with only 4 inflation lags in x_t	(17)
AR-BVS-SBB-SBV	Specification of AR-BMA-SBB-SBV with highest posterior probability	(17)
BVS-SBB-SBV	Specification of BMA-SBB-SBV with highest posterior probability	(2)
UCSV	Unobserved component model with SV	(18)

h-quarter ahead quarter-to-quarter inflation rate in a given vintage of data, and use only this model to predict with stochastic breaks y_{t+h}. This approach will be labeled as the BVS-SBB-SBV model.

Another option is an inflation forecast model that has been successfully used by Stock and Watson (2007, 2008) to predict inflation. They propose an unobserved components model with stochastic volatility [UCSV] specification for the unobserved component of inflation as well as the temporary deviation from it, i.e.,

$$\begin{array}{ll} y_t = \beta_t + \sigma_t \varepsilon_t & \ln \sigma_t^2 = \ln \sigma_{t-1}^2 + u_{1t} \\ \beta_t = \beta_{t-1} + \omega_t \eta_t & \ln \omega_t^2 = \ln \omega_{t-1}^2 + u_{2t}, \end{array} \tag{18}$$

where $\varepsilon_t \sim \text{NID}(0,1)$, $\eta_t \sim \text{NID}(0,1)$ and $u_t = (u_{1t}\ u_{2t})' \sim \text{NID}(\mathbf{0}, \rho I_2)$ with ρ a scalar parameter controlling the smoothness of the stochastic volatility processes, and where ε_t, η_t and u_t are independent. We follow Stock and Watson (2007) and set in (18) $\rho = 0.04$; Stock and Watson (2007) motivate their choice for ρ based on the fit of (18) for U.S. inflation rates over the 1955–2004 sample. The resulting forecast for y_{t+h} is set equal to the filtered estimate of β_t from (18) for a given data vintage of quarterly inflation rates.

Table 5 summarizes the above described range of models that serve as alternatives to our BMA family of inflation forecasting models.

4.2 Forecast Evaluation Methodology

We use the models described in the previous subsection to obtain and evaluate one-quarter and one-year ahead forecasts for the quarter-on-quarter inflation rate of both the PCE deflator and the GDP deflator in the United States. For computational reasons we obtain the one-year ahead forecasts through direct forecasting.[9] Each forecast is based on a re-estimation of the underlying model using an expanding window of historical data. For example, suppose the first one-quarter and one-year ahead forecasts are produced in quarter t_0. As we want to evaluate the forecasts in real-time, we use the original vintage of data available at t_0, containing data up to $t_0 - 1$, to re-estimate the range of models on the

[9] Whether an iterative procedure provides more accurate forecasts than a direct approach is a matter of ongoing debate, see the discussion in Marcellino *et al.* (2006).

sample $t = 1, \dots, t_0 - h$, with the forecast horizon $h = 1$ or 5. This results in a quarterly inflation nowcast for quarter t_0 (horizon $h = 1$) and a direct quarterly inflation forecast for quarter $t_0 + 4$ (horizon $h = 5$) using data on x_{jt} for $t = 1, \dots, t_0 - 1$ and the parameter estimates or the posterior draws from the estimation up to $t_0 - h$[10]. These forecasts are evaluated against the vintage of inflation data that is available h quarters ahead, i.e., the vintage at $t_0 + h$. We repeat this process of re-estimation and forecast generation for $t_0 + 1, \dots, T - h$. In the end we will have time series of forecasts (and in most cases their distribution) and forecast errors for $t = t_0, \dots, T - h$, which we use to compute several forecast evaluation criteria.

We follow the usual practise in the literature, and evaluate the point forecasts from the different models using both the square root of the mean squared forecast errors [RMSE] and the mean of the absolute forecast errors [MAE]. In algebraic terms these can be represented as, respectively,

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{T - t_0 - h} \sum_{s=t_0-1}^{T-h} \hat{\epsilon}_{s+h}^2}, \tag{19}$$

and

$$\text{MAE} = \frac{1}{T - t_0 - h} \sum_{s=t_0-1}^{T-h} |\hat{\epsilon}_{s+h}|, \tag{20}$$

where the out-of-sample forecast error of a model for y_{t+h} is defined $\hat{\epsilon}_{t+h}$. The analysis in Gneiting (2011) indicates that the RMSE is only a consistent evaluation measure when the point forecast equals the mean of the distribution of forecasts, whereas the MAE is consistent if the point forecast equals the median of the predictive distribution.

In case of a conditionally Gaussian distribution such as that implied by the RW model (12) this distinction is not relevant as both the mean and median of the forecasts at t equals $\frac{1}{4} \sum_{i=0}^{3} y_{t-i}$. However, when the forecasts are based on a non-Gaussian distribution or posterior draws from a Gibbs sampler, such as in case of our family of BMA inflation models, we will have to base the RMSE and MAE on, respectively, the mean and median of the distribution of inflation predictions.

Any model is a mere approximation of reality, and thus to some degree misspecified, and we therefore assess the forecasting performance of the models relative to a benchmark model. Given the results in Stock and Watson (2007, 2008) it appears that the UCSV model produces forecasts for U.S. inflation rates that are hard to beat by Phillips curve and naive time series models (including the RW model). We therefore assess real-time

[10] Except for the RW, BIC-AR and BIC-AR-UR models, all other models are estimated using MCMC simulation procedures. In case of the UCSV model, we use the Gibbs sampler as has been kindly made available on Mark Watson's homepage.

inflation point forecasts in terms of the RMSE and MAE ratios for a model relative to those of the UCSV, where a RMSE or MAE ratio smaller than 1 suggests that the mean or median forecast of a model is better than that generated by the UCSV model.

Any assessment whether the resulting differences in either RMSE or MAE are statistically significant or not is complicated by the fact that a lot of the models outlined in Section 4.1 are nested, and in combination with parameter updating using an expanding window of historical data this means that a test statistic like the one proposed by Diebold and Mariano (1995) will have a non-standard limiting distribution; see Clark and McCracken (2012) for an overview. Like Faust and Wright (2011) we follow a more pragmatic approach, however, and build on Monte Carlo results in Clark and McCracken (2012). They find that comparing the Harvey *et al.* (1997) small sample correction of the Diebold and Mariano (1995) statistic to standard normal critical values results in a good sized test of the null hypothesis of equal finite sample forecast precision for both nested and non-nested models, including cases with expanded window-based model updating.

More specifically, we follow Harvey *et al.* (1997) and construct for each model relative to forecast errors from the UCSV model the statistic:

$$t_{HLN} = \sqrt{\frac{(T-t_0-h)+1-2h+(T-t_0-h)^{-1}h(h-1)}{(T-t_0-h)}}\, t_{DM};$$

$$t_{DM} = \sqrt{T-t_0-h}\left(\frac{\bar{d}_l - \bar{d}_{\mathrm{UCSV}}}{\hat{\sigma}_{rect}}\right) \quad (21)$$

with for each model l $\bar{d}_l = \mathrm{MSE}_l$ or MAE_l and similarly for $\bar{d}_{\mathrm{UCSV}}$, and t_{DM} is the Diebold and Mariano (1995) test statistic. In (21) $\hat{\sigma}_{rect}$ is a standard deviation based on a heteroskedasticity and autocorrelation corrected [HAC] estimate of the variance $\hat{\sigma}^2_{rect}$ of either the squared forecast error differential (in case of RMSEs) or the absolute forecast error differential (in case of MAEs). The HAC variance estimate encompasses a rectangular window of lags of the aforementioned differentials with a lag truncation parameter set to h (see, e.g., Clark and McCracken (2012) for more details). Once we constructed t_{HLN} for a model, we compare its value to one-sided critical values from the standard normal distribution, and thus we test the null of equal finite sample forecast accuracy based on a certain evaluation measure $H_0: \bar{d}_l - \bar{d}_{\mathrm{UCSV}} = 0$ *vis-à-vis* the alternative that model l has been more accurate, i.e., $H_1: \bar{d}_l - \bar{d}_{\mathrm{UCSV}} < 0$.

Point forecasts of inflation, however, cannot provide insight into the uncertainty that is associated with producing these forecasts, and inflation density forecasts are more useful for this purpose. Density forecasts of inflation can also be used to assess the ability of a model to predict unusual developments, such as the likelihood of deflation or accelerating inflation given current information. We therefore will also evaluate density inflation forecasts from our family of BMA inflation models (see Table 1) as well as the range of alternative models summarized in Table 5.

There are several measures available for density forecast evaluation of which the log score, or the logarithm of the predictive density evaluated in the realization, is often used, as it suggests a close relationship with the likelihood function of a model. There are, however, also several drawbacks to using the log score in this context: it does not reward values from the predictive density that are close but not equal to the realization (see, e.g., Gneiting and Raftery (2007)), it is very sensitive to outliers, and Gneiting and Ranjan (2011) show that it is invalid to use weighted log scores to emphasize certain areas of the distribution (such as the tails) in density forecast evaluation.

Gneiting and Raftery (2007) and Gneiting and Ranjan (2011) show that the continuous ranked probability score (CRPS) is able to address the aforementioned drawbacks of the log-score and we therefore use this measure to evaluate density forecasts from our inflation models. The CRPS reads like:

$$\begin{aligned} \text{CRPS}(t+h,l) &= \int_{-\infty}^{\infty} (F(z) - \mathtt{I}\{y_{t+h} \leq z\})^2 \, \mathrm{d}z \\ &= \mathtt{E}_f|Y_{t+h,l} - y_{t+h}| - \frac{1}{2}\mathtt{E}_f|Y_{t+h,l} - Y'_{t+h,l}|, \end{aligned} \tag{22}$$

where F is the cumulative distribution function (CDF) that corresponds to the predictive density f of model l estimated at time t, $\mathtt{I}(.)$ takes a value 1 if $y_{t+h} \leqq z$ and equals 0 otherwise, $\mathtt{E}_f$ is the expectation operator, and $Y_{t+h,l}$ and $Y'_{t+h,l}$ are independent random variables with common sampling density equal to the posterior predictive density of model l for y_{t+h} estimated at time t.

It is clear from (22) that the CRPS provides a measure of the distance between the predictive CDF as implied by a model and the CDF of realizations h periods ahead, and thus a relatively low CRPS value reflects a relatively good density forecast. Moreover, the second equality in (22) suggests that the CRPS for a model l in practice can be computed fairly easily in discrete terms by measuring the average absolute distance between the empirical CDF of y_{t+h}, which is simply a step function in y_{t+h}, and the empirical CDF that is associated with model l's predictive density. The latter, in turn, can easily be computed from the posterior draws of a MCMC sampler (draws from Y'_{t+h} can be obtained by random resampling from these draws) or can be computed analytically when a Gaussian approximation is used (as in the case of the RW model).

To assess the density forecasting performance of model l over a certain evaluation sample, one can determine the historical average of (22) across the forecasts, i.e.,

$$\text{avCRPS}_l = \frac{1}{T - t_0 - h} \sum_{s=t_0-1}^{T-h} \widehat{\text{CRPS}}(s+h,l). \tag{23}$$

The real-time inflation density forecasts for a model relative to those of the UCSV is therefore assessed using the ratio of the avCRPS of a model l relative to the avCRPS of the UCSV model, where, a ratio smaller than 1 suggests that density forecasts from model l are better than those from the UCSV model.

Finally, we want to assess how our range of models fares when different areas of their predictive densities are emphasized in the forecast evaluation, such as the tails, as forecasters are often very interested in predicting 'risk events' for inflation, such as the probability of deflation given the current data or the likelihood of accelerating inflation. To do that we will compute weighted averages of Gneiting and Raftery (2007) quantile scores (QS) that are based on quantile forecasts that correspond to the predictive densities from the different models, i.e.,

$$\mathrm{QS}(\alpha, t+h, l) = \left(\mathrm{I}\{y_{t+h} \leqq F^{-1}(\alpha, l)\} - \alpha\right)\left(F^{-1}(\alpha, l) - y_{t+h}\right) \tag{24}$$

with $F^{-1}(\alpha, l)$ is the quantile forecast h periods ahead using model l for level $\alpha \in (0,1)$. It can be shown that integrating (24) over $\alpha \in (0,1)$ will result in the CRPS measure (22), see Gneiting and Ranjan (2011).

Gneiting and Ranjan (2011) therefore propose to integrate weighted versions of (24) over α, with these weights being fixed functions of α chosen such to emphasize in the forecast evaluation a certain area of the underlying forecast density. In practice we will have to use a discrete approximation to this integration and use weights that emphasize the center, the right tail and the left tail of the predictive density of model l relative to the inflation realization y_{t+h}, which are than averaged over the evaluation sample. Thus, we have

$$\begin{aligned}
\text{avQS-C}_l &= \frac{1}{T-t_0-h}\sum_{s=t_0-1}^{T-h}\left(\frac{1}{99}\sum_{j=1}^{99}\alpha_j(1-\alpha_j)\mathrm{QS}(\alpha_j, s+h, l)\right), \\
\text{avQS-R}_l &= \frac{1}{T-t_0-h}\sum_{s=t_0-1}^{T-h}\left(\frac{1}{99}\sum_{j=1}^{99}\alpha_j^2\mathrm{QS}(\alpha_j, s+h, l)\right), \\
\text{avQS-L}_l &= \frac{1}{T-t_0-h}\sum_{s=t_0-1}^{T-h}\left(\frac{1}{99}\sum_{j=1}^{99}(1-\alpha_j)^2\mathrm{QS}(\alpha_j, s+h, l)\right),
\end{aligned} \tag{25}$$

where $\alpha_j = j/100$ and $\mathrm{QS}(\alpha_j, s+h, l)$ is defined in (24). In (25), avQS-C emphasizes the center, avQS-R the right tail, and avQS-L the left tail of the predictive density relative to the realization h periods ahead. We will report ratios of avQS-C, avQS-R and avQS-L measures of a certain model relative to those of the UCSV, where a ratio smaller than 1 indicates that a model generates more precise predictions for certain parts of the (unknown) distribution of future inflation rates.

Based on the theoretical considerations in Amisano and Giacomini (2007), Gneiting and Ranjan (2011) show that a statistic like t_{DM} in (21) can be used to test whether the difference in average CRPS and QS measures for two models is statistically significant by comparing it to standard normal critical values. One of the most important conditions for the Amisano and Giacomini (2007)-Gneiting and Ranjan (2011) result to hold is the presence of non-vanishing parameter uncertainty, as the null is that of equal finite sample forecast accuracy. This is typically only achieved by updating the different forecasting models using fixed, rolling windows of data rather than expanding data windows, as in our analysis. However, as in the case of the RMSE and MAE measures, we take a pragmatic route in which we take the stance that comparing the t_{HLN} from (21), based on either the CRPS, QS-C, QS-R or QS-L differentials relative to the UCSV model, to (one-sided) standard normal critical values will result in a decent sized test for the null-hypothesis of equal finite sample forecasting performance.[11]

4.3 Out-of-Sample Results

The models discussed in Section 4.1 are used to generate quarter-on-quarter PCE deflator and GDP deflator inflation forecasts for the current quarter ($h = 1$) and one-year ahead ($h = 5$). These are evaluated by computing the forecast evaluation measures discussed in Section 4.2 across two periods: 1980Q1–2011Q2 and 1985Q1–2011Q2. These evaluation samples span a number of large events that potentially could have caused time-variation in the dynamics of inflation rates, like, e.g., the 'monetarist experiment' by the Federal Reserve under Volcker, the 'Great Moderation' during the late-1980s, the 1990s and early 2000s, the 9/11 catastrophe in 2001 and, of course, the 'Great Recession' of 2007–2009.

The longer period of 1980Q1–2011Q2 was chosen based on computational considerations: 1979Q4 (i.e., the data from the 1980Q1 vintage) was the furthest we could have gone back without compromising the proper estimation of models with stochastic structural breaks as well as due to the availability of proper real-time data. However, this evaluation sample does encompass the early 1980s, which represents the end of the Great

[11] In an unreported exercise, we ran a Monte Carlo experiment similar to the one in Amisano and Giacomini (2007, Section 5). Amisano and Giacomini (2007) use a DGP in which a random walk model and a Phillips curve model, calibrated on data, have equal finite sample predictive power for the conditional mean of inflation, while imposing identical conditional variance across all models and inflation. We follow their set-up, where we use (12) and (14), recursively estimated in real-time on our data using an expanding window, and asses the size of the t_{DM} and t_{HLN} statistics from (21) for the CRPS and avCRPS statistics using standard normal critical values across these 2 artificial inflation forecasting models. We ran these experiments for both $h = 1$ and $h = 5$ (we impose for the latter horizon a $MA(4)$ in the innovation variance of the artificial inflation rate) at different sample sizes, using a nominal size of 5% and 10000 iterations. In general, the t_{HLN} statistic kept the finite sample size fairly close to the nominal size, whereas in case of the t_{DM} statistic it became quite oversized at $h = 5$ and smaller samples. More detailed results are available from the authors.

Table 6: Forecast Evaluation PCE deflator inflation

	Horizon: $h=1$			Horizon: $h=5$		
	RMSE	MAE	avCRPS	RMSE	MAE	avCRPS
	Forecast evaluation sample: 1980Q1-2011Q2					
BMA-SBB-SBV	**0.91**	**0.95**	**0.87****	1.00	1.04	**0.92**
	(-0.85)	(-0.76)	(-1.90)	(0.02)	(0.39)	(-0.81)
BMA-SBV	1.01	**0.98**	**0.81*****	1.06	1.09	**0.88**
	(0.14)	(-0.32)	(-3.44)	(0.43)	(0.58)	(-0.95)
BMA-SBB	1.03	1.01	**0.83*****	1.13	1.13	**0.99**
	(0.79)	(0.02)	(-3.21)	(1.04)	(0.78)	(-0.11)
BMA	1.09	1.09	**0.89*****	1.37	1.32	1.10
	(2.30)	(1.95)	(-2.40)	(1.56)	(1.56)	(0.58)
BMA-RWB-RWV	**0.99**	**0.95***	**0.82*****	**0.99**	**0.99**	**0.83****
	(-0.31)	(-1.39)	(-4.39)	(-0.12)	(-0.01)	(-2.28)
RW	1.00	**0.97**	1.00	**0.99**	1.03	1.25
	(0.12)	(-0.63)	(0.02)	(-0.03)	(1.18)	(3.05)
BIC-AR	1.04	1.02	1.05	1.03	1.02	1.00
	(3.47)	(0.55)	(0.92)	(0.41)	(0.28)	(0.06)
BIC-AR-UR	1.06	1.03	1.06	1.05	1.03	1.02
	(2.75)	(0.83)	(1.04)	(0.78)	(0.46)	(0.22)
Ridge	**0.99**	**0.95**	3.84	1.15	1.34	21.06
	(-0.22)	(-1.16)	(9.95)	(1.45)	(2.80)	(15.24)
BVS-SBB-SBV	1.17	1.03	**0.97**	1.18	1.02	**0.91**
	(1.66)	(0.51)	(-0.43)	(1.23)	(0.16)	(-0.85)
AR-BMA-SBB-SBV	**0.98**	**0.93**	**0.83*****	**0.96**	**0.99**	**0.84*****
	(-0.84)	(-1.21)	(-3.36)	(-0.95)	(-0.16)	(-2.39)
AR-BVS-SBB-SBV	**0.98**	**0.93***	**0.83*****	**0.94**	**0.99**	**0.79*****
	(-0.90)	(-1.33)	(-3.32)	(-1.13)	(-0.02)	(-3.06)
	Forecast evaluation sample: 1985Q1-2011Q2					
BMA-SBB-SBV	**0.88**	**0.94**	**0.88***	**0.88***	**0.93****	**0.83****
	(-0.94)	(-0.84)	(-1.55)	(-1.45)	(-1.78)	(-1.89)
BMA-SBV	1.03	1.01	**0.85*****	1.03	1.02	**0.85**
	(0.52)	0.18)	(-2.49)	(0.19)	(0.12)	(-1.22)
BMA-SBB	1.02	1.00	**0.86*****	1.02	**0.98**	**0.92**
	(0.80)	(0.01)	(-2.63)	(0.17)	(-0.24)	(-0.87)
BMA	1.08	1.09	**0.90****	1.13	1.11	**0.93**
	(2.19)	(1.71)	(-1.89)	(0.74)	(0.62)	(-0.41)
BMA-RWB-RWV	**0.96****	**0.95***	**0.83*****	**0.93***	**0.93***	**0.78*****
	(-1.68)	(-1.64)	(-3.62)	(-1.53)	(-1.57)	(-3.47)
RW	1.04	1.02	1.08	**0.98**	**0.99**	1.31
	(0.91)	(0.39)	(1.10)	(-0.56)	(-0.25)	(-3.56)
BIC-AR	1.05	1.03	1.12	1.08	1.05	1.06
	(4.01)	(0.77)	(1.83)	(1.22)	(0.67)	(0.80)
BIC-AR-UR	1.06	1.04	1.12	1.11	1.08	1.07
	(2.59)	(1.00)	(1.86)	(1.53)	(0.99)	(1.02)
Ridge	1.03	**0.99**	3.41	1.27	1.35	20.72
	(1.10)	(-0.06)	(11.14)	(2.50)	(2.54)	(26.26)
BVS-SBB-SBV	1.13	1.00	**0.96**	1.00	**0.94**	**0.84****
	(1.13)	(0.02)	(-0.55)	(0.06)	(-0.71)	(-1.93)
AR-BMA-SBB-SBV	1.01	**0.99**	**0.87****	**0.96**	**0.94**	**0.82*****
	(0.30)	(-0.23)	(-2.30)	(-0.68)	(-0.92)	(-2.42)
AR-BVS-SBB-SBV	1.01	**0.99**	**0.88****	**0.92***	**0.92***	**0.76*****
	(0.51)	(-0.28)	(-2.15)	(-1.44)	(-1.28)	(-3.56)

Notes: The numbers are ratios of the RMSE (19), MAE (20) or avCRPS (23) measure relative that of the USCV model; **bold** numbers indicates a better performance than UCSV. In parentheses we report the t_{HLN} statistic (21) for the null hypothesis of equal finite sample prediction accuracy *versus* the alternative hypothesis that a model outperforms UCSV for either of these measures, where *, ** and *** indicates rejection of this null at the 10%, 5% and 1% level, respectively, based on one-sided standard normal critical values. For model mnemonics, see Tables 1 and 5.

Table 7: Forecast Evaluation GDP deflator inflation

	Horizon: $h = 1$			Horizon: $h = 5$		
	RMSE	MAE	avCRPS	RMSE	MAE	avCRPS
	Forecast evaluation sample: 1980Q1-2011Q2					
BMA-SBB-SBV	**0.91*****	**0.90****	**0.97**	**0.99**	**0.99**	**0.93**
	(-2.59)	(-1.84)	(-0.71)	(-0.08)	(-0.02)	(-0.85)
BMA-SBV	**0.95**	**0.92**	**0.83*****	1.04	1.03	**0.86**
	(-1.18)	(-1.23)	(-3.54)	(0.29)	(0.20)	(-1.09)
BMA-SBB	**0.97**	**0.99**	**0.90****	1.40	1.27	1.12
	(-0.62)	(-0.21)	(-1.80)	(1.25)	(1.28)	(1.07)
BMA	1.15	1.08	**0.94**	1.63	1.47	1.24
	(1.37)	(0.90)	(-0.74)	(1.65)	(1.69)	(0.96)
BMA-RWB-RWV	**0.96**	1.00	1.08	**0.96**	1.01	**0.96**
	(-0.81)	(0.03)	(1.22)	(-0.36)	(0.09)	(-0.30)
RW	**0.91***	**0.89***	1.21	**0.96**	**0.96**	1.43
	(-1.48)	(-1.51)	(3.02)	(-0.57)	(-0.58)	(4.43)
BIC-AR	**0.97**	**0.98**	1.32	**0.99**	1.02	1.13
	(-0.52)	(-0.28)	(5.14)	(-0.07)	(0.24)	(1.14)
BIC-AR-UR	**0.96**	**0.97**	1.32	**0.99**	1.00	1.12
	(-0.73)	(-0.40)	(5.10)	(-0.15)	(0.02)	(1.14)
Ridge	**0.96**	1.00	5.76	1.19	1.33	20.83
	(-0.91)	(0.04)	(7.45)	(1.80)	(2.86)	(16.23)
BVS-SBB-SBV	2.18	1.40	1.41	1.42	1.29	1.18
	(13.42)	(3.09)	(4.37)	(2.21)	(2.34)	(1.76)
AR-BMA-SBB-SBV	**0.96**	**0.98**	**0.89****	**0.98**	1.05	**0.89***
	(-0.88)	(-0.30)	(-1.87)	(-0.28)	(0.46)	(-1.39)
AR-BVS-SBB-SBV	2.09	**0.98**	1.03	**0.94**	1.03	**0.89***
	(14.28)	(-0.26)	(0.54)	(-0.56)	(0.28)	(-1.38)
	Forecast evaluation sample: 1985Q1-2011Q2					
BMA-SBB-SBV	**0.85*****	**0.89****	**0.95**	**0.85***	**0.86***	**0.84****
	(-3.71)	(-1.81)	(-0.90)	(-1.63)	(-1.54)	(-1.82)
BMA-SBV	**0.93***	**0.92**	**0.83*****	**0.99**	**0.95**	**0.83**
	(-1.51)	(-1.05)	(-2.80)	(-0.03)	(-0.26)	(-1.07)
BMA-SBB	**0.92***	**0.97**	**0.92***	**0.96**	**0.96**	**0.99**
	(-1.58)	(-0.48)	(-1.35)	(-0.34)	(-0.36)	(-0.09)
BMA	1.03	**0.99**	**0.88***	1.19	1.17	**0.96**
	(0.36)	(-0.03)	(-1.60)	(0.78)	(0.79)	(-0.20)
BMA-RWB-RWV	**0.93**	**0.99**	1.12	**0.94**	**0.97**	**0.99**
	(-0.99)	(-0.06)	(1.62)	(-0.71)	(-0.27)	(-0.10)
RW	**0.88****	**0.91**	1.30	**0.88****	**0.89****	1.56
	(-1.66)	(-1.78)	(3.76)	(-1.82)	(-0.37)	(5.03)
BIC-AR	**0.99**	1.02	1.43	1.05	1.05	1.22
	(-0.18)	(0.21)	(5.94)	(0.67)	(0.42)	(2.06)
BIC-AR-UR	**0.99**	1.04	1.42	1.08	1.06	1.23
	(-0.18)	(0.43)	(5.70)	(0.86)	(0.55)	(2.20)
Ridge	**0.96**	1.03	5.01	1.28	1.30	21.79
	(-0.59)	(0.34)	(6.41)	(2.63)	(2.28)	(21.40)
BVS-SBB-SBV	2.39	1.27	1.34	1.27	1.15	1.11
	(2.39)	(2.32)	(4.10)	(2.02)	(1.80)	(1.10)
AR-BMA-SBB-SBV	**0.96**	1.01	**0.91**	**0.98**	**0.98**	**0.87***
	(-0.69)	(0.07)	(-1.23)	(-0.21)	(-0.17)	(-1.30)
AR-BVS-SBB-SBV	2.25	1.07	1.06	1.00	**0.99**	**0.89**
	(12.88)	(0.80)	(0.90)	(0.04)	(-0.06)	(-1.09)

Note: See the notes for Table 6.

Inflation era but does not convey any information about the start of the Great Inflation. It is for this reason that often in the literature, such as in, e.g., Atkeson and Ohanian (2001) and Faust and Wright (2011), more attention is paid to inflation forecasting in the post-1984 period, also because inflation was deemed to be hard to forecast during the Great Moderation period. Therefore, we also analyze the predictive performance of our inflation models over the 1985Q1–2011Q2 sample. The forecasts of our models are based on estimates of the model parameters and, if relevant, the latent variables computed using an expanding window of data, starting with 1960Q1–1979Q4 using the original 1980Q1 data vintages for the 1980Q1–2011Q2 sample and commencing with 1960Q1–1984Q4 using the original 1985Q1 data vintages for the 1985Q1–2011Q2 sample. The range of forecasts are used to assess the real-time out-of-sample performance of our models relative to the Stock and Watson (2007, 2008) UCSV model (18).

Table 6 report for PCE deflator inflation ratios of the RMSE, MAE, and avCRPS measures from Section 4.2 for our BMA family of inflation models and those summarized in Table 5 relative to these measures produced by the UCSV model. When we focus first on the point forecast evaluation based on the RMSE and MAE measures over the 1980Q1-2011Q2 evaluation sample in the upper panel of this table, it appears that for $h = 1$ only the BMA-RWB-RWV and AR-BVS-SBB-SBV specifications significantly outperformed the UCSV model. For one-year ahead forecasts, the picture is even more bleak as none of the models can produce more precise forecasts than the UCSV-based ones. In terms of density forecasts, however, there is for both horizons evidence based on the CRPS measure that a number of models are producing more precise density forecasts than the UCSV model. All our BMA-based inflation models perform better when it comes to current quarter inflation density predictions, and also the two autoregressive specifications with stochastic structural breaks (AR-BMA-SBB-SBV and AR-BVS-SBB-SBV) perform well in that case. At $h = 5$, the BMA-RWB-RWV, AR-BMA-SBB-SBV and AR-BVS-SBB-SBV models exhibit significantly better density forecasts.

The lower panel of Table 6 reports on the forecast evaluation for PCE deflator inflation over the 1985Q1–2011Q2 period. Over this period, the BMA-RWB-RWV model outperforms the UCSV model both in terms of point forecast accuracy and density forecast accuracy at the two forecast horizons. The BMA-SBB-SBV member of our BMA-range of inflation models performs at least as well as the BMA-RWB-RWV one when it comes to one-year ahead prediction. Although these two specifications model structural instability in a different manner, this only seems to matter *vis-à-vis* the UCSV model for point forecasts of current quarter PCE deflator inflation. The remaining BMA-based inflation models only do well in terms of $h = 1$ inflation density predictions. Of the alternatives to our BMA-based approaches, only the AR-BVS-SBB-SBV model seems to do well relative to point and density forecasts from the UCSV model.

The results of a similar evaluation for GDP deflator inflation is summarized in Table 7. For the 1980Q1–2011Q2 period both our BMA-SBB-SBV specification significantly outperformed the UCSV model in terms of $h = 1$ point forecasts; see the upper panel of Table 7. Interestingly, in terms of current quarter GDP deflator inflation density forecasts the BMA-based specifications that shut down either of the two channels for structural instability, BMA-SBV and BMA-SBB, are the most successful ones according to the avCRPS ratios. Of the alternative forecasting models, the RW model performs well for point forecasting and AR-BMA-SBB-SBV do so for density forecasting. As for PCE deflator inflation, one-year head GDP deflator inflation forecasting for the 1980Q1–2011Q2 sample is challenging for the models under consideration, as they generally seem not to be able to improve upon point and density predictions from the UCSV model. Our BMA-based approach dominates when we conduct the forecast evaluation for the 1985Q1–2011Q2 sample, on which we report in the lower panel of Table 7. When trading off point forecasts (as summarized by the RMSE and MAE ratios) *versus* density forecasts (summarized by the avCRPS ratios) our BMA-SBV model performs best relative to the UCSV model at $h = 1$, whereas our BMA-SBB-SBV specification most clearly dominates UCSV-based point and density inflation forecasts at $h = 5$.

In sum, the results in Tables 6 and 7 make it clear that the early 1980s appears to be challenging in terms of inflation point forecast for both BMA-based and non-BMA-based competitors of the UCSV model, but much less so in terms of PCE deflator inflation density forecasts. When the focus is on the post-1984 sample, which still includes major events, our BMA family of inflation models does quite well for predicting different inflation series at both forecast horizons. For PCE deflator inflation the BMA-based specifications that model time-variation in both the regression parameters and error variance, BMA-SBB-SBV and BMA-RWB-RWV, generally seem to have the best forecasting performance. In case of GDP deflator inflation, BMA with only stochastic structural breaks in the error variance (BMA-SBV) is most dominant relative to the UCSV specification at $h = 1$, whereas the BMA-SBB-SBV model does that for one-year ahead predictions. Finally, the Tables 6 and 7 results seem to suggest that BMA in conjunction with some form break modeling is particularly useful when the forecast density is evaluated. This warrants a more detailed density forecast evaluation for both inflation rates, and the results of which are summarized in Tables 8 and 9.

In Table 8 we report on density forecast evaluation of our BMA range of models for PCE deflator inflation relative to UCSV-based density forecasts when different parts of

Table 8: Detailed Density Forecast Evaluation PCE deflator inflation

	Horizon: $h=1$			Horizon: $h=5$		
	avQS-C	avQS-R	avQS-L	avQS-C	avQS-R	avQS-L
	Forecast evaluation sample: 1980Q1-2011Q2					
BMA-SBB-SBV	**0.90***	**0.94**	**0.80****	**0.97**	1.11	**0.74****
	(-1.57)	(-0.76)	(-2.15)	(-0.28)	(0.61)	(-1.99)
BMA-SBV	**0.88****	**0.80*****	**0.81*****	**0.97**	**0.97**	**0.74***
	(-2.26)	(-2.87)	(-3.44)	(-0.25)	(-0.19)	(-1.57)
BMA-SBB	**0.91****	**0.84****	**0.76*****	1.07	1.15	**0.80***
	(-1.86)	(-2.14)	(-3.57)	(0.56)	(0.80)	(-1.47)
BMA	**0.95**	**0.84****	**0.82*****	1.18	1.13	**0.94**
	(-0.96)	(-2.23)	(-3.09)	(1.05)	(0.63)	(-0.46)
BMA-RWB-RWV	**0.88*****	**0.84****	**0.75*****	**0.91**	**0.91**	**0.71*****
	(-3.21)	(-2.32)	(-3.71)	(-1.14)	(-0.58)	(-2.83)
	Forecast evaluation sample: 1985Q1-2011Q2					
BMA-SBB-SBV	**0.91***	**0.91**	**0.83***	**0.88****	**0.88**	**0.79***
	(-1.32)	(-1.02)	(-1.47)	(-1.76)	(-0.94)	(-1.57)
BMA-SBV	**0.92***	**0.81*****	**0.85*****	**0.92**	**0.80***	**0.83**
	(-1.42)	(-2.39)	(-2.49)	(-0.59)	(-1.38)	(-1.18)
BMA-SBB	**0.93***	**0.84****	**0.81*****	**0.97**	**0.94**	**0.86**
	(-1.45)	(-1.96)	(-2.63)	(-0.30)	(-0.40)	(-1.08)
BMA	**0.96**	**0.84****	**0.86*****	1.03	**0.95**	**0.84**
	(-0.67)	(-1.88)	(-2.44)	(0.18)	(-0.25)	(-1.02)
BMA-RWB-RWV	**0.89*****	**0.83****	**0.79*****	**0.86*****	**0.76****	**0.75*****
	(-2.83)	(-2.18)	(-2.79)	(-2.73)	(-1.79)	(-2.61)

Note: See the notes for Table 6, but now for the avQS-C, avQS-R and avQS-L measures in (25).

Table 9: Detailed Density Forecast Evaluation GDP deflator inflation

	Horizon: $h=1$			Horizon: $h=5$		
	avQS-C	avQS-R	avQS-L	avQS-C	avQS-R	avQS-L
	Forecast evaluation sample: 1980Q1-2011Q2					
BMA-SBB-SBV	**0.95**	1.18	**0.83*****	**0.96**	1.25	**0.70****
	(-1.13)	(2.59)	(-2.46)	(-0.47)	(1.45)	(-2.23)
BMA-SBV	**0.87*****	**0.96**	**0.70*****	**0.93**	1.09	**0.67****
	(-2.51)	(-0.64)	(-4.85)	(-0.51)	(0.54)	(-2.20)
BMA-SBB	**0.95**	1.06	**0.74*****	1.20	1.43	**0.86**
	(-0.62)	(-0.21)	(-1.80)	(1.39)	(2.37)	(-1.09)
BMA	**0.99**	1.03	**0.79*****	1.33	1.40	1.02
	(-0.05)	(0.30)	(-2.47)	(1.33)	(1.41)	(0.11)
BMA-RWB-RWV	1.09	1.33	**0.87****	1.01	1.29	**0.70****
	(1.54)	(3.25)	(-1.74)	(0.12)	(1.55)	(-1.69)
	Forecast evaluation sample: 1985Q1-2011Q2					
BMA-SBB-SBV	**0.93**	1.14	**0.83****	**0.86****	**0.99**	**0.71****
	(-1.26)	(1.68)	(-2.10)	(-1.75)	(-0.00)	(-1.83)
BMA-SBV	**0.87****	**0.95**	**0.71*****	**0.88**	**0.95**	**0.70****
	(-2.01)	(-0.55)	(-3.89)	(-0.68)	(-0.29)	(-1.75)
BMA-SBB	**0.96**	1.07	**0.76*****	1.01	1.19	**0.83**
	(-0.69)	(0.83)	(-2.91)	(0.10)	(1.37)	(-0.99)
BMA	**0.95**	1.01	**0.74*****	1.06	1.14	**0.77**
	(-0.72)	(0.11)	(-3.10)	(0.30)	(0.48)	(-1.21)
BMA-RWB-RWV	1.13	1.36	**0.93**	1.02	1.20	**0.80**
	(1.81)	(2.86)	(-0.94)	(0.20)	(0.94)	(-1.14)

Note: See the notes for Table 6, but now for the avQS-C, avQS-R and avQS-L measures in (25).

the underlying predictive densities are emphasized, i.e., the center, the right (upper) tail, and the left (lower) tail. It is clear that for the 1980Q1-2011Q2 sample in the upper panel of Table 8, the BMA family of inflation models provide more precise forecasts for both tails of the current quarter inflation density, with in particular the BMA-SBV and BMA-RWB-RWV models also being able to outperform the UCSV for the center of this density. The improvement over UCSV density forecasts at $h = 5$ for this evaluation sample is limited to the left (lower) tail for BMA models that incorporate some form of structural instability. For the 1985Q1-2011Q2 sample, reported in the lower panel of Table 8, we observe similar results at $h = 1$ relative to the 1980Q1-2011Q2 sample. At the one-year ahead horizon, however, the results are now a bit more mixed: the BMA-SBB-SBV specification outperforms the UCSV model when the center and the lower tail of the forecast densities are emphasized, BMA-SBV based density forecasts are more precise than UCSV ones for the upper tail, and the BMA-RWB-RWV model performs better across the whole density.

Results of a similar detailed density forecast analysis for GDP deflator inflation can be found in Table 9. As in the case of PCE deflator inflation, our BMA-based inflation models provide significantly more precise predictions for the lower tail in terms of the $h = 1$ inflation forecast density than the UCSV model, for both the 1980Q1-2011Q2 and the 1985Q1-2011Q evaluation samples. Across both these evaluation periods only the BMA-SBB-SBV and the BMA-SBV members of our BMA-range of inflation models are able to outperform the UCSV model at the one-year ahead horizon ($h = 5$) when we consider the lower tail of the respective forecast densities for GDP deflator inflation.

So what can we conclude from the results reported in Tables 6-9? Conditioning BMA across a large number of potential inflation predictors on a form of break modeling that at the very least incorporates structural variance breaks can result in very good inflation point forecasts, especially for the post-1984 period. Moreover, the biggest pay-off of such an approach occurs when one wants to predict the likelihood of excessive inflation movements, particularly one-year ahead when emphasizing the likelihood of lower-than-usual inflation rates.

The aforementioned results are reminiscent of the discussion in Stock and Watson (2008) who link the predictive content of economic activity variables for inflation to extremes in economic activity over the business cycle. In fact, Stock and Watson (2010) make this explicit and propose a forecasting model that relates inflation dynamics to recessional downturns in unemployment and they thus effectively use a nonlinear inflation-slack relationship. The time-variation in the predictive content of activity variables for inflation as implied by specifications like BMA-SBB-SBV and BMA-SBV therefore seem to pick up similar nonlinear inflation-slack tradeoffs. The main difference is that specifications like BMA-SBB-SBV and BMA-SBV incorporate model uncertainty, which possibly makes these more robust when economic downturns in different periods imply different drivers for inflation dynamics.

5 Conclusion

In this paper we have proposed an approach that entails Bayesian model averaging of regression-based models with possible structural breaks to forecast U.S. inflation using real time data. Bayesian model averaging is accomplished with a variable selection approach where we use about 15 potential predictors as well as lagged inflation rates to predict quarterly PCE and GDP deflator inflation series. Within the resulting model average we consider flexible discrete break specifications as well as a random walk specification to model different channels of structural instability, by allowing time-variation in the regression parameters, the error variance of the overall model average, or both.

Posterior inference on the full 1960Q1 − 2011Q2 sample shows that there is substantial uncertainty in the specification of the individual regressions across our different Bayesian model averaging approaches. We find that combining Bayesian model averaging with a form of structural breaks in the regression parameters and/or the innovation variance results in a much smaller number of predictor variables in the individual regression specifications of the model average than when we do not incorporate these structural breaks. In addition, Bayesian model averaging with structural breaks in both the regression parameters and the innovation variance seems to adapt the parsimony of the individual regression specifications to the forecast horizon, with more parsimony when one wants to forecast further out. Our different Bayesian model averaging with breaks specifications captures well the time-varying properties of inflation reported in previous studies.

The real time inflation forecasting performance of our different Bayesian model averaging specifications are compared with several rival approaches including simple autoregressive specifications, the random walk specification and ridge regression. The well-known Stock and Watson (2007, 2008) UCSV model is taken as the benchmark for the real-time forecasting exercise. We evaluate the relative forecasting performance across these different approaches in terms of both point and density forecasts for quarterly (GDP or PCE deflator) inflation rates in the current quarter as well as the quarter one year beyond the current quarter. This comparison shows that Bayesian model averaging over a large number of predictors in a model that at least allows for structural breaks in the innovation variance results in very accurate inflation point and density forecasts, especially for the post-1984 period. Bayesian model averaging with breaks in the regression parameters and error variance performs especially well when predicting, in real-time, the possibility of excessive negative inflation movements over the medium term.

References

Amisano, G. and R. Giacomini, 2007, Comparing Density Forecasts via Weighted Likelihood Ratio Tests, *Journal of Business & Economic Statistics* **25**, 177–189.

Ang, A., G. Bekaert and M. Wei, 2007, Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?, *Journal of Monetary Economics* **54**, 1163–1212.

Ang, A., M. Piazzesi and M. Wei, 2006, What Does the Yield Curve Tell us about GDP Growth?, *Journal of Econometrics* **131**, 359–403.

Atkeson, A. and L. E. Ohanian, 2001, Are Phillips Curves Useful for Forecasting Inflation?, *Federal Reserve Bank of Minneapolis Quarterly Review* **25**, 2–11.

Benati, L., 2004, Evolving Post-World War II U.K. Economic Performance, *Journal of Money, Credit, and Banking* **36**, 691–718.

Carter, C. and R. Kohn, 1994, On Gibbs Sampling for State-Space Models, *Biometrika* **81**, 541–553.

Carter, C. and R. Kohn, 1997, Semiparametric Bayesian Inference for Time Series with Mixed Spectra, *Journal of the Royal Statistical Society, Series B* **B**, 255–268.

Clark, T. E., 2011, Real-Time Density Forecasts from VARs with Stochastic Volatility, *Journal of Business & Economic Statistics* . Forthcoming.

Clark, T. E. and M. W. McCracken, 2012, Advances in Forecast Evaluation, *in* A. Timmermann and G. Elliott (editors), *Handbook of Economic Forecasting*, Elsevier, Amsterdam.

Clark, T. E. and T. Davig, 2011, Decomposing the Declining Volatility of Long-Term Inflation Expectations, *Journal of Economic Dynamics and Control* **35**, 981–999.

Cogley, T. and A. M. Sbordone, 2008, Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve, *American Economic Review* **98**, 2101–2126.

Cogley, T., G. E. Primiceri and T. J. Sargent, 2010, Inflation-Gap Persistence in the U.S., *American Economic Journal: Macroeconomics* **2**, 43–69.

Cogley, T. and T. J. Sargent, 2002, Evolving Post World War II U.S. Inflation Dynamics, *in* B. S. Bernanke and K. Rogoff (editors), *NBER Macroeconomics Annual 2001*, MIT Press, Cambridge, U.S.

Cogley, T. and T. J. Sargent, 2005, Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S., *Review of Economic Dynamics* **8**, 262–302.

De Mol, C., D. Giannone and L. Reichlin, 2008, Forecasting with a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?, *Journal of Econometrics* **146**(5829), 318–328.

Diebold, F. X., G. D. Rudebusch and B. Aruoba, 2006, The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach, *Journal of Econometrics* **131**, 309–338.

Diebold, F.X. and R.S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* **13**, 253–263.

Elliott, G., T. J. Rothenberg and J. H. Stock, 1996, Efficient Tests for an Autoregressive Unit Root, *Econometrica* **64**, 813–836.

Fama, E. F. and R. R. Bliss, 1987, The Information in Long-Maturity Forward Rates, *American Economic Review* **77**, 680–692.

Faust, J. and J. H. Wright, 2011, Forecasting Inflation, *Handbook of Forecasting*, Forthcoming.

Geman, S. and D. Geman, 1984, Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**, 721–741.

George, E. I. and R. E. McCulloch, 1993, Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* **88**, 881–889.

Gerlach, R., C. Carter and R. Kohn, 2000, Efficient Bayesian Inference for Dynamic Mixture Models, *Journal of the American Statistical Association* **95**, 819–828.

Geweke, J. F., 1992, Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, *in* J. Berger, J. Bernardo, A. Dawid and A. Smith (editors), *Bayesian Statistics*, Oxford University Press, Oxford.

Geweke, J. F., 2005, *Contemporary Bayesian Econometrics and Statistics*, Wiley, Chichester.

Giordani, P. and M. Villani, 2010, Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction, *International Journal of Forecasting* **26**, 312–325.

Giordani, P. and R. Kohn, 2008, Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models, *Journal of Business and Economic Statistics* **26**, 66–77.

Giordani, P., R. Kohn and D. van Dijk, 2007, A Unified Approach to Nonlinearity, Outliers and Structural Breaks, *Journal of Econometrics* **137**, 112–137.

Gneiting, T., 2011, Making and Evaluating Point Forecasts, *Journal of the American Statistical Association* **106**, 746–762.

Gneiting, T. and A.E. Raftery, 2007, Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association* **102**, 359–378.

Gneiting, T. and R. Ranjan, 2011, Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules, *Journal of Business & Economic Statistics* **29**, 411–422.

Groen, J. J. J. and H. Mumtaz, 2008, Investigating the Structural Stability of the Phillips Curve Relationship, *Working Paper 350*, Bank of England.

Haldane, A. and D. Quah, 1999, UK Phillips Curves and Monetary Policy, *Journal of Monetary Economics* **44**, 259–278.

Harvey, D., S. Leybourne and P. Newbold, 1997, Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting* **13**, 281–291.

Jacquier, E., N. G. Polson and P. E. Rossi, 1994, Bayesian Analysis of Stochastic Volatility Models, *Journal of Business and Economic Statistics* **12**, 371–418.

Jore, A.S., J. Mitchell and S.P. Vahey, 2010, Combining Forecast Densities from VARs with Uncertain Instabilities, *Journal of Applied Econometrics* **25**, 621–634.

Justiniano, A. and G. E. Primiceri, 2008, The Time-Varying Volatility of Macroeconomic Fluctuations, *American Economic Review* **98**, 604–641.

Kim, S., S. Shephard and S. Chib, 1998, Stochastic Volatility: Likelihood Inference and Comparison with ARCH models, *Review of Economic Studies* **65**, 361–393.

Koop, G. and S. Potter, 2007, Estimation and forecasting in models with multiple breaks, *Review of Economic Studies* pp. 763–789.

Kuo, L. and B. Mallick, 1998, Variable Selection for Regression Models, *The Indian Journal of Statistics* **60**(Series B), 65–81.

Levin, A. T. and J. M. Piger, 2004, Is Inflation Persistence Intrinsic in Industrial Economies?, *Working Paper 334*, European Central Bank.

Marcellino, M., J. H. Stock and M. W. Watson, 2006, A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics* **127**(1-2), 499–526.

Newey, W. and K. D. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703–708.

Orphanides, A. and S. van Norden, 2005, The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time, *Journal of Money, Credit and Banking* **37**, 583–601.

Pesaran, M.H., D. Pettenuzzo and A. Timmermann, 2006, Forecasting Time Series Subject to Multiple Structural Breaks, *Review of Economic Studies* **73(4)**, 1057–1084.

Primiceri, G., 2005, Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies* **72**, 821–852.

Sensier, M. and D. J. C. van Dijk, 2004, Testing for Volatility Changes in U.S. Macroeconomic Time Series, *Review of Economics and Statistics* **86**, 833–839.

Shephard, N.E., 1994, Partial non-Gaussian state-space models, *Biometrika* **81**, 115–131.

Sims, C. A. and T. Zha, 2006, Were there Regime Switches in U.S. Monetary Policy?, *American Economic Review* **96**, 54–81.

Stock, J. H. and M. W. Watson, 2010, Modeling Inflation After the Crisis, *Macroeconomic Policy: Post-Crisis and Risks Ahead*, FRB Kansas City Symposium, Jackson Hole, Wyoming.

Stock, J.H. and M.W. Watson, 1999, Forecasting Inflation, *Journal of Monetary Economics* **44**, 293–335.

Stock, J.H. and M.W. Watson, 2007, Has Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking* **39**, 3–33.

Stock, J.H. and M.W. Watson, 2008, Phillips Curve Inflation Forecasts, *Working Paper 14322*, National Bureau of Economic Research.

Tanner, M.A. and W.H. Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* **82**, 528–550.

Wright, J. H., 2009, Forecasting U.S. Inflation by Bayesian Model Averaging, *Journal of Forecasting* **28**, 131–144.

Appendix

A Gibbs Sampler

In this section we derive the full conditional posterior distributions of the latent variables and the model parameters as discussed in Section 2.3.

Step 1: Sampling the variable selection parameters in D

We follow Kuo and Mallick (1998), which is a simplified version of the George and McCulloch (1993) algorithm. Starting from the previous iteration, the variable D is drawn from its full conditional posterior distribution. We compute the value of the posterior density (10) for $\delta_j = 0$ and $\delta_j = 1$ given the value of the other parameters which results in p_{j0} and p_{j1}, respectively. The full conditional posterior simplifies to

$$\begin{aligned}&\Pr[\delta_j = 1|D_{-j}, \pi, q^2, B, S, K, y, x] = \\ &\frac{\lambda_j \prod_{t=1}^{T-h} p(y_{t+h}|D_{-j}, x_t, \beta_t, \sigma_t^2)|_{\delta_j=1}}{(1-\lambda_j)\prod_{t=1}^{T-h} p(y_{t+h}|D_{-j}, x_t, \beta_t, \sigma_t^2)|_{\delta_j=0} + \lambda_j \prod_{t=1}^{T-h} p(y_{t+h}|D_{-j}, x_t, \beta_t, \sigma_t^2)|_{\delta_j=1}},\end{aligned} \tag{A.1}$$

for $j = 1, \ldots, k$, where $D_{-j} = (\delta_1, \ldots, \delta_{j-1}, \delta_{j+1}, \ldots, \delta_k)'$ and where the density of y_{t+h} is given in (9). We randomly choose the order in which we sample the k δ_j parameters. As starting value of the Gibbs sampler we consider a model which includes all k x_t variables.

Step 2: Sampling K_β

The structural breaks in the regression parameters B, measured by the latent variable κ_{jt}, are drawn using the algorithm of Gerlach *et al.* (2000, Section 3), which derives its efficiency from generating κ_{jt} without conditioning on the states β_{jt}. The conditional posterior density for κ_{jt}, $t = 1, \ldots, T$, $j = 0, \ldots, k$ unconditional on B is

$$\begin{aligned}&p(\kappa_{0t}, \ldots, \kappa_{kt}|K_{\beta,-t}, K_\sigma, S, \theta, y, x) \\ &\quad\propto p(y|K, S, \theta, x)p(\kappa_{0t}, \ldots, \kappa_{kt}|K_{\beta,-t}, K_\sigma, S, \theta, x) \\ &\quad\propto p(y_{t+h+1}, \ldots, y_{T-h}|y_{h+1}, \ldots, y_{t+h}, K, S, \theta, x) \\ &\qquad\quad p(y_{t+h}|y_{h+1}, \ldots, y_{t+h-1}, \kappa_1, \ldots, \kappa_t, K_\sigma, S, \theta, x)p(\kappa_{0t}, \ldots, \kappa_{kt}|K_{\beta,-t}, K_\sigma, S, \theta, x),\end{aligned} \tag{A.2}$$

where $K_{\beta,-t} = \{\{\kappa_{js}\}_{j=0}^{k}\}_{s=1,s\neq t}^{T-h}$. The density $p(\kappa_{0t}, \ldots, \kappa_{kt}|K_{\beta,-t}, K_\sigma, S, \theta, x)$ is equal to $\prod_{j=0}^{k} \pi_j^{\kappa_{jt}}(1-\pi_j)^{1-\kappa_{jt}}$ since κ_{jt} does not depend on δ_j. The two remaining densities $p(y_{t+h+1}, \ldots, y_{T-h}|y_{h+1}, \ldots, y_{t+h}, K, S, \theta, x)$ and $p(y_{t+h}|y_{h+1}, \ldots, y_{t+h-1}, \kappa_1, \ldots, \kappa_t, K_\sigma, S, \theta, x)$ can easily be evaluated as shown in Gerlach *et al.* (2000, Section 3). Because κ_t can take a finite number of values, the integrating constant can easily be computed by normalization. When $\delta_j = 0$, κ_{jt} for $t = 1, \ldots, T-h$ is sampled using (4).

Step 3: Sampling the regression parameters in B

The full conditional posterior density for the latent regression parameters B is computed using a simulation smoother. We follow Carter and Kohn (1994). The Kalman smoother is applied to derive the conditional mean and variance of the latent factors. For the initial values of $\beta_0, \ldots, \beta_k$ we use a multivariate normal prior with the mean equal to the corresponding OLS parameter estimate and a covariance matrix equal to the diagonal matrix of the covariance matrix of the OLS parameter estimates. These OLS parameter estimates result from estimating a model that includes all potential predictor variables, and is re-estimated in real-time when forecasting out-of-sample. When we have $\delta_j = 0$, β_{jt} is recursively simulated according to (9) conditional on the values of κ_{jt} and the variance q_j^2.

Steps 4 and 5: Sampling the variance parameters K_σ and S

To draw K_σ and S we want to follow a similar approach as above. As the model for $\ln \sigma_t^2$ does not result in a linear state space model the Kalman filter cannot be applied. Therefore, we apply the approach of Giordani and Kohn (2008) and rewrite the model (2)–(3) as

$$\begin{aligned} \ln(y_{t+h} - \beta_{0t} - \sum_{j=1}^{k} \delta_j \beta_{jt} x_{jt})^2 &= \ln \sigma_t^2 + u_t \\ \ln \sigma_t^2 &= \ln \sigma_{t-1}^2 + \kappa_{k+1,t} \eta_{k+1,t}, \end{aligned} \tag{A.3}$$

where $u_t = \ln \varepsilon_t^2$ has a log χ^2 distribution with 1 degree of freedom. We follow Carter and Kohn (1994, 1997), Shephard (1994) and Kim *et al.* (1998) and approximate the $\ln \chi^2(1)$ distribution by a finite mixture of normal distributions. We consider a mixture of five normal distributions such that the density of u_t is given by

$$f(u_t) = \sum_{s=1}^{5} \varphi_s \frac{1}{\omega_s \sqrt{2\pi}} \exp\Big(- \frac{(u_t - \mu_s)^2}{2\omega_s} \Big) \tag{A.4}$$

with $\sum_{s=1}^{5} \varphi_s = 1$. The appropriate values for μ_s, ω_s^2 and φ_s can be found in Carter and Kohn (1997, Table 1). In each step of the Gibbs sampler we simulate for each observation t a component of the mixture distribution from the distribution of the mixing distribution. Given the value of the mixture component we can apply standard Kalman filter techniques. Hence, the variables K_σ and S can be sampled in a similar way as K_β and B in step 2 and 3. [2] For $\ln \sigma_0^2$ we take a normal prior with mean -1 and variance 0.1.

[2]If we consider the case were $\kappa_{jt} = 1$ for $j = 0, \ldots, k$ and all t, we use the Metropolis-within-Gibbs MCMC algorithm as in Cogley and Sargent (2005), which combines Gibbs sampling steps for model co-

Step 6: Sampling $\boldsymbol{\pi}$

The full conditional posterior density of π is given by

$$p(\pi|D, q^2, B, S, K, y, x) \propto \prod_{j=1}^{k+1} \pi_j^{a_j-1}(1-p_j)^{b_j-1} \prod_{t=1}^{T-h} \pi_j^{\kappa_{jt}}(1-\pi_j)^{(1-\kappa_{jt})} \tag{A.5}$$

and hence the individual π_j parameter can be sampled from Beta distributions with parameters $a_j + \sum_{t=1}^{T-h} \kappa_{jt}$ and $b_j + \sum_{t=1}^{T-h}(1-\kappa_{jt})$ for $j = 0, \ldots, k+1$.

Step 6: Sampling of $\boldsymbol{q^2}$

The full conditional posterior density of q_j^2 is given by

$$p(q_j^2|D, \pi, B, S, K, y, x) \propto q_j^{-\nu_j} \exp(-\frac{\omega_j}{2q_j^2}) \prod_{t=1}^{T-h} \left(\frac{1}{q_j} \exp(\frac{-(\beta_{jt} - \beta_{j,t-1})^2}{2q_j^2}) \right)^{\kappa_{jt}} \tag{A.6}$$

for $j = 0, \ldots, k$ and

$$p(q_{k+1}^2|D, \pi, B, S, K, y, x) \propto q_{k+1}^{-\nu_{k+1}} \exp(-\frac{\omega_{k+1}}{2q_{k+1}^2}) \prod_{t=1}^{T-h} \left(\frac{1}{q_{k+1}} \exp(\frac{-(\ln\sigma_t^2 - \ln\sigma_{t-1}^2)^2}{2q_{k+1}^2}) \right)^{\kappa_{k+1,t}} \tag{A.7}$$

and hence q_j^2 can be sampled from an inverted Gamma-2 distribution with the scale parameter set equal to $\omega_j + \sum_{t=1}^{T-h} \kappa_{jt}(\beta_{jt} - \beta_{j,t-1})^2$ for $j = 0, \ldots, k$ or $\omega_{k+1} + \sum_{t=1}^{T-h} \kappa_{k+1,t}(\ln\sigma_t^2 - \ln\sigma_{t-1}^2)^2$ and degrees of freedom equal to $\nu_j + \sum_{t=1}^{T-h} \kappa_{jt}$ for $j = 0, \ldots, k+1$.

B Data Sources and Construction

Inflation rates

Our two dependent variables are inflation rates based on the gross domestic product (GDP) deflator as well as the personal consumption expenditures (PCE) deflator. Both measures get revised on a regular basis and we therefore do not retrieve our data from the usual data sources. Instead, we get the original vintages of the underlying data from the 'Real-Time Data Set for Macroeconomists' (RTDSM) at the Federal Reserve Bank of Philadelphia (http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data). The RTDSM proxies the original vintages for each quarter by selecting the data that was originally available around the middle of that quarter (as close as possible to the 15th day of the

efficients with the Metropolis algorithm as in Jacquier *et al.* (1994). Giordani and Kohn (2008) use a Metropolis-within-Gibbs MCMC algorithm where K_β and K_σ are sampled by an adaptive Metropolis algorithm using a proper candidate. Our experiments did not show substantial increase in computing time for the Cogley and Sargent (2005) approach, so we choose to work with the exact sampling version.

middle month of a quarter). Vintages of inflation rates are constructed as the percentage quarterly changes of the respective deflator series.[3]

Explanatory variables

We use in this paper an extensive set of activity and expectations measures to model inflation dynamics. Like the aforementioned inflation rates, the bulk of these variables gets revised so we strive to use as much as possible the original vintages of underlying data. Some of the measures can be directly retrieved from the respective real-time databases, others need to be constructed. When necessary we transform the variables to render them $I(0)$ by computing the percentage quarterly change of a series. The decision to transformation is based on the following: for each variable we randomly take 30% of the available vintages used in the forecasting analysis and apply the Elliott *et al.* (1996) unit root test on the (log of) the level for each of these selected vintages, and we transform the variable if in case of half or more of selected the vintages we cannot reject the null of non-stationarity.

Real output growth - ROUT We take the original quarterly data vintages for GDP in volume terms from the RTDSM at the Federal Reserve Bank of Philadelphia. Based on these we construct real output growth rates, i.e., the percentage quarterly change in real GDP.

Real durable consumption growth - RCON We take the original quarterly data vintages for real durable personal consumption expenditures (PCE) from the RTDSM at the Federal Reserve Bank of Philadelphia. Based on these we construct real durable consumption growth rates, i.e., percentage quarterly change in real durable PCE.

Real residential investment growth - RINV We take the original quarterly data vintages for real residential investment from the RTDSM at the Federal Reserve Bank of Philadelphia. Based on these we construct real residential investment growth rates, i.e., percentage quarterly change in the real residential investment level.

Import price inflation - PIMP We take the original quarterly data vintages for the imports deflator from the RTDSM at the Federal Reserve Bank of Philadelphia. Based on these we construct import price inflation, i.e., percentage quarterly change in the imports deflator.

Non-farm payrolls growth rate - NFPR From the RTDSM database, we take as quarterly vintages those monthly data vintages of non-farm payrolls employment that are closest to the middle of quarter (where the quarterly vintage is the vintage of monthly

[3] We define percentage quarterly change as 100 times the quarterly change of the logarithm of the original series.

data for the second month of the quarter). We transform these data to the quarterly frequency through averaging; finally, the non-farm payrolls growth rate is constructed as the percentage quarterly change in non-farm payrolls.

Housing starts - HSTS We take the original quarterly data vintages of monthly housing starts (where the quarterly vintage is the vintage of monthly data for the second month of the quarter) from the RTDSM at the Federal Reserve Bank of Philadelphia. We transform these data to the quarterly frequency through averaging; finally, we take the natural logarithm of the 'raw' quarterly average of housing starts.

M2 growth rate - M2 The RTDSM database contains quarterly data vintages of monthly M2 data (where the quarterly vintage is the vintage of monthly data for the second month of the quarter). However, for the 1981Q1 and 1981Q2 vintages these data are incomplete and we replace these vintages with M2 vintage for the first months of these quarters, which we retrieve from the ALFRED® real-time database at the Federal Reserve Bank of St. Louis. We transform these data to the quarterly frequency through averaging; finally, the M2 growth rate is constructed as the percentage quarterly change in the M2 level.

Unemployment ratio - UNEMPL We take the original quarterly data vintages for unemployment as a percentage of the labor force (UNEMP) from the RTDSM at the Federal Reserve Bank of Philadelphia.

Level term structure factor - YL This is a proxy for the level factor describing the dynamics in the term structure of interest rates. The term structure is approximated by seven interest rates: the 3-month Treasury bill rate, the 6-month Treasury bill rate, both are secondary market rates retrieved from the Haver Analytics database, as well as the Fama and Bliss (1987) 1-year, 2-year, 3-year, 4-year and 5-year zero-coupon bond yields from the CRSP database at Wharton Research Data Services. These are monthly data, which are not revised as they are financial data. In order to get quarterly data we select the aforementioned interest rates at the end of the second month of a quarter. The level term structure factor equals the cross-sectional average across the above seven interest rates for each quarter.

Slope term structure factor - TS This is a proxy for the slope factor describing the dynamics in the term structure of interest rates. We use the same interest rates as for the level term structure factor - see above. These are monthly data; in order to get quarterly data we select the aforementioned interest rates at the end of the second month of a quarter. The slope term structure factor equals the spread between the 5-year zero-coupon bond yield and the 3-month T-bill rate for each quarter.

Curvature term structure factor - CS This is a proxy for the curvature factor describing the dynamics in the term structure of interest rates. We use the same interest rates as for the level term structure factor - see above. These are monthly data; in order to get quarterly data we select the aforementioned interest rates at the end of the second month of a quarter. The curvature term structure factor equals for each quarter the spread between two times the 2-year zero-coupon bond yield and the sum of the 3-month Treasury bill rate and the 5-year zero-coupon bond yield.

Real oil price inflation - OIL To construct real oil prices, we first retrieve nominal oil prices - for this we use the West Texas Intermediate oil spot price from the Haver Analytics data base. Quarterly observations result by selecting in each quarter the observed oil spot price closest to the middle of the quarter; as these data are market prices they are not prone to revisions. Quarterly data vintages of real oil prices are constructed by deflating the aforementioned oil spot price, which is unrevised, by either the GDP deflator or PCE deflator for that vintage, depending on which inflation rate one wants to model. Vintages of real oil price inflation are then equal to the percentage quarterly change in the constructed real oil price level.

Real food commodities inflation - FOOD Vintages of real food commodities inflation are constructed in a similar manner as those for real oil price inflation - see above. Only now the construction is based on the Commodities Research Bureau (CRB) Index of Foodstuffs commodity prices, which is based on the spot prices for butter, cocoa beans, corn, cottonseed oil, hogs, lard, steers, sugar and wheat. The CRB Foodstuffs price index is retrieved from the Haver Analytics data base.

Real raw industrial commodities inflation - RAW Vintages of real raw industrial commodities inflation are constructed in a similar manner as those for real oil price inflation - see above. Only now the construction is based on the CRB Index of Raw Industrials commodity prices, which is based on the spot prices for burlap, copper scrap, cotton, hides, lead scrap, print cloth, rosin, rubber, steel scrap tallow, tin, wool tops and zinc. The CRB Raw Industrials price index is retrieved from the Haver Analytics data base.

Reuters/University of Michigan Survey of Consumers' inflation expectations - MS The Reuters/University of Michigan Survey of Consumers asks members of the general public, amongst other, to give a quantitative assessment of expected inflation in a year's time. As this is a one-year ahead measure, we lag these series, which are never revised, with four-quarters as to make them properly real-time. The quarterly data are retrieved from http://www.sca.isr.umich.edu/main.php at the University of Michigan.

C Prior Sensitivity Analysis

We investigate in this appendix the properties of the MCMC algorithm outlined in Section 2.3 and detailed in Appendix A and discuss the influence of prior values on posterior results. We base our results on the following data generating process [DGP]

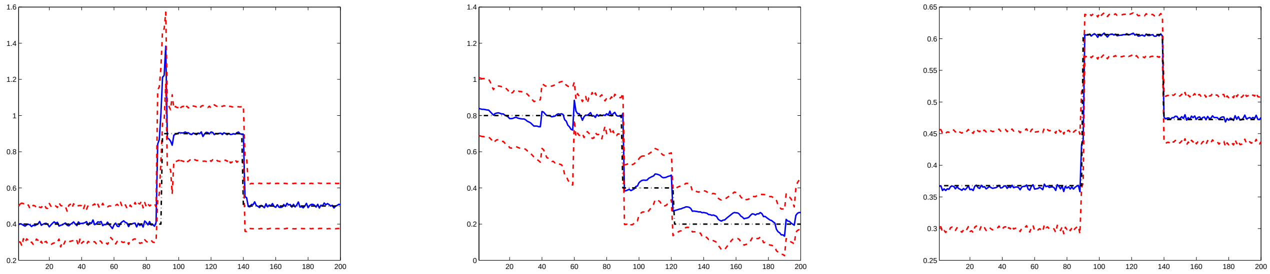
$$y_{t+1} = \beta_{0,t} + \beta_{1,t}x_{1t} + \beta_{2,t}x_{2t} + \beta_{3,t}x_{3t} + \sigma_t\varepsilon_{t+1}, \text{ for } t = 1, \ldots, 200 \tag{C.1}$$

with $\varepsilon_{t+1} \sim \text{NID}(0, 1)$ and $x_{j,t} \sim \text{NID}(0, 1)$ for $j = 1, \ldots, 3$. The $\beta_{3,t} = 0$ for $t = 1, \ldots, 200$ and hence $x_{3,t}$ is not included in the model. For the first regressor we take as parameters $\beta_{1,t} = 0.4$ for $t = 1, \ldots, 89$, $\beta_{1,t} = 0.9$ for $t = 90, \ldots, 139$ and $\beta_{1,t} = 0.5$ for $t = 140, \ldots, 200$. For the second regressor we have $\beta_{2,t} = 0.8$ for $t = 1, \ldots, 89$, $\beta_{2,t} = 0.4$ for $t = 90, \ldots, 119$ and $\beta_{2,t} = 0.2$ for $t = 140, \ldots, 200$. Furthermore, $\beta_{0,t} = 1/2$ for all t and $\ln \sigma_t^2 = -2$ for $t = 1, \ldots, 89$, $\ln \sigma_t^2 = -1$ for $t = 90, \ldots, 139$ and $\ln \sigma_t^2 = -1.5$ for $t = 140, \ldots, 200$. Hence, we allow for breaks in the parameters at different points in time but we also include breaks which occur at the same time.

We apply our Bayesian model averaging framework with structural breaks (2)–(3) with $h = 1$, 24000 posterior draws (with a burn-in of 4000 draws and a thinning of 2), and different prior settings to investigate the sensitivity of posterior results with respect to prior specification. We consider x_{1t}, x_{2t} and x_{3t} as potential regressors and allow for breaks in all parameters including the variance. Note that the intercept is always included. In the base case we take the prior parameter λ_j in (5) equal to 50% for $j = 1, \ldots, 3$. We set $a_0 = 0.50$ and $b_0 = 100$ in (6), and $\omega_0 = 0.85$ and $\nu_0 = 100$ in (7) for the intercept parameter. For the other regression parameters we choose $a_j = 0.5$, $b_1 = 100$, $b_2 = b_3 = 5$, $\omega_j = 0.75$ and $\nu_j = 50$ for $j = 1, ..., 3$, which implies a smaller expected size of breaks and, often, a lower break probability than for the intercept. The prior parameters concerning the variance equation are $a_4 = 0.8$, $b_4 = 5$, $\omega_4 = 0.2$ and $\nu_4 = 50$, respectively. The base case prior settings are such that the BMA-SBB-SBV approach provides the best fit for the simulated data. For example, the posterior inclusion probabilities for our base case prior settings are well in line with DGP (C.1) for the simulated data, i.e., they equal 0.988, 0.980, and 0.030 for x_{1t}, x_{2t} and x_{3t}, respectively. As a further illustration of this, we report in Figure C.1 posterior estimates of parameters β_{1t}, β_{2t} and σ_{1t} ($\beta_{3t} = 0$ always as it is basically never selected) together with the corresponding DGP parameters. The results from this figure show that our approach is quite accurate in estimating both the timing and the size of the breaks, where the estimate of β_{2t} is slightly more volatile due to our prior choice for b_2 that is lower than b_1.

In our prior sensitivity analysis, we will consider four alternative prior specifications where we decrease or increase the prior probability of a break and decrease or increase the prior expectation of the size of the break. These changes in the priors are applied to, respectively, the intercept, the regression parameters and the variance specification, which implies that we consider 12 different prior specifications in total. A lower probability of a break than in the base case means that we multiply b_j by 10 and, correspondingly, a

Figure C.1: Posterior estimates of the time-varying parameters implied by DGP (C.1): Base case prior settings



Notes: The solid lines represent the posterior medians of the β_{1t} parameter (first column), β_{2t} parameter (second column) and σ_t parameter (third column). The dashed lines denote the 25th and 75th percentiles of the posterior distributions. The dark solid line displays the values used to generate the data generating process in DGP (C.1).

higher probability means that we divide b_j by 10 for $j = 1, ...4$.[4] A higher expected prior break size than in the base case is obtained by multiplying ω_j and ν_j $\forall j$ by 5 and, thus, a lower expected prior break size is obtained by dividing these parameters by 5. Table C.1 summarizes the prior settings. Note, some of the prior settings are quite extreme but they serve to illustrate our prior sensitivity analysis.

Table C.1: Summary of the prior settings for the different cases

break	exp.	prior intercept				prior regressors				prior variance			
prob.	size	a_0	b_0	ω_0	ν_0	$a_{1:3}$	$b_1, b_{2:3}$	$\omega_{1:3}$	$\nu_{1:3}$	a_4	b_4	ω_4	ν_4
base	base	0.50	100	0.85	100	0.5	100,5	0.75	50	0.8	5	0.2	50
low	small	0.50	100	0.17	20	0.5	1000,50	0.15	10	0.8	50	0.04	10
low	large	0.50	100	4.25	500	0.5	1000,50	3.75	250	0.8	50	1	250
high	small	0.50	10	0.17	20	0.5	10,0.50	0.15	10	0.8	0.50	0.04	10
high	large	0.50	10	4.25	500	0.5	10,0.50	3.75	250	0.8	0.50	1	250

We first focus on the posterior inclusion probabilities. Table C.2 report these probabilities for the standard prior setting, which are shown in the first line of the table, together with the 12 different cases. As mentioned earlier, the posterior inclusion probabilities for x_{1t} and x_{2t} are close to 1 and for x_{3t} are close to zero which corresponds with our DGP. In columns 3–5 we consider situations where we only change the prior settings for the intercept parameters. We see that the posterior inclusion are hardly affected by these changes, even if the inclusion probabilities of x_{3t} is in some examples slightly larger than

[4] The prior parameters b_0 and b_1 are already large so in the case of a lower probability of a break in the base case we multiply these parameters by 1.5.

for the base case. In the final three columns of the table we display the results where we only change the prior settings of the variance parameters. Again, the posterior inclusion probabilities are not affected much by these prior changes, except maybe for the case where we consider both an increase in the prior break probability and the prior break size, as the inclusion probabilities for x_{1t} and x_{2t} decline to levels just below 0.90. Columns 6–8, finally, show that the posterior inclusion probabilities for x_{1t} and x_{2t} become substantially smaller when we assume a larger expected size of the break in x-variables than for the base case, especially when we also have a higher prior probability of a break.

Table C.2: Posterior variable inclusion probabilities for different prior specifications

		prior sens. intercept			prior sens. regressors			prior sens. variance		
break prob.	exp. size	x_{1t}	x_{2t}	x_{3t}	x_{1t}	x_{2t}	x_{3t}	x_{1t}	x_{2t}	x_{3t}
base	base	0.988	0.980	0.030	0.988	0.980	0.030	0.988	0.980	0.030
low	small	0.995	0.992	0.000	1.000	0.992	0.056	0.988	0.974	0.123
low	large	0.989	0.972	0.085	0.621	0.648	0.005	0.976	0.946	0.036
high	small	0.995	0.977	0.004	1.000	0.993	0.000	0.997	0.989	0.021
high	large	0.988	0.974	0.123	0.499	0.571	0.002	0.839	0.878	0.098

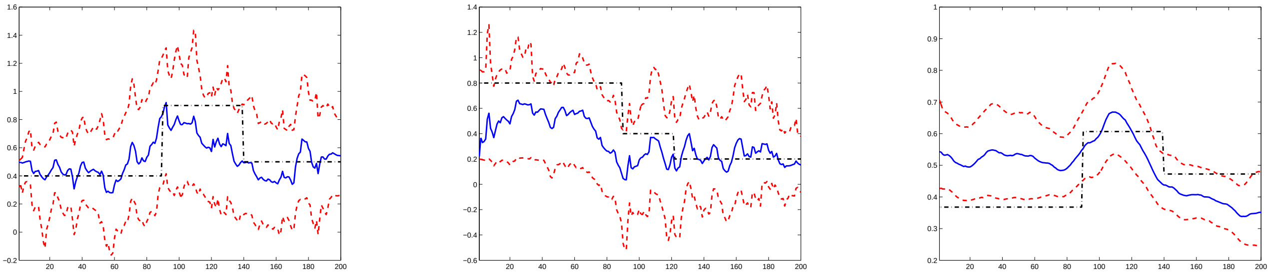
Posterior results for different prior break probability and expected prior size of a break, see Table C.1

From Table C.2 it is clear that when we increase the prior expected break size for the predictor variables, irrespective of the prior break probability, relative to the base case, the impact is the most substantial in terms of the posterior inclusion probabilities. This coincides with a substantial deterioration of the posterior medians of the parameters in (C.1) relative to the base case. For example, in case of a higher prior break probability and a higher expected prior break size for the predictor variables, Figure C.2 reports the posterior estimates of β_{1t}, β_{2t} and σ_t.[5] The figure makes it clear that for this case the posterior medians of the parameter are quite off in terms of the timing of the breaks, with a large uncertainty for the posterior estimates of β_{1t} and β_{2t} as well as a σ_t that is significantly lower towards the end of the sample than is implied by the DGP.

Another interesting case that emerges from the posterior inclusion results in Table C.2 is if we assume for the error variance σ_t^2 a higher prior probability of breaks of larger prior expected size than in the base case, as this leads to a slightly downward bias in the posterior inclusion probabilities. Again, as before, this seems to be a symptom of severely imprecise posterior parameter estimation results when the prior settings for, in this case, the error variance is changed in such a way, and Figure C.3 summarizes them. From

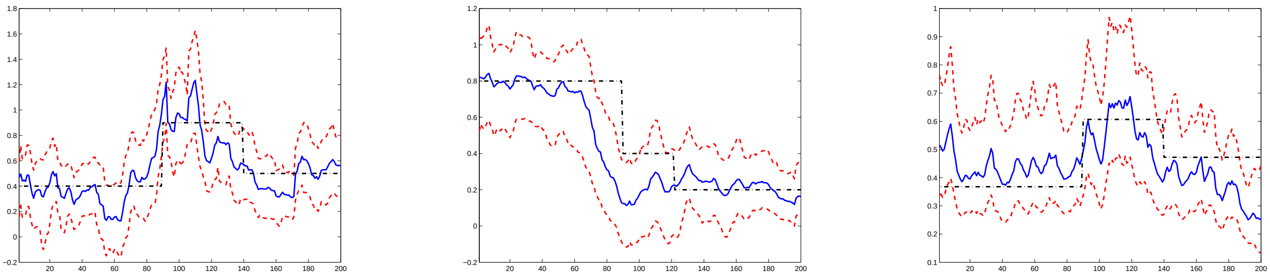
[5] The posterior estimates for the case of a lower prior break probability and a higher expected prior break size for the x variables are very similar.

Figure C.2: Posterior estimates of the time-varying parameters implied by DGP (C.1): Higher prior break probabilities and higher prior break sizes for the x-variables



Note: See Figure C.1, but now the posterior results are based on a higher prior break probability as well as a higher prior break size for the regressors x_{1t}, x_{2t}, and x_{3t} than in the base case.

Figure C.3: Posterior estimates of the time-varying parameters implied by DGP (C.1): Higher prior break probabilities and higher prior break sizes for σ_t^2

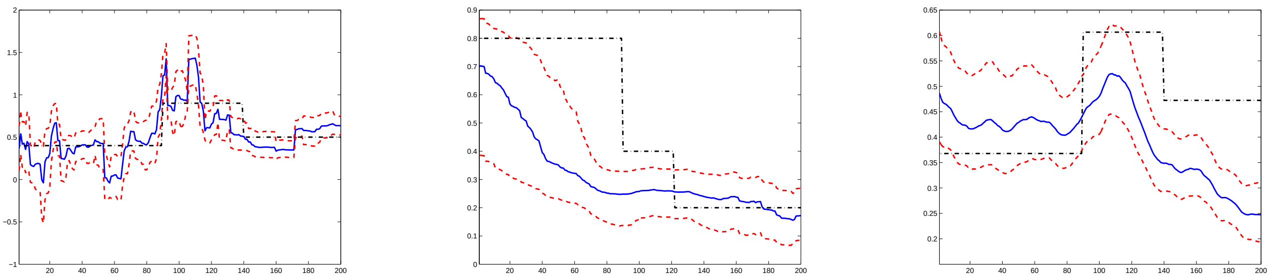


Note: See Figure C.1, but now the posterior results are based on a higher prior break probability as well as a higher prior break size for σ_t^2 than in the base case.

Figure C.3 it is clear that assuming *a priori* large sized breaks in σ_t^2 will result in a very uncertain posterior estimate of the error standard deviation and that the timing of breaks in both β_{1t} and β_{2t} are biased relative to those implied by DGP (C.1).

Altering the prior assumptions for the intercept also can lead to peculiar posterior estimation results. In Figure C.4 we report the posterior median and interquantile range for β_{1t} β_{2t} and σ_t when we impose a low prior break probability of breaks of larger prior expected size than in the base case. What becomes clear from this figure is that lowering the prior break probability of the intercept and attempting to compensate for that by increasing the corresponding prior expected break size, results in a process for β_{1t} that exhibits frequent smaller sized breaks than what we impose in the underlying DGP, which thus often result in false break signals for this parameter. For β_{2t} and σ_t not only the posterior estimates of the break dates are wrong but also the parameter levels themselves are often inconsistent with those from the DGP.

Figure C.4: Posterior estimates of the time-varying parameters implied by DGP (C.1): Lower prior break probabilities and higher prior break sizes for the intercept



Note: See Figure C.1, but now the posterior results are based on a lower prior break probability and a higher prior break size for the intercept than in the base case.

For the remaining prior sensitivity exercises,[6] the effect on the posterior parameter estimation results are less strong but nonetheless noticeable. Especially the estimated sizes of the breaks in the variance are substantially affected. In several of these alternative prior cases, the posterior medians of the variance parameters do not correspond to the true value after a break has occurred, even if the timing of the break is determined correctly. The regression parameters $\beta_{0t}, \ldots, \beta_{2t}$ seem to be less affected by the same prior cases, although we did notice much more uncertainty in the estimate of the timing of the breaks and there is often more posterior uncertainty in the estimated parameters. A general pattern we observe is that when the prior settings correspond to a higher probability of larger or smaller breaks than in the base case, the posterior medians of β_{1t} and β_{2t} are much more volatile over time than in Figure C.1, and a lower prior probability of larger or smaller breaks than in the base case increases the posterior uncertainty of these regression parameter estimates.

D MCMC Convergence Analysis

To analyze how well the MCMC sampler from Section 2.3 and Appendix A converges, we will report in this Appendix on the application of several MCMC convergence analysis approaches on this sampler for the full BMA-SBB-SBV specification estimated for both PCE and GDP deflator inflation rates at $h = 1$ and $h = 5$. More specifically, we followed the procedures utilized in, e.g., Primiceri (2005), Justiniano and Primiceri (2008) and Clark and Davig (2011). They consider computing inefficiency factors and t-tests for equality of the means across subsamples of the MCMC chain.

[6]We do not report them in order to preserve space, but they are available upon request from the authors.

For each individual parameter and latent variable, the simulation inefficiency factor is estimated as $(1 + \sum_{f=1}^{\infty} \rho_f)$, where ρ_f is the f-th order autocorrelation of the chain of draws. This inefficiency factor equals the variance of the mean of the posterior draws from the MCMC sampler, divided by the variance of the mean assuming independent draws. Then, if we require that the variance of the mean of the MCMC posterior draws should be limited to be at most 1% of the variation due to the data (measured by the posterior variance), the inefficiency factor provides an indication of the minimum number of MCMC draws to achieve this, see Kim *et al.* (1998). So, for example, an inefficiency factor of 20 for a parameter suggests that one needs in theory at least 2000 draws from the MCMC sampler for a reasonably accurate analysis of this parameter from the model. When estimating these inefficiency factors, we use the Bartlett kernel as in Newey and West (1987), with a bandwidth set to 4% of the sample of draws. Finally, we also compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed with the first 20 percent and last 40 percent of the sample of retained draws. For this particular convergence diagnostic test we compute the variances of the respective means using the Newey and West (1987) heteroskedasticity and autocorrelation robust variance estimator with a bandwidth set to 4% of the utilized sample sizes.

The two aforementioned sets of statistics were applied on a range of choices for the total number of posterior draws as well as burn-in period lengths and thinning for the BMA-SBB-SBV specification for both inflation rates and forecasting horizons. Based on this comparison we felt most comfortable that with the number of posterior draws set equal to 24000 with a burn-in period of 2000 draws and thinning value of 2, yielding 10000 retained posterior draws, our MCMC sampler would perform satisfactorily. Tables D.1 and D.2 provide a summary of, respectively, the corresponding inefficiency factors and Geweke (1992) diagnostic tests for this choice of the number of retained posterior draws for the BMA-SBB-SBV specification. These inefficiency factors and convergence diagnostic tests are computed for the full (1960Q1-2011Q2) sample estimates of the parameters and latent variables, as well as the real-time estimates of the predictive densities at $h = 1$ and $h = 5$ for each quarter in the 1980Q1-2011Q2 evaluation sample.

For most parameters and latent variables as well as the real-time estimated predictive densities, the inefficiency factors in Table D.1 suggest that our MCMC sampler is very efficient and that it requires far less than 10000 retained posterior draws to be able to do a reasonably accurate inferential analysis. In case of the time-varying parameters B at $h = 5$, with likely values in the 0.9-22 range, and, in particular, for the variable selection parameters D our sampler is less efficient.[7] Nonetheless, the corresponding inefficiency

[7]For both inflation rates at horizon $h = 5$, Figure 1 indicate that the real oil price and the two real

factors suggest a minimum number of draws of less than 4000 to achieve an accurate analysis of these parameters, less than our choice of 10000 retained draws. We nonetheless felt that accurate inference for the density forecast evaluation in Section 4.3 would be served better with our choice of 10000 retained draws. The convergence diagnostic tests in Table D.2 indeed confirm our conclusions regarding efficiency based on the results in Table D.1. For example, in the case of the D parameters the null hypothesis of equal means across subsamples of these 10000 retained draws is hardly ever rejected.

Thus, inference in our BMA framework appears to be reasonably accurate when we base posterior inference on 24000 draws with a burn-in of 4000 and thin value of 2. This also helped us to reduce computing time, as our forecasting exercise with an expanding data window and real-time data implied that we have to rerun our MCMC sampler many times. We use a similar choice for the posterior draws for most of the other variants of our BMA family of models, as unreported results of a similar convergence analysis as discussed above for these models reached similar conclusions. However, in case of the BMA-RWB-RWV specification, which assumes that κ_{jt} and $\kappa_{k+1,t}$ are always equal to 1 for all j, the inefficiency factors and convergence diagnostic tests pointed to much less efficient estimation when using only 24000 posterior draws. Hence, for the BMA-RWB-RWV model we increased the number of draws to 44000, with 4000 initial draws and selection of every 4th draw.

commodity price indices are basically never selected, which makes it impossible to check the three corresponding δ_j parameters for convergence. They are therefore not part of the convergence analysis at $h = 5$.

Table D.1: Summary of simulation inefficiency factors: BMA-SBB-SBV model

		Parameters	Median	Mean	Min	Max	5%	95%
				PCE Deflator Inflation				
$h = 1$	B	4020	2.964	3.806	0.889	13.822	1.531	9.829
	S	201	1.186	1.228	1.039	1.601	1.057	1.500
	$K_\beta,\ K_\sigma$	4221	4.377	4.425	0.765	9.466	1.051	8.202
	D	19	28.760	29.008	22.625	38.255	22.997	36.308
	$p(y_{T+h+1}\|y,x)$	126	1.174	1.201	0.821	2.072	0.946	1.604
$h = 5$	B	4020	3.371	5.122	0.813	27.227	0.939	22.700
	S	201	1.308	1.315	0.983	1.696	1.107	1.552
	$K_\beta,\ K_\sigma$	4221	4.014	3.817	0.743	12.329	0.937	6.697
	D	16	28.829	28.949	19.865	38.823	20.049	38.004
	$p(y_{T+h+1}\|y,x)$	122	1.426	1.676	0.960	4.592	1.031	3.446
				GDP Deflator Inflation				
$h = 1$	B	4020	3.897	4.331	0.680	12.073	1.035	9.928
	S	201	1.229	1.224	0.985	1.496	1.031	1.403
	$K_\beta,\ K_\sigma$	4221	4.170	4.259	0.782	10.578	0.983	8.626
	D	19	29.721	28.795	21.269	35.331	21.704	35.311
	$p(y_{T+h+1}\|y,x)$	126	1.107	1.147	0.789	1.904	0.916	1.495
$h = 5$	B	4020	2.795	4.612	0.791	27.251	0.972	20.652
	S	201	1.294	1.291	1.036	1.742	1.109	1.436
	$K_\beta,\ K_\sigma$	4221	3.834	3.609	0.754	10.792	0.922	7.599
	D	16	28.961	27.524	11.968	38.907	14.308	38.040
	$p(y_{T+h+1}\|y,x)$	122	1.337	1.606	0.954	5.093	1.066	3.103

Note: The table summarizes the simulation inefficiency factors, $(1 + \sum_{f=1}^{\infty} \rho_f)$, for the posterior values of $B = \{\beta_t\}_{t=1}^{T-h}$ with $\beta_t = (\beta_{0t}, \beta_{1t}, \ldots, \beta_{kt})'$, $S = \{\sigma_t^2\}_{t=1}^{T-h}$, $K_\beta = \{\kappa_{0t}, \ldots, \kappa_{kt}\}_{t=1}^{T-h}$ and $K_\sigma = \{\kappa_{k+1,t}\}_{t=1}^{T-h}$, the variable inclusion parameters $D = (\delta_1, \ldots, \delta_k)'$, see Section 2.3, estimated over the 1960Q1-2011Q2 sample, as well as the predictive densities $p(y_{T+h+1}|y,x)$ in (11), estimated in real-time for each quarter for the 1980Q1-2011Q2 sample. The estimated inefficiency factors are based on the Bartlett kernel as in Newey and West (1987) with a bandwidth equal to 4% of the 10000 retained draws.

Table D.2: Summary of convergence diagnostic tests: BMA-SBB-SBV model

		Parameters	10% reject rate	5% rejection rate
		PCE Deflator Inflation		
$h = 1$	B	4020	0.000	0.000
	S	201	0.000	0.000
	$K_\beta,\ K_\sigma$	4221	0.000	0.000
	D	19	0.000	0.000
	$p(y_{T+h+1}\|y, x)$	126	0.000	0.000
$h = 5$	B	4020	0.000	0.000
	S	201	0.000	0.000
	$K_\beta,\ K_\sigma$	4221	0.000	0.000
	D	16	0.000	0.000
	$p(y_{T+h+1}\|y, x)$	122	0.000	0.000
		GDP Deflator Inflation		
$h = 1$	B	4020	0.001	0.000
	S	201	0.005	0.000
	$K_\beta,\ K_\sigma$	4221	0.007	0.003
	D	19	0.063	0.063
	$p(y_{T+h+1}\|y, x)$	126	0.000	0.000
$h = 5$	B	4020	0.001	0.000
	S	201	0.000	0.000
	$K_\beta,\ K_\sigma$	4221	0.000	0.000
	D	16	0.000	0.000
	$p(y_{T+h+1}\|y, x)$	122	0.000	0.000

Note: The table summarizes the convergence test results by reporting the percentage for which the null hypothesis is rejected at significance levels of 10% and 5%. This is done for the posterior values of $B = \{\beta_t\}_{t=1}^{T-h}$ with $\beta_t = (\beta_{0t}, \beta_{1t}, \ldots, \beta_{kt})'$, $S = \{\sigma_t^2\}_{t=1}^{T-h}$, $K_\beta = \{\kappa_{0t}, \ldots, \kappa_{kt}\}_{t=1}^{T-h}$ and $K_\sigma = \{\kappa_{k+1,t}\}_{t=1}^{T-h}$, the variable inclusion parameters $D = (\delta_1, \ldots, \delta_k)'$, see Section 2.3, estimated over the 1960Q1-2011Q2 sample, as well as the predictive densities $p(y_{T+h+1}|y, x)$ in (11), estimated in real-time for each quarter for the 1980Q1-2011Q2 sample. For each of these, we compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed for the first 20% and the last 40% of the retained 10000 draws. The variances of the means are estimated with the Newey and West (1987) variance estimator using a bandwidth of 4% of the respective sample sizes.