Online Appendix for "Real-Time Inflation Forecasting in a Changing World"

Jan J. J. Groen
Federal Reserve Bank of New York

Richard Paap
Erasmus University Rotterdam

Francesco Ravazzolo
Norges Bank and BI Norwegian Business School

Outline

This Online Appendix provides additional detail regarding our methodology, on data used in the article and some additional results. In Section A we describe in detail the Gibbs sampler used for estimating our BMA-based models. More specifics about the data are provided in Section B. Convergence properties for our MCMC approach can be found in Section C, whereas Section D reports on a prior sensitivity analysis for our framework that uses a simulated data set. Finally, in Section E we investigate some properties regarding the in-sample fit of our BMA-based specifications. Note that when we refer in this Appendix numerically to equations, tables, sections and so on, these pertain to the ones in the main article. For example, when we refer below to (8) this is in reference to equation (8) in the main article. All notations and model definitions are similar to those in the main article.

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. Before we describe in detail the different steps of sampler, we need to define the densities that make up the joint density of the data and the latent variables (8). These densities are given by

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

where $\Delta\beta_{jt} = \beta_{jt} - \beta_{j,t-1}$, $\Delta\ln\sigma_t^2 = \ln\sigma_t^2 - \ln\sigma_{t-1}^2$ and $\boldsymbol{\delta}(\cdot)$ is a Dirac delta function.[1] The densities for β_t and $\ln\sigma_t^2$ in (A.1) each consist of two parts. First one where breaks occurs ($\kappa_{jt} = 1$, $j = 0, 1, \ldots, k+1$) and these are drawn from their corresponding distributions. The second component is the case of no break ($\kappa_{jt} = 0$) which results in a degenerate distribution of either the β_{jt}'s or $\ln\sigma_t^2$ at their previous period values, i.e., $\Pr[\beta_{jt} = \beta_{j,t-1}] = 1$ and $\Pr[\ln\sigma_t^2 = \ln\sigma_{t-1}^2] = 1$, represented with a Dirac delta function.

[1] A Dirac delta function $\boldsymbol{\delta}(x) = 0$ if $x \neq 0$ and $\int \boldsymbol{\delta}(x)dx = 1$, and one can interpret the value $\boldsymbol{\delta}(0)$ as probability 1 at $x = 0$. The latter follows from the fact that the Dirac delta function can be regarded as the limit of a sequence of zero-mean normal distributions with variance σ^2 as $\sigma^2 \to 0$. It is similar to using an indicator function that equals 1 when $x = 0$ and zero otherwise as a distribution for x. See, e.g., Kanwal (1998) for more details.

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 (9) 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

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

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 (A.1). 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

$$\begin{aligned} &p(\kappa_{0t}, \ldots, \kappa_{kt} | K_{\beta,-t}, K_\sigma, S, \theta, y, x) \\ &\quad \propto p(y|K, S, \theta, x) p(\kappa_{0t}, \ldots, \kappa_{kt} | K_{\beta,-t}, K_\sigma, S, \theta, x) \\ &\quad \propto p(y_{t+h+1}, \ldots, y_{T-h} | y_{h+1}, \ldots, y_{t+h}, K, S, \theta, x) \\ &\qquad p(y_{t+h}|y_{h+1}, \ldots, y_{t+h-1}, \kappa_1, \ldots, \kappa_t, K_\sigma, S, \theta, x) p(\kappa_{0t}, \ldots, \kappa_{kt} | K_{\beta,-t}, K_\sigma, S, \theta, x), \end{aligned} \quad \text{(A.3)}$$

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). There is no closed form representation for the integrating constant. However, because κ_t can take a finite number of values, we follow Gerlach et al. (2000, Section 3.2) and compute the integrating constant 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 (A.1) 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

$$\begin{aligned} \ln(y_{t+h} - \beta_{0t} - \sum_{j=1}^{k} \delta_j \beta_{jt} x_{jt})^2 &= \ln \sigma_t^2 + u_t \\ \ln \sigma_t^2 &= \ln \sigma_{t-1}^2 + \kappa_{k+1,t} \eta_{k+1,t}, \end{aligned} \tag{A.4}$$

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

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

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 coefficients 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.

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

The full conditional posterior density of π is given by

$$p(\pi|D, q^2, B, S, K, y, x) \propto \prod_{j=1}^{k+1} \pi_j^{a_j-1}(1-p_j)^{b_j-1} \prod_{t=1}^{T-h} \pi_j^{\kappa_{jt}}(1-\pi_j)^{(1-\kappa_{jt})} \tag{A.6}$$

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

$$p(q_j^2|D, \pi, B, S, K, y, x) \propto q_j^{-\nu_j} \exp(-\frac{\bar{\omega}_j}{2q_j^2}) \prod_{t=1}^{T-h} \left(\frac{1}{q_j} \exp(\frac{-(\beta_{jt} - \beta_{j,t-1})^2}{2q_j^2}) \right)^{\kappa_{jt}} \tag{A.7}$$

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

$$p(q_{k+1}^2|D, \pi, B, S, K, y, x) \propto q_{k+1}^{-\nu_{k+1}} \exp(-\frac{\bar{\omega}_{k+1}}{2q_{k+1}^2}) \prod_{t=1}^{T-h} \left(\frac{1}{q_{k+1}} \exp(\frac{-(\ln \sigma_t^2 - \ln \sigma_{t-1}^2)^2}{2q_{k+1}^2}) \right)^{\kappa_{k+1,t}} \tag{A.8}$$

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

Both our inflation rates and the bulk of our predictor variables get revised on a regular basis and we therefore do retrieve the original vintages of the underlying data from, largely, the Federal Reserve Bank of Philadelphia's RTDSM. The RTDSM proxies the original vintages for each quarter by selecting the data that was originally available around the middle of that quarter. When necessary we transform the variables to render them $I(0)$. The decision to transform a variable 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. We summarize the sources, transformation and construction of our data in Table B.1.

Table B.1: Data Sources and Construction: 1960Q1 - 2011Q2

Label	Frequency	Source	Transformation	Definition
Inflation rates				
GDP deflator	Q	RTDSM	$100\Delta \ln$	Gross domestic product deflator, vintage: second month of quarter
PCE deflator	Q	RTDSM	$100\Delta \ln$	Personal consumption expenditures deflator, vintage: second month of quarter
Predictor variables				
ROUT	Q	RTDSM	$100\Delta \ln$	GDP in volume terms, vintage: second month of quarter
RCON	Q	RTDSM	$100\Delta \ln$	Real durable PCE, vintage: second month of quarter
RINV	Q	RTDSM	$100\Delta \ln$	Real residential investment, vintage: second month of quarter
PIMP	Q	RTDSM	$100\Delta \ln$	Imports deflator, vintage: second month of quarter
NFPR	M	RTDSM	$100\Delta \ln$	Non-farm payrolls employment, quarterly average, vintage: second month of quarter
HSTS	M	RTDSM	$\ln$	Housing starts, quarterly average, vintage: second month of quarter
M2	M	ALFRED®, RTDSM	$100\Delta \ln$	M2 monetary aggregate, quarterly average, vintage: second month of quarter
UNEMPL	Q	RTDSM	level	Unemployment as a percentage of labor force, vintage: second month of quarter
YL	M	CRSP, Haver Analytics	level	Level term structure factor: cross-section average of 3-month and 6-month Treasury bill rates, as well as the Fama and Bliss (1987) 1-year, 2-year, 3-year, 4-year and 5-year zero-coupon bond yields, rates samples at end of second month of quarter, unrevised
TS	M	CRSP, Haver Analytics	level	Slope term structure factor: 5-year yield – 3-month T-bill rate, rates samples at end of second month of quarter, unrevised
CS	M	CRSP, Haver Analytics	level	Curvature term structure factor: (2×2-year yield) – (3-month T-bill rate + 5-year yield), rates samples at end of second month of quarter, unrevised
OIL	D & Q	Haver Analytics, RTDSM	$100\Delta \ln$	Real oil price: West Texas Intermediate oil spot price/GDP or PCE deflator for each vintage, oil price sampled at end of second month of quarter, oil price unrevised
FOOD	D & Q	Haver Analytics, RTDSM	$100\Delta \ln$	Real food commodities price: CRB Foodstuffs Price Index /GDP or PCE deflator for each vintage, food commodities price sampled at end of second month of quarter, food commodities price unrevised
RAW	D & Q	Haver Analytics, RTDSM	$100\Delta \ln$	Real raw industrial commodities price: CRB Raw Industrials Price Index/GDP or PCE deflator for each vintage, raw commod. price sampled at end of second month of quarter, raw commod. price unrevised
MS	Q	University of Michigan	level	Reuters/University of Michigan Consumer Survey one-year ahead expected inflation, unrevised.

Notes: In the table, Q (M) indicates that the original data are quarterly (monthly), and D & Q indicates that one component (commodities price) are originally daily and the other quarterly. For sources mnemonics, see Section 3.1, and RTDSM (http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data), ALFRED® (http://alfred.stlouisfed.org/) and University of Michigan (http://www.sca.isr.umich.edu/main.php) data can be had directly online. In case of M2, the RTDSM 1981Q1 and 1981Q2 vintages are incomplete and we replace these M2 vintages with those from the first month of these quarters from ALFRED®. Transformations are: *(i)* level: $X_t = S_t$, *(ii)* $\ln$: $X_t = \ln(S_t)$, and *(iii)* $100\Delta \ln$: $X_t = 100(\ln(S_t) - \ln(S_{t-1}))$. CRB stands for Commodities Research Bureau.

C MCMC Convergence Analysis

To analyze how well the MCMC sampler from Section 2.3 and Appendix A converges, we will report in this section 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.

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 C.1 and C.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

Table C.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 (10), 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.

$h = 5$ 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 C.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.[3] Nonetheless, the corresponding inefficiency 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 4.4 would be served better with our choice of 10000 retained draws. The convergence diagnostic tests in Table C.2 indeed confirm our conclusions regarding efficiency based on the results in Table C.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.

D Prior Sensitivity Analysis

We investigate in this section 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]

$$y_{t+1} = \beta_{0,t} + \beta_{1,t}x_{1t} + \beta_{2,t}x_{2t} + \beta_{3,t}x_{3t} + \sigma_t\varepsilon_{t+1}, \text{ for } t = 1, \ldots, 200 \qquad \text{(D.1)}$$

[3] For both inflation rates at horizon $h = 5$, Figure E.3 indicate that the real oil price and the two real 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 C.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 (10), 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.

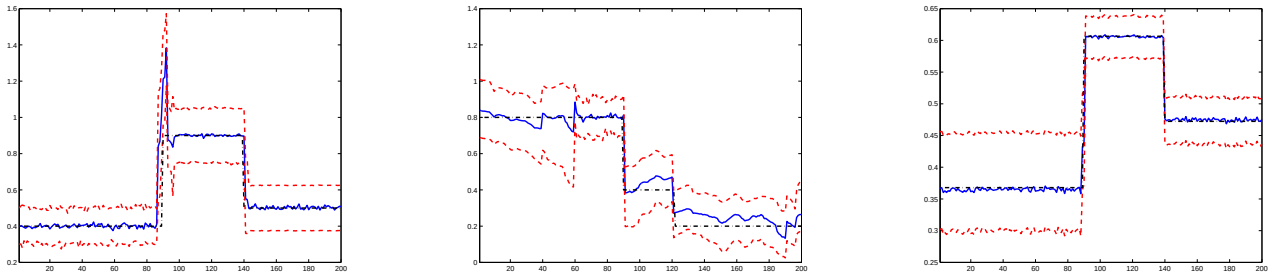
with $\varepsilon_{t+1} \sim \text{NID}(0,1)$ and $x_{j,t} \sim \text{NID}(0,1)$ for $j = 1, \ldots, 3$. The $\beta_{3,t} = 0$ for $t = 1, \ldots, 200$ and hence $x_{3,t}$ is not included in the model. For the first regressor we take as parameters $\beta_{1,t} = 0.4$ for $t = 1, \ldots, 89$, $\beta_{1,t} = 0.9$ 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 (D.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 D.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 higher probability means that we divide b_j by 10 for $j = 1, ...4$.[4] A higher expected prior

[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

Figure D.1: Posterior estimates of the time-varying parameters implied by DGP (D.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 (D.1).

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 D.1 summarizes the prior settings. Note, some of the prior settings are quite extreme but they serve to illustrate our prior sensitivity analysis.

Table D.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 D.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 for the base case. In the final three columns of the table we display the results where we

base case we multiply these parameters by 1.5.

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 D.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

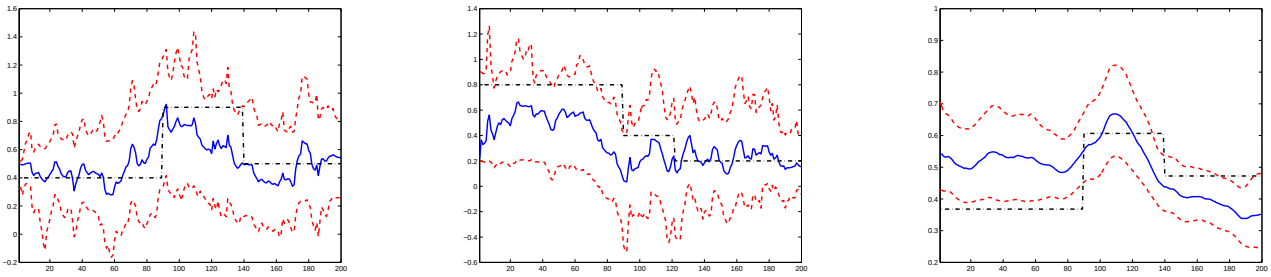
Posterior results for different prior break probability and expected prior size of a break, see Table D.1

From Table D.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 (D.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 D.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 D.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 D.3 summarizes them. From Figure D.3 it is clear that assuming *a priori* large sized breaks in σ_t^2 will result in a very

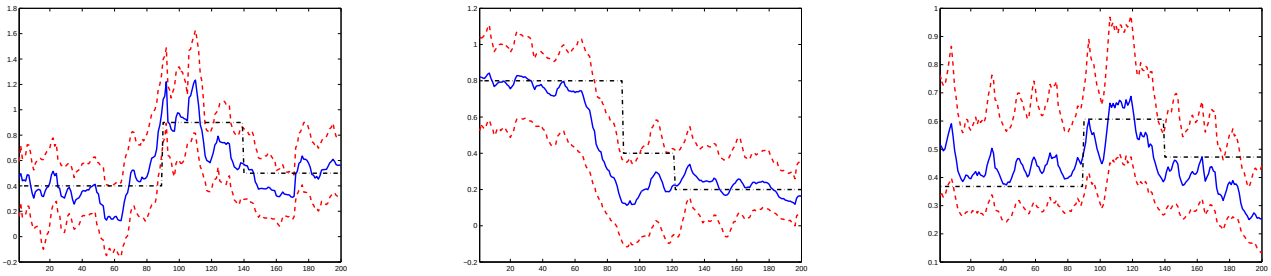
[5]The posterior estimates for the case of a lower prior break probability and a higher expected prior break size for the x variables are very similar.

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



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

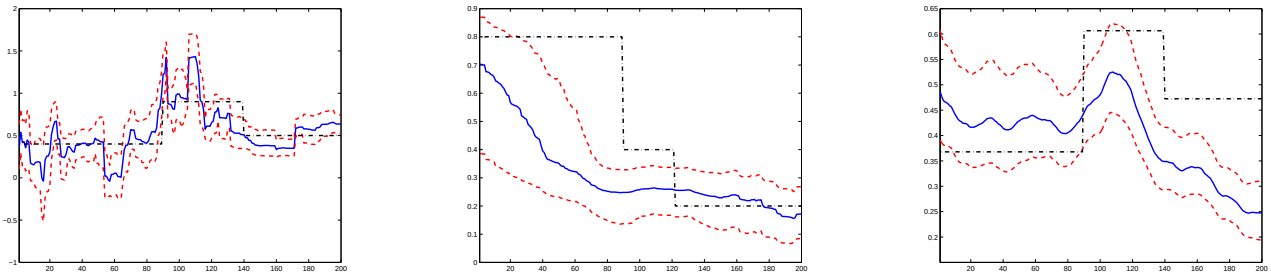


Note: See Figure D.1, but now the posterior results are based on a higher prior break probability as well as a higher prior break size for σ_t^2 than in the base case.

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 (D.1).

Altering the prior assumptions for the intercept also can lead to peculiar posterior estimation results. In Figure D.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 D.4: Posterior estimates of the time-varying parameters implied by DGP (D.1): Lower prior break probabilities and higher prior break sizes for the intercept



Note: See Figure D.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 D.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.

E In-Sample Posterior Inference

In this section we report on some in-sample properties of different specifications of our general forecasting model (2)–(3) based on our prior choices in Table 2 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 how informative the data are for posterior inference on time-variation (Section E.1), 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 (Section E.2), and an analysis on how

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

the combination of model averaging and differing degrees of structural instability affects predictor variable selection in our framework (Section E.3). For the in-sample estimation, we use the 2011Q3 data vintage for all variables, which contains data up to in 2011Q2. We focus here 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 ($h = 5$).

E.1 Posterior Inference on Time-Variation

The marginal posterior distributions of the regression parameters and latent breaks for all 19 predictor variables within our BMA-based specifications are averaged over all possible model specifications, and 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 therefore be different. A proper interpretation is only possible when we condition on each individual regression specification, which is an arduous task given the number of specifications we consider in this paper.[7] But, as we mentioned in Section 3.2, analyzing the posterior means of the individual break probabilities π_j can still be useful for determining how informative the data has been for the break estimation in our BMA-based specifications given our prior choices (see Table 2). We report these in Table E.1 for the BMA-SBB-SBV model.[8]

What immediately becomes clear from Table E.1 is that at 1% the posterior means of the break probability of the intercept π_0 are well below the mean 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. Only for the error variance the posterior mean break probability remains fairly close to its prior counterpart. The patterns in Table E.1 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.

[7] For example, the analysis in Section E.3 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.

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

Table E.1: BMA-SBB-SBV marginal posterior break probabilities π_j: 1960Q1-2011Q2

		PCE Deflator Inflation		*GDP Deflator Inflation*	
		$h=1$	$h=5$	$h=1$	$h=5$
	π_j^{prior}	π_j^{post}	π_j^{post}	π_j^{post}	π_j^{post}
Intercept	0.09	0.01	0.01	0.00	0.02
y_t	0.02	0.18	0.26	0.19	0.23
y_{t-1}	0.02	0.18	0.24	0.15	0.21
y_{t-2}	0.02	0.12	0.25	0.13	0.19
y_{t-3}	0.02	0.12	0.19	0.09	0.15
ROUTP	0.02	0.14	0.23	0.15	0.20
RCONS	0.02	0.14	0.37	0.14	0.23
RINVR	0.02	0.12	0.22	0.10	0.14
PIMP	0.02	0.10	0.21	0.09	0.15
UNEMPL	0.02	0.08	0.21	0.18	0.19
HSTS	0.02	0.09	0.46	0.37	0.22
NFPR	0.02	0.13	0.15	0.09	0.17
M2	0.02	0.14	0.16	0.11	0.22
YL	0.02	0.10	0.13	0.09	0.22
TS	0.02	0.17	0.21	0.17	0.17
CS	0.02	0.15	0.20	0.13	0.14
OIL	0.02	0.26	0.16	0.06	0.17
FOOD	0.02	0.15	0.27	0.08	0.21
RAW	0.02	0.10	0.29	0.04	0.04
MS	0.02	0.08	0.11	0.04	0.10
σ_t	0.28	0.29	0.29	0.29	0.29

Notes: See Section 3.1 and Table B.1 for variable mnemonics. The values in the column headed 'π_j^{prior}' represent the prior break probabilities for each $j = 0, \ldots, k+1$ implied by (6) based on the prior choices from Table 2, whereas the columns headed 'π_j^{post}' report on the posterior means of π_j for each $j = 0, \ldots, k+1$ across different inflation rates and horizons h.

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 E.1. 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 based on the settings in Table 2. Given this, it therefore appears that in case of BMA-SBB-SBV, which uses similar prior settings, the error variance break probability is less well identified than in case of BMA-SBV. 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.

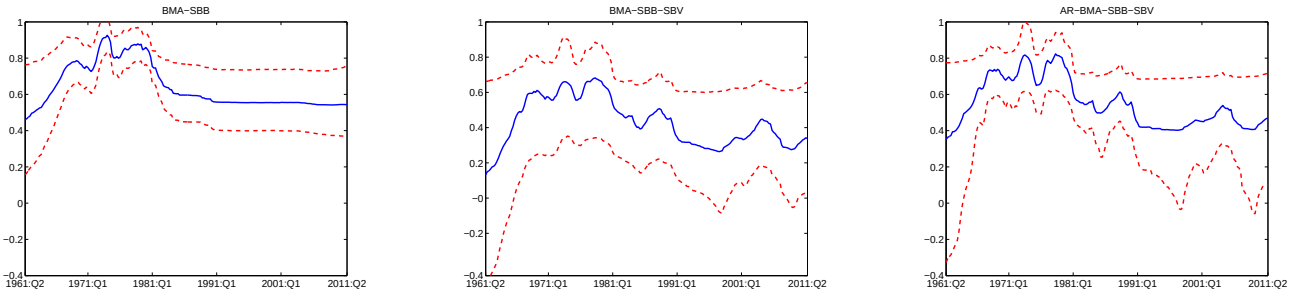
E.2 Some Implied Inflation Characteristics

We now investigate whether our BMA-based framework from Section 2.1 indeed implies similar inflation properties as identified by the existing literature. Figure E.1 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.[9] In this figure we also display similar posterior results for the AR-BMA-SBB-SBV specification (16) described in Section 4.1. 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. Additionally, Figure E.2 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.

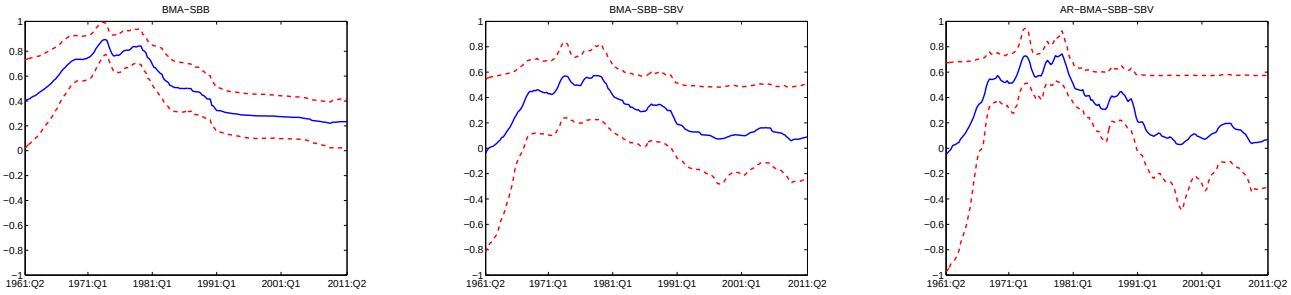
Although there are some differences in the level and variation of inflation persistence for BMA-SBB versus the other models, Figure E.1 overall suggests a similar pattern for inflation persistence across models: relatively low in the 1960's, a substantial increase in the 1970's, and reducing drastically in the second part of the 1980's. In case of the shock variance (Figure E.2), all models exhibit a downward shift from the late 1980s, early 1990s onwards, albeit that the innovation variance for BMA-SBV varies in a more smooth manner. The shock variances move 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

[9] 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 E.1: Posterior estimates of time-varying inflation persistence



GDP deflator, $h = 1$

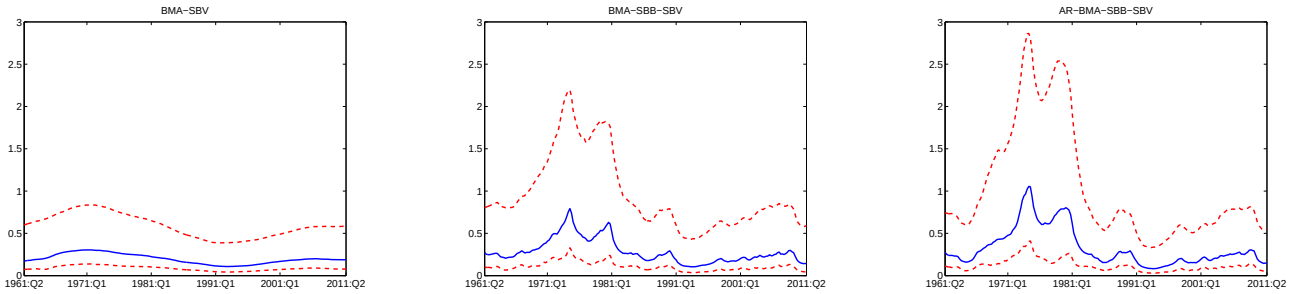


PCE deflator, $h = 1$

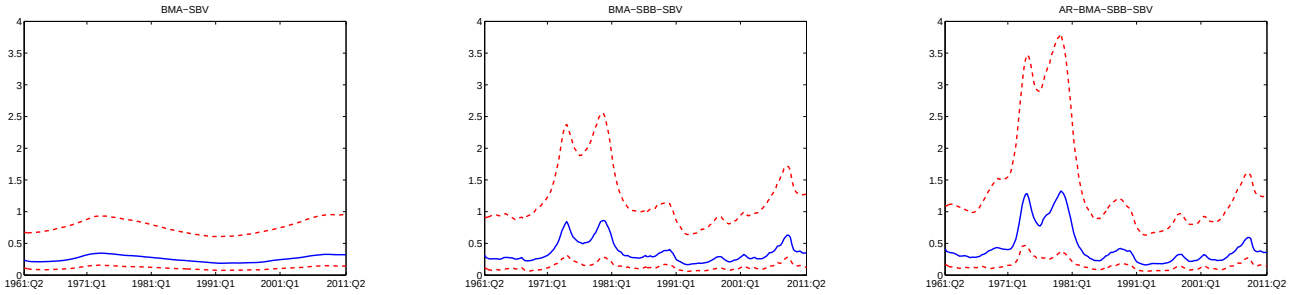
Note: The solid lines represent the posterior medians of the persistence parameters for the BMA-SBB (first column), BMA-SBB-SBV (second column) and AR-BMA-SBB-SBV (third column). Persistence is computed by averaging the sum of the included autoregressive parameters across all model specifications using the posterior model probabilities. The dashed lines denote the 16th and 84th percentiles of the posterior distributions.

able to reproduce in a satisfactorily manner the stylized facts on inflation persistence and volatility for the post-WWII U.S. period.

Figure E.2: 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 16th and 84th percentiles of the posterior distributions.

E.3 Posterior Variable Selection

In Figure E.3 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.[10] 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 16 to 17 predictors, including the intercept. 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

[10] To preserve space as well as the fact that we want to focus on the impact of ignoring instability in certain parts of the specification on variable selection, we did not report the results for the BMA-RWB-RWV model. They are qualitatively the same as for the the other BMA models with structural instability.

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 3 to 4 (including intercept).

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 E.3 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 (including the intercept) is in the 6 to 7 range for the current quarter horizon across inflation measures, whereas for $h = 5$ this number fluctuates between 3 and 4. One can conclude from Figure E.3 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.

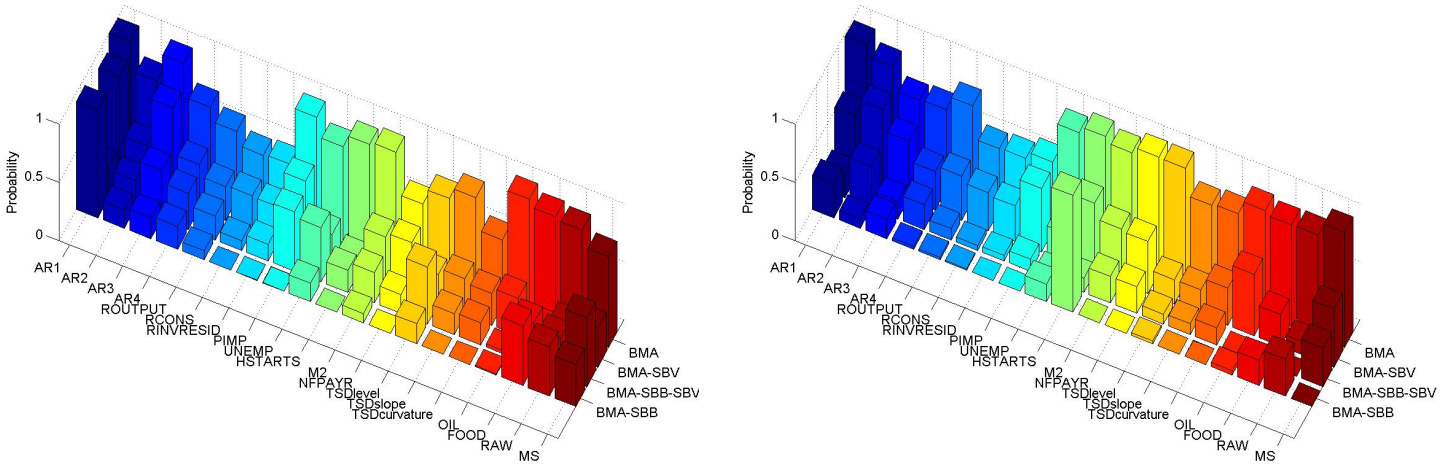
To shed more light on what combinations of explanatory variables dominate the Bayesian model averaging, we can consider the top 10 models with the highest posterior model probability. In the interest of brevity we report these only for the BMA-SBB-SBV variant of (2), see Table E.2. In general, the conclusions drawn from the results in Figure E.3 are confirmed, that is, the most selected variables do show up most frequently amongst those top 10 models. 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. Finally, we notice a fair amount of variability in the composition of these specifications.

Unreported results[11] 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 3 and 15 (including intercept) with a posterior mode of 7 selected predictors. For the quarterly inflation rate one-year ahead, the model size fluctuates for GDP deflator inflation between 2 and 10 (2 and 9 for PCE deflator inflation) and the posterior mode is 5 (4 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.

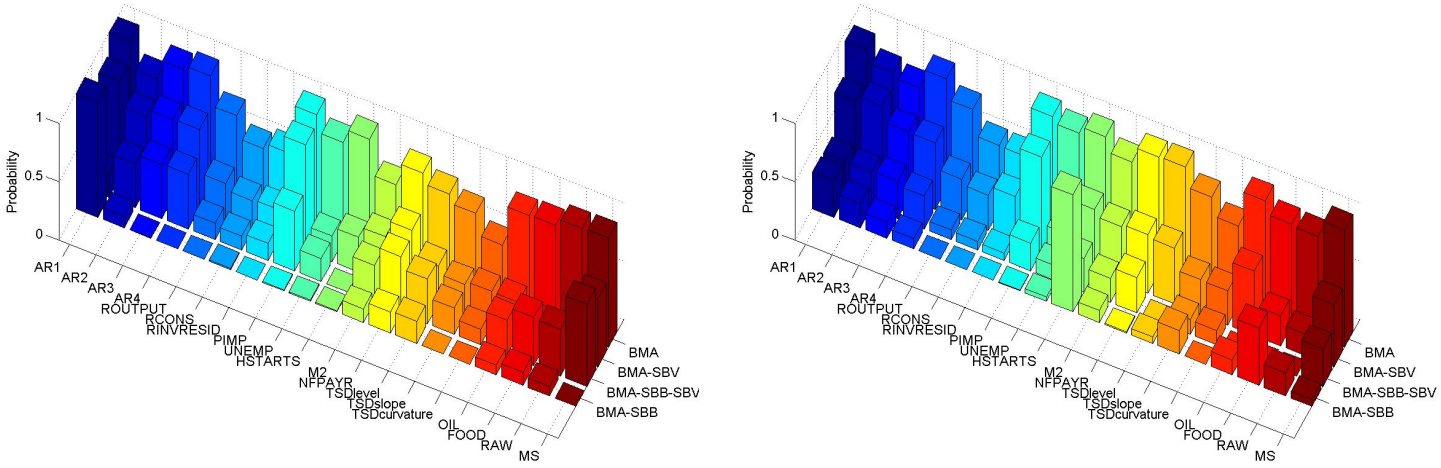
So what does the varying degrees of parsimony of our different BMA-based specifica-

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

Figure E.3: 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.

Table E.2: 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 and Table B.1 for variable mnemonics.

Table E.3: In-sample fit of the BMA-based specifications: 1960Q1-2011Q2

	PCE Deflator Inflation		*GDP Deflator Inflation*	
	$h = 1$	$h = 5$	$h = 1$	$h = 5$
BMA-SBB-SBV	-394.15	-346.41	-495.48	-429.42
BMA-SBV	-458.62	-327.35	-570.25	-409.51
BMA-SBB	-380.52	-263.58	-470.26	-291.68
BMA	-360.70	-255.14	-487.66	-354.03
BMA-RWB-RWV	-319.38	-321.47	-336.46	-346.33

Notes: See Table 1 for model mnemonics. The values in the table are posterior means of the Bayesian information criterion [BIC] for each model.

tions mean for the in-sample fit of these models? One way to get insight into this is to compute the Bayesian information criterion [BIC] of each model. Posterior marginal BIC values are reported in Table E.3. The results in this table suggest that the BMA-SBB-SBV and BMA-SBV specifications generally seem to be the preferred specifications for the full 1960-2011 sample, with BMA-SBV having an edge at the nowcasting horizon and BMA-SBB-SBV at $h = 5$.

References

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

— (1997), "Semiparametric Bayesian Inference for Time Series with Mixed Spectra," *Journal of the Royal Statistical Society, Series B*, B, 255–268.

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

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

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

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

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

Gerlach, R., Carter, C., and Kohn, R. (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 *Bayesian Statistics*, eds. Berger, J., Bernardo, J., Dawid, A., and Smith, A., Oxford: Oxford University Press.

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

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

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

Kanwal, R. P. (1998), *Generalized Functions: Theory and Technique, 2nd Ed.*, Boston: Birkhäuser.

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

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

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

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

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