Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting

Kenichiro McAlinn, Knut Are Aastveit, Jouchi Nakajima, & Mike West

Supplementary Material

A Appendix: Summary of MCMC for Dynamic BPS

A.1 Overview and Initialization

This appendix summarizes algorithmic details of implementation of the MCMC computations for dynamic BPS model fitting of Section 2.4. This involves a standard set of steps in a customized three-component block Gibbs sampler: the first component samples the latent agent states, the second, the second samples the dynamic BPS model states/parameters, and the third component samples the observation variance. The latter two involves a modified FFBS algorithm central to MCMC in all conditionally normal DLMs (Frühwirth-Schnatter 1994; West and Harrison 1997, Sect 15.2; Prado and West 2010, Sect 4.5).

In our sequential learning and forecasting context, the full MCMC analysis is performed anew at each time point as time evolves and new data are observed. We detail MCMC steps for analysis based on data over times $t = 1{:}T$ for any chosen T.

Standing at time $t = 0$, the decision maker has initial information summarized via in terms of $\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$ and $\boldsymbol{V}_0 \sim IW(n_0, \boldsymbol{D}_0)$, independently. Here the $q \times q$ variance matrix $\boldsymbol{V}_0$ has the inverse Wishart distribution with $n > 0$ degrees of freedom and prior "sum-of-squares" matrix $\boldsymbol{D}_0$. Equivalently, the precision matrix $\boldsymbol{V}_0^{-1} \sim W(h_0, \boldsymbol{D}_0^{-1})$, the Wishart distribution with $h_0 = n_0 + q - 1$ and mean $h_0 \boldsymbol{D}_0^{-1}$ so that the initial estimate $\boldsymbol{D}_0/h_0$ is the prior harmonic mean of $\boldsymbol{V}_0$. Model specification is completed with two chosen discount factors: β, defining the extent of time variation in the evolution of the states $\boldsymbol{\theta}_t$, and δ defining levels of variation in the evolution of the volatility matrices $\boldsymbol{V}_t$.

At time T, the decision maker has accrued information $\{y_{1:T}, \mathcal{H}_{1:T}\}$. The MCMC analysis is then run iteratively as follows.

Initialization: First, initialize by setting

$$\boldsymbol{F}_t = \begin{pmatrix} 1 & \boldsymbol{f}'_{t1} & 0 & \boldsymbol{0}' & \cdots & \cdots & 0 & \boldsymbol{0}' \\ 0 & \boldsymbol{0}' & 1 & \boldsymbol{f}'_{t2} & & & & \vdots \\ \vdots & & & & \ddots & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & \boldsymbol{f}'_{tq} \end{pmatrix}, \tag{8}$$

for each $t = 1{:}T$, with elements set at some chosen initial values of the latent agent states. Initial values can be chosen arbitrarily. One obvious and appropriate choice– our recommended default choice– is to simply generate agent states from their priors, i.e., from the agent forecast distributions, $\boldsymbol{x}_{tj} \sim h_{tj}(\boldsymbol{x}_{tj})$ independently

for all $t = 1{:}T$ and $j = 1{:}J$. This is easily implemented in cases when the agent forecasts are T or normal distributions, or can be otherwise directly sampled; we use this in our analyses reported in the paper, and recommend it as standard. An obvious alternative initialization is to simply set $\boldsymbol{x}_{tj} = \boldsymbol{y}_t$ for each t, j, though we prefer to initialize with some inherent dispersion in starting values. Ultimately, since the MCMC is rapidly convergent, choice of initial values is not critical. Given initial values of agent factor vector $\boldsymbol{x}_{tj} = (x_{t1j}, x_{t2j}, ..., x_{tqj})'$ for each agent $j = 1{:}J$ and each time t, the $\boldsymbol{F}_t$ matrices are initialized with series-specific row entries from $\boldsymbol{f}_{tr} = (x_{tr1}, x_{tr2}, ..., x_{trJ})'$ for each $r = 1{:}q$.

A.2 Three Sampling Steps in Each MCMC Iterate

Following initialization, the MCMC iterates repeatedly to resample three sets of conditional posteriors to generate the MCMC samples from the target posterior $p(\boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{1:T}, \boldsymbol{V}_{1:T} | \boldsymbol{y}_{1:T}, \mathcal{H}_{1:T})$. These conditional posteriors and algorithmic details of their simulation are as follows.

A.2.1 Per MCMC Iterate Step 1: Sampling BPS DLM parameters $\boldsymbol{\theta}_{1:T}$

Conditional on any values of the latent agent states and observation error, we are in the setting of a conditionally normal, multivariate DLM with the agent states as

known predictors based on their specific values. The BPS DLM form,

$$\boldsymbol{y}_t = \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \nu_t \sim N(0, \boldsymbol{V}_t),$$
$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{W}_t),$$

has known elements $\boldsymbol{F}_t, \boldsymbol{W}_t$ and specified initial prior at $t = 0$. The implied conditional posterior for $\boldsymbol{\theta}_{1:T}$ then does not depend on $\mathcal{H}_{1:T}$, reducing to $p(\boldsymbol{\theta}_{1:T}|\boldsymbol{X}_{1:T}, \boldsymbol{V}_{1:T}, \boldsymbol{y}_{1:T})$. This is simulated using the efficient and standard FFBS algorithm (e.g. Frühwirth-Schnatter 1994; West and Harrison 1997, Sect 15.2; Prado and West 2010, Sect 4.5). In detail, this proceeds as follows.

Forward filtering: For each $t = 1{:}T$ in sequence, perform the standard one-step filtering updates to compute and save the sequence of sufficient statistics for the on-line posteriors $p(\boldsymbol{\theta}_t|\boldsymbol{X}_{1:t}, \boldsymbol{V}_{1:t}, \boldsymbol{y}_{1:t})$ at each t. The summary technical details are as follows:

1. *Time $t-1$ posterior:*

$$\boldsymbol{\theta}_{t-1}|\boldsymbol{X}_{1:t-1}, \boldsymbol{V}_{1:t-1}, \boldsymbol{y}_{1:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}),$$

   with point estimate $\boldsymbol{m}_{t-1}$ of $\boldsymbol{\theta}_{t-1}$.

2. *Update to time t prior:*

$$\boldsymbol{\theta}_t|\boldsymbol{X}_{1:t-1}, \boldsymbol{V}_{1:t-1}, \boldsymbol{y}_{1:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{R}_t) \quad \text{with} \quad \boldsymbol{R}_t = \boldsymbol{C}_{t-1}/\delta,$$

with (unchanged) point estimates $\boldsymbol{m}_{t-1}$ of $\boldsymbol{\theta}_t$, but with increased uncertainty relative to the time $t-1$ posteriors, the level of increased uncertainty being defined by the discount factors.

3. *1-step predictive distribution:* $\boldsymbol{y}_t|\boldsymbol{X}_{1:t}, \boldsymbol{V}_{1:t}, \boldsymbol{y}_{1:t-1} \sim T_{\beta n_{t-1}}(\boldsymbol{f}_t, \boldsymbol{Q}_t)$ where

$$\boldsymbol{f}_t = \boldsymbol{F}_t \boldsymbol{m}_{t-1} \quad \text{and} \quad \boldsymbol{Q}_t = \boldsymbol{F}_t \boldsymbol{R}_t \boldsymbol{F}_t' + \boldsymbol{V}_t.$$

4. *Filtering update to time t posterior:*

$$\boldsymbol{\theta}_t|\boldsymbol{V}_{1:t}, \boldsymbol{X}_{1:t}, \boldsymbol{y}_{1:t} \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t),$$

with defining parameters $\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t \boldsymbol{e}_t$ and $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t'$, based on 1-step forecast error $\boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{f}_t$ and the state adaptive coefficient vector (a.k.a. "Kalman gain") $\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t' \boldsymbol{Q}_t^{-1}$.

Backward sampling: Having run the forward filtering analysis up to time T, the backward sampling proceeds as follows.

a. *At time T:* Simulate $\boldsymbol{\theta}_T$ from the final multivariate normal posterior

$$p(\boldsymbol{\theta}_T|\boldsymbol{X}_{1:T}, \boldsymbol{V}_{1:T}, \boldsymbol{y}_{1:T}) = N(\boldsymbol{m}_T, \boldsymbol{C}_T).$$

b. *Recurse back over times $t = T-1, T-2, \ldots, 0$:* At each time t, simulate the state $\boldsymbol{\theta}_t$ from the conditional posterior $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{X}_{1:t}, \boldsymbol{V}_{1:t}, \boldsymbol{y}_{1:t})$; this

is multivariate normal with mean vector $\boldsymbol{m}_t + \delta(\boldsymbol{\theta}_{t+1} - \boldsymbol{m}_t)$ and variance matrix $\boldsymbol{C}_t(1-\delta)$.

A.2.2 Per MCMC Iterate Step 2: Sampling BPS DLM parameters $\boldsymbol{V}_{1:T}$

Conditional on the sampled values of the BPS DLM parameters $\boldsymbol{\theta}_{1:T}$ and latent agent states $\boldsymbol{X}_{1:T}$, the next step in the MCMC iterate samples the full conditional posterior of the sequence of volatility matrices, generating a draw from $p(\boldsymbol{V}_{1:T}|\boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{1:T}, \boldsymbol{y}_{1:T})$.

Forward filtering: For each $t = 1{:}T$ in sequence, update and save the forward filtering summaries $(n_t, \boldsymbol{D}_t)$ of on-line posteriors

$$\boldsymbol{V}_t|\boldsymbol{X}_{1:t}, \boldsymbol{\theta}_{1:t}, \boldsymbol{y}_{1:t} \sim IW(n_t, \boldsymbol{D}_t),$$

given by $n_t = h_t - q + 1$ where $h_t = \beta h_{t-1} + 1$, and $\boldsymbol{D}_t = \beta \boldsymbol{D}_{t-1} + (\boldsymbol{y}_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t)(\boldsymbol{y}_t - \boldsymbol{F}_t'\boldsymbol{\theta}_t)'$.

Backwards sampling: Having run the forward filtering analysis up to time T, the backward sampling proceeds as follows.

a. *At time T:* Simulate $\boldsymbol{V}_T$ from the final inverse Wishart posterior $IW(n_T, \boldsymbol{D}_T)$.

b. *Recurse back over times* $t = T-1, T-2, \ldots, 0$: At time t, sample $\boldsymbol{V}_t$ from the conditional posterior $p(\boldsymbol{V}_t|\boldsymbol{V}_{t+1}, \boldsymbol{X}_{1:t}, \boldsymbol{\theta}_{1:t}, \boldsymbol{y}_{1:t})$. Algorithmically, this

is achieved via

$$\boldsymbol{V}_t^{-1} = \beta \boldsymbol{V}_{t+1}^{-1} + \boldsymbol{\Upsilon}_t \quad \text{where} \quad \boldsymbol{\Upsilon}_t \sim W((1-\beta)h_t, \boldsymbol{D}_t^{-1}),$$

and where the Υ_t are independent over t.

A.2.3 Per MCMC Iterate Step 3: Sampling the latent agent states $\boldsymbol{X}_{1:T}$

Conditional on most recently sampled values of the BPS DLM parameters $\boldsymbol{\Phi}_{1:T}$, the MCMC iterate completes with resampling of the latent agent states from their full conditional posterior $p(\boldsymbol{X}_{1:t}|\boldsymbol{\Phi}_{1:t}, \boldsymbol{y}_{1:t}, \mathcal{H}_{1:t})$. It is immediate that the $\boldsymbol{X}_t$ are conditionally independent over time t in this conditional distribution, with time t conditionals

$$p(\boldsymbol{X}_t|\boldsymbol{\Phi}_t, \boldsymbol{y}_t, \mathcal{H}_t) \propto N(\boldsymbol{y}_t|\boldsymbol{F}_t\boldsymbol{\theta}_t, \boldsymbol{V}_t) \prod_{j=1:J} h_{tj}(\boldsymbol{x}_{tj}). \tag{9}$$

Several comments are relevant to studies with different forms of the agent forecast densities.

1. *Multivariate normal agent forecast densities:* In cases when each of the agent forecast densities is normal, the posterior in eqn. (9) yields a multivariate normal distribution for vectorized $\boldsymbol{X}_t$. Computation of its defining parameters and then drawing a new sample vector $\boldsymbol{X}_t$ are trivial.

2. In some cases, as in our study in this paper, the agent forecast densities will be those of Student T distributions. In our case study the five agents rep-

resent conjugate exchangeable dynamic linear models in which all forecast densities are multivariate T, with parameters varying over time and with step-ahead forecast horizon. In such cases, standard Bayesian augmentation methods apply to enable simulation. Each multivariate T distribution is expressed as a scale mixture of multivariate normals, with the mixing scale parameters introduced as inherent latent variables with inverse gamma distributions. This expansion of the parameter space makes the multivariate T distributions conditional multivariate normals, and the mixing scales are resampled (from implied conditional posterior inverse gamma distributions) each MCMC iterate along with the agent states. This is again a standard MCMC approach and much used in Bayesian time series, as in other areas (e.g. Frühwirth-Schnatter 1994; West and Harrison 1997, Chap 15). Then, conditional on the current values of these latent scales, sampling the $\boldsymbol{X}_t$ reduces technically to that conditional normals above.

Specifically, suppose that $h_{tj}(\boldsymbol{x}_{tj})$ is density of the normal $T_{n_{tj}}(\boldsymbol{h}_{tj}, \boldsymbol{H}_{tj})$; the notation means that $(\boldsymbol{x}_{tj} - \boldsymbol{h}_{tj})/\sqrt{\boldsymbol{H}_{tj}}$ has a standard multivariate Student T distribution with n_{tj} degrees of freedom. Then latent scale factors ϕ_{tj} exist such that: *(i)* conditional on ϕ_{tj}, latent agent factor $\boldsymbol{x}_{tj}$ has a conditional multivariate normal density $\boldsymbol{x}_{tj}|\phi_{tj} \sim N(\boldsymbol{h}_{tj}, \boldsymbol{H}_{tj}/\phi_{tj})$ independently over t, j; *(ii)* the ϕ_{tj} are independent over t, j with gamma distributions, $\phi_{tj} \sim G(n_{tj}/2, n_{tj}/2)$. Then, at each MCMC step, the above normal update for latent agent states is replaced by normal simulations conditional on the

ϕ_{tj}. Following this, we resample values of the ϕ_{tj} from their trivially implied conditional gamma posteriors.

3. In some cases, agent densities may be more elaborate mixtures of normals, such as (discrete or continuous) location and/or scale mixtures that represent asymmetric distributions. The same augmentation strategy can be applied in such cases, with augmented parameters including location shifts in place of, or in addition to, scale shifts.

4. In other cases, we may be able to directly simulate the agent forecast distributions and evaluate forecast density functions at any point, but do not have access to analytic forms. One class of examples is when the agents are simulation models, e.g., DSGE models. Another involves forecasts in terms of histograms. In such cases, MCMC will proceed using some form of Metropolis-Hastings algorithm, or accept/reject methods, or importance sampling for the latent agent states.

   For example, suppose we only have access to simulations from the agent forecast distributions, in terms of I independent draws from each collated in the simulated matrix $\boldsymbol{X}_t^{(i)}$ for $i = 1{:}I$. We can apply importance sampling as follows: *(a)* compute the marginal likelihood values $p(\boldsymbol{y}_t|\boldsymbol{\Phi}_t, \boldsymbol{X}_t^{(i)}, \mathcal{H}_t)$ for each $i = 1{:}I$; *(b)* compute and normalize the implied importance sampling weights $w_{ti} \propto N(\boldsymbol{y}_t|\boldsymbol{\Phi}_t, \boldsymbol{X}_t^{(i)}, \mathcal{H}_t)$, and then *(c)* resample latent agent states for this MCMC stage according to the probabilities these weights define.

B Appendix: Additional Graphical Summaries from Macroeconomic Analysis

This appendix lays out additional graphical summaries of results from the macroeconomic forecasting analysis in the paper, providing material supplementary to that discussed in Section 3.

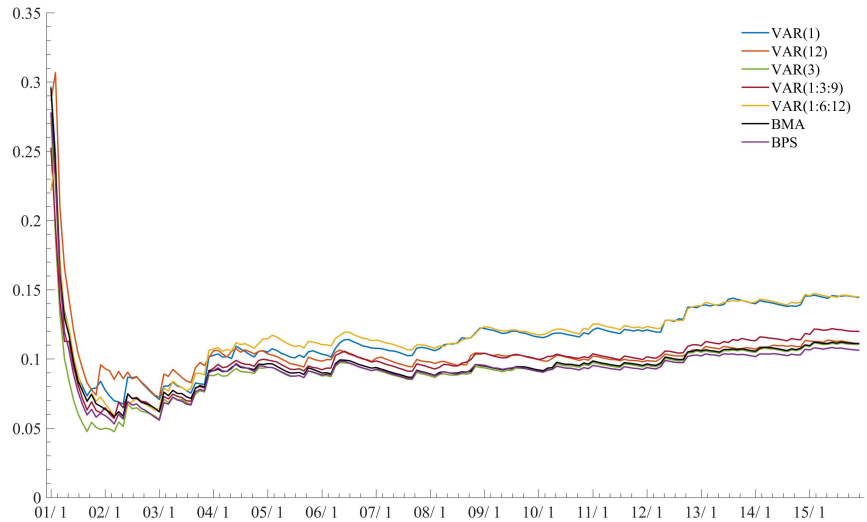

Figure C1: US macroeconomic forecasting 2001/1-2015/12: Mean squared 1-step ahead forecast errors $\text{MSFE}_{1:t}(1)$ of wage (w) sequentially revised at each of the $t = 1{:}180$ months.

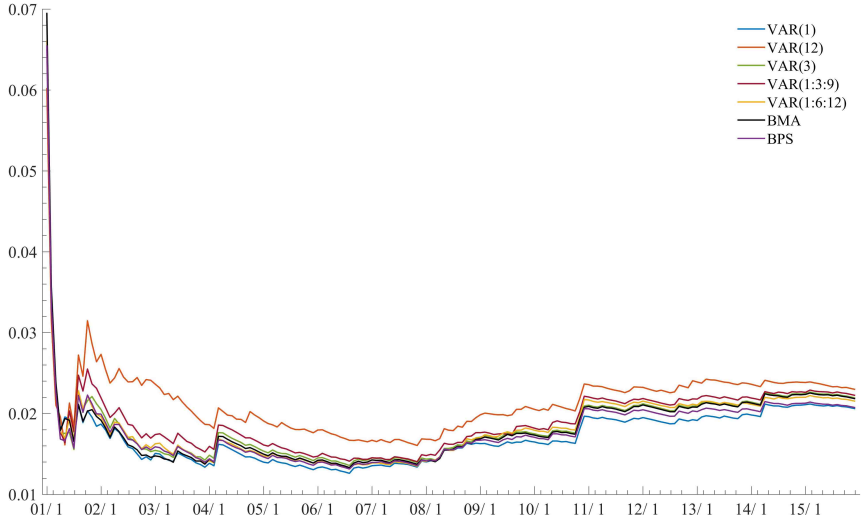

Figure C2: US macroeconomic forecasting 2001/1-2015/12: Mean squared 1-step ahead forecast errors $\text{MSFE}_{1:t}(1)$ of unemployment rate (u) sequentially revised at each of the $t = 1{:}180$ months.

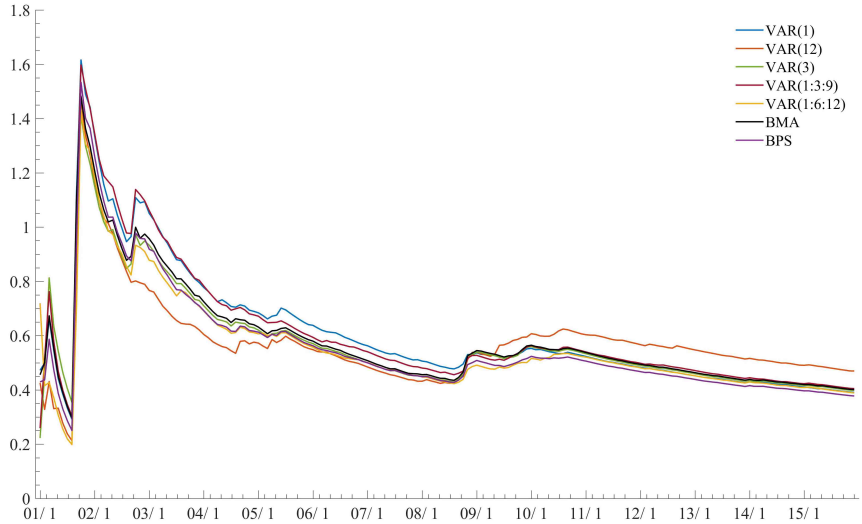

Figure C3: US macroeconomic forecasting 2001/1-2015/12: Mean squared 1-step ahead forecast errors $\text{MSFE}_{1:t}(1)$ of consumption (c) sequentially revised at each of the $t = 1{:}180$ months.

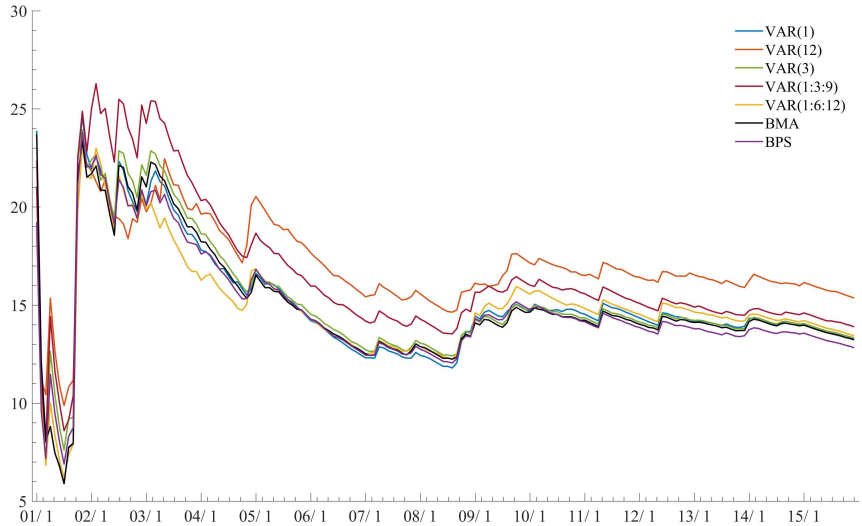

Figure C4: US macroeconomic forecasting 2001/1-2015/12: Mean squared 1-step ahead forecast errors $\text{MSFE}_{1:t}(1)$ of investment (i) sequentially revised at each of the $t = 1{:}180$ months.

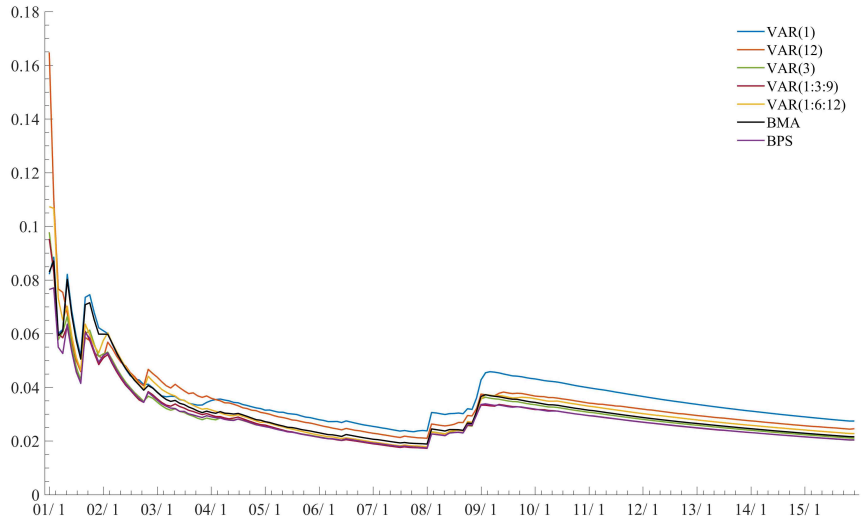

Figure C5: US macroeconomic forecasting 2001/1-2015/12: Mean squared 1-step ahead forecast errors $\text{MSFE}_{1:t}(1)$ of interest rate (r) sequentially revised at each of the $t = 1{:}180$ months.

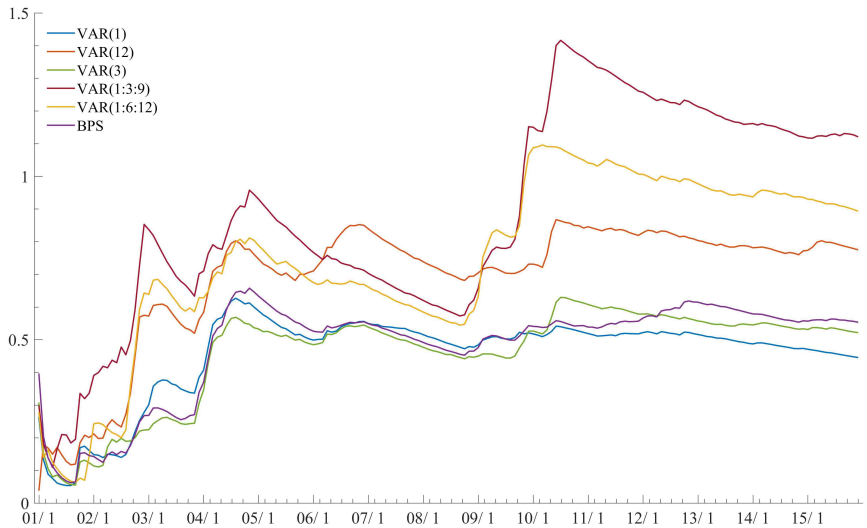

Figure C6: US macroeconomic forecasting 2001/1-2015/12: Mean squared 12-step ahead forecast errors $\text{MSFE}_{1:t}(12)$ of wage (w) sequentially revised at each of the $t = 1{:}180$ months.

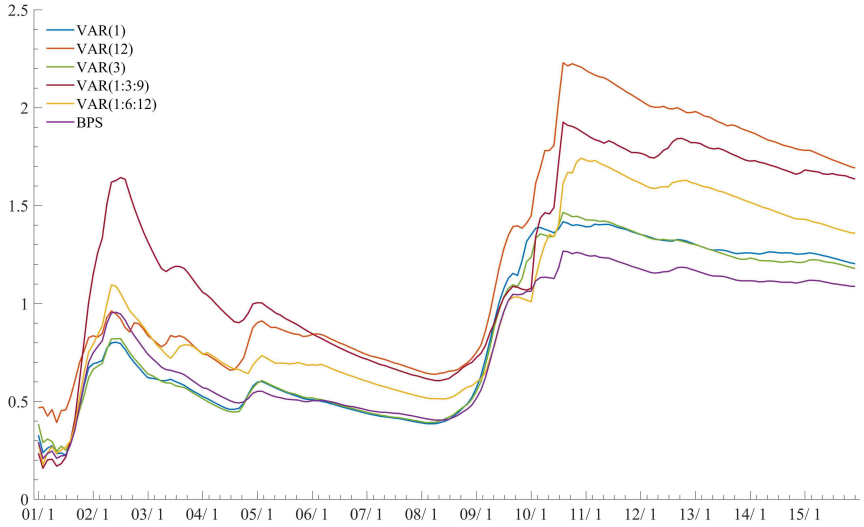

Figure C7: US macroeconomic forecasting 2001/1-2015/12: Mean squared 12-step ahead forecast errors $\text{MSFE}_{1:t}(12)$ of unemployment rate (u) sequentially revised at each of the $t = 1{:}180$ months.

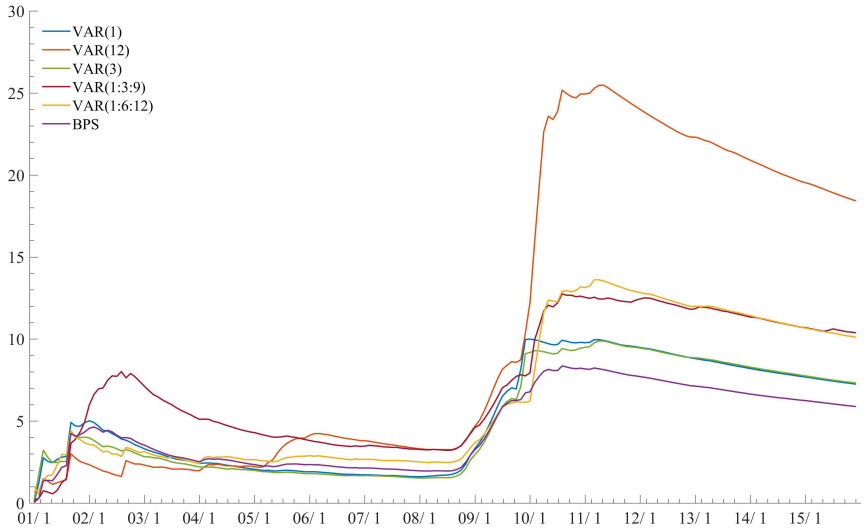

Figure C8: US macroeconomic forecasting 2001/1-2015/12: Mean squared 12-step ahead forecast errors $\text{MSFE}_{1:t}(12)$ of consumption (c) sequentially revised at each of the $t = 1{:}180$ months.

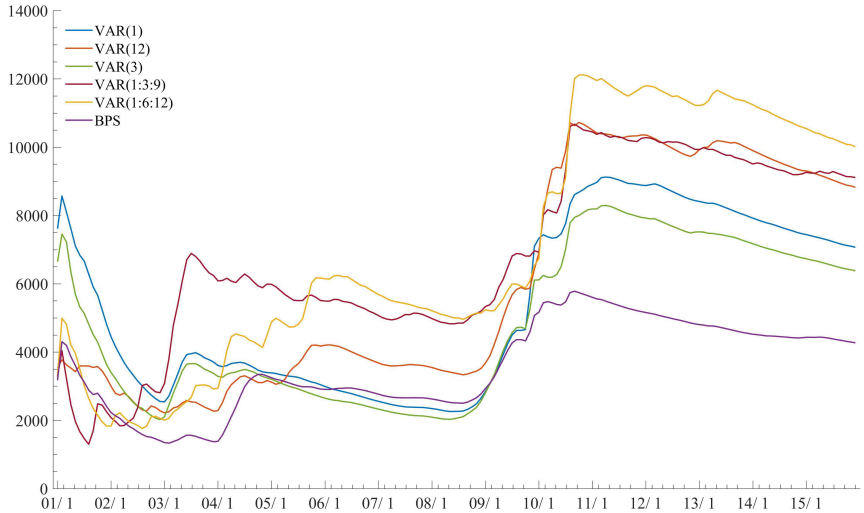

Figure C9: US macroeconomic forecasting 2001/1-2015/12: Mean squared 12-step ahead forecast errors $\text{MSFE}_{1:t}(12)$ of investment (i) sequentially revised at each of the $t = 1{:}180$ months.

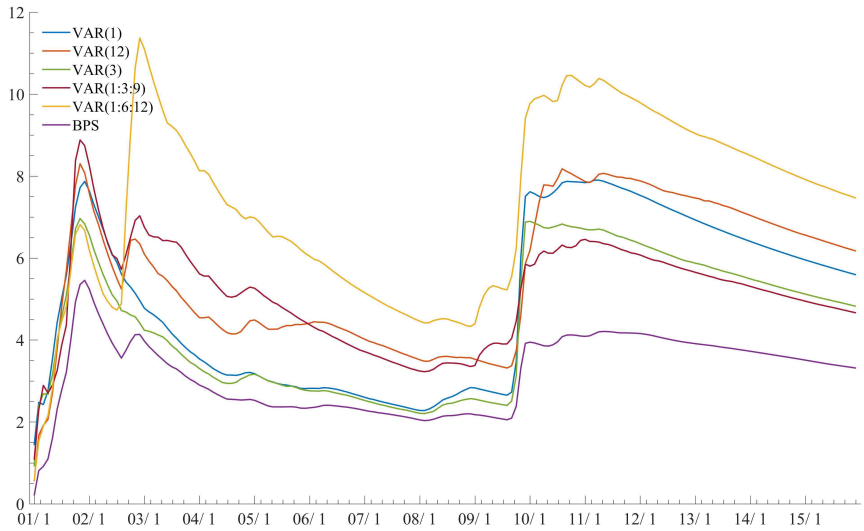

Figure C10: US macroeconomic forecasting 2001/1-2015/12: Mean squared 12-step ahead forecast errors $\text{MSFE}_{1:t}(12)$ of interest rate (r) sequentially revised at each of the $t = 1{:}180$ months.

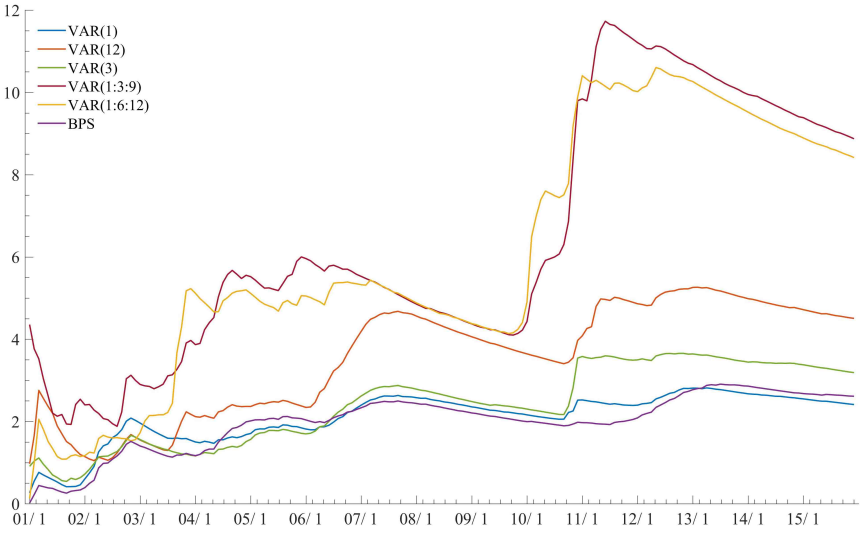

Figure C11: US macroeconomic forecasting 2001/1-2015/12: Mean squared 24-step ahead forecast errors $\text{MSFE}_{1:t}(24)$ of wage (w) sequentially revised at each of the $t = 1{:}180$ months.

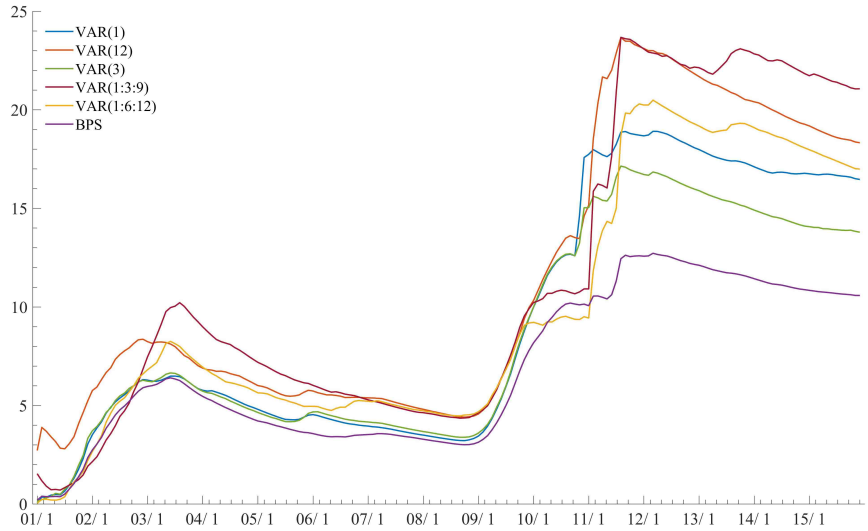

Figure C12: US macroeconomic forecasting 2001/1-2015/12: Mean squared 24-step ahead forecast errors $\text{MSFE}_{1:t}(24)$ of unemployment rate (u) sequentially revised at each of the $t = 1{:}180$ months.

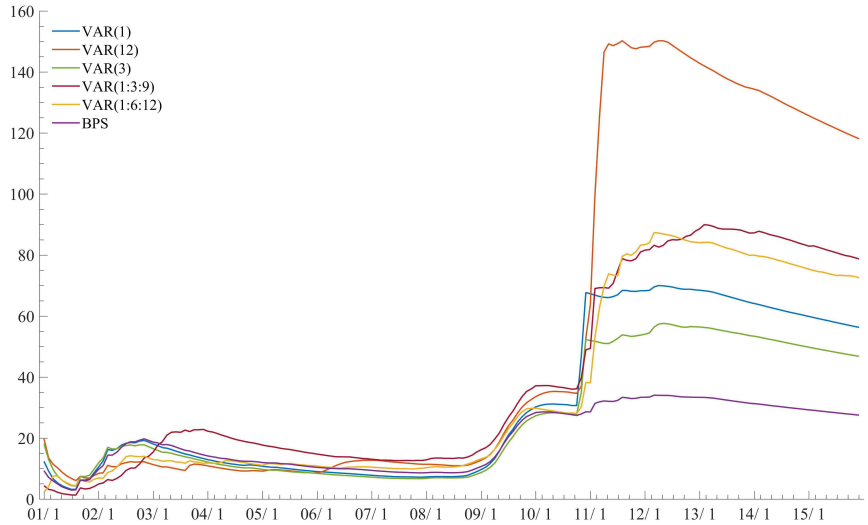

Figure C13: US macroeconomic forecasting 2001/1-2015/12: Mean squared 24-step ahead forecast errors $\text{MSFE}_{1:t}(24)$ of consumption (c) sequentially revised at each of the $t = 1{:}180$ months.

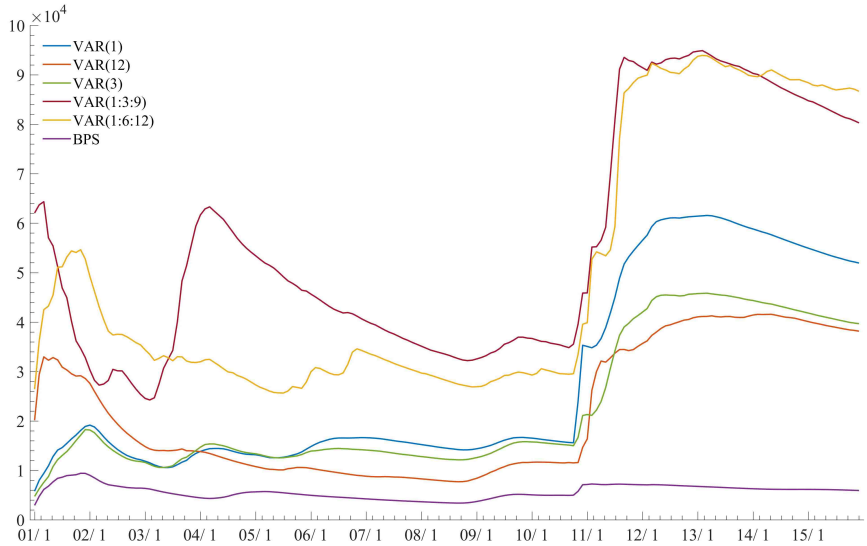

Figure C14: US macroeconomic forecasting 2001/1-2015/12: Mean squared 24-step ahead forecast errors $\text{MSFE}_{1:t}(24)$ of investment (i) sequentially revised at each of the $t = 1{:}180$ months.

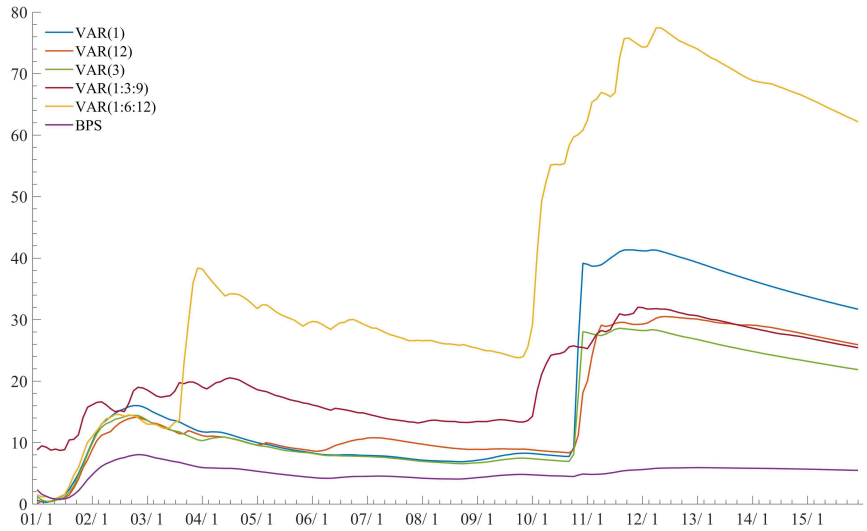

Figure C15: US macroeconomic forecasting 2001/1-2015/12: Mean squared 24-step ahead forecast errors $\text{MSFE}_{1:t}(24)$ of interest rate (r) sequentially revised at each of the $t = 1{:}180$ months.

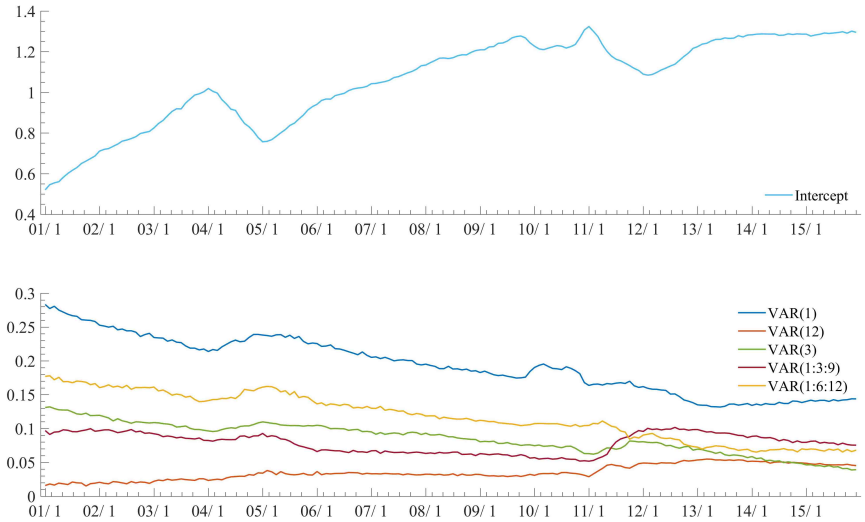

Figure C16: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for inflation (p) sequentially computed at each of the $t = 1{:}180$ months.

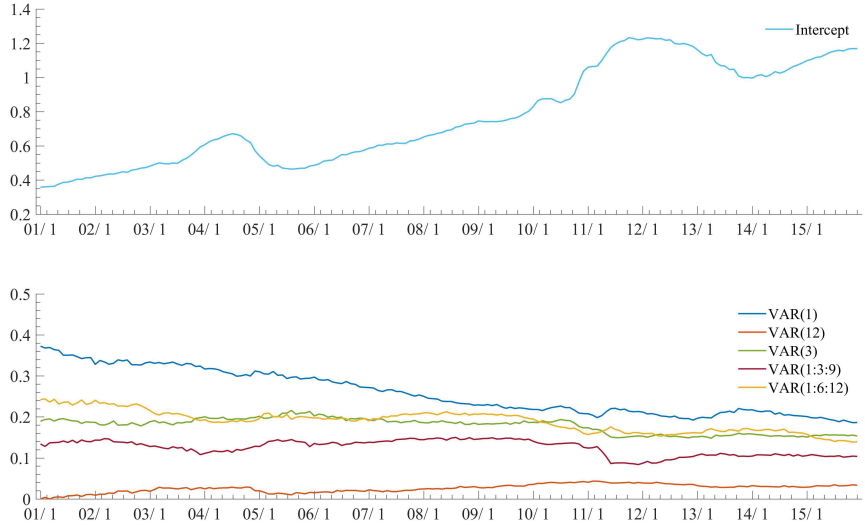

Figure C17: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for wage (w)sequentially computed at each of the $t = 1{:}180$ months.

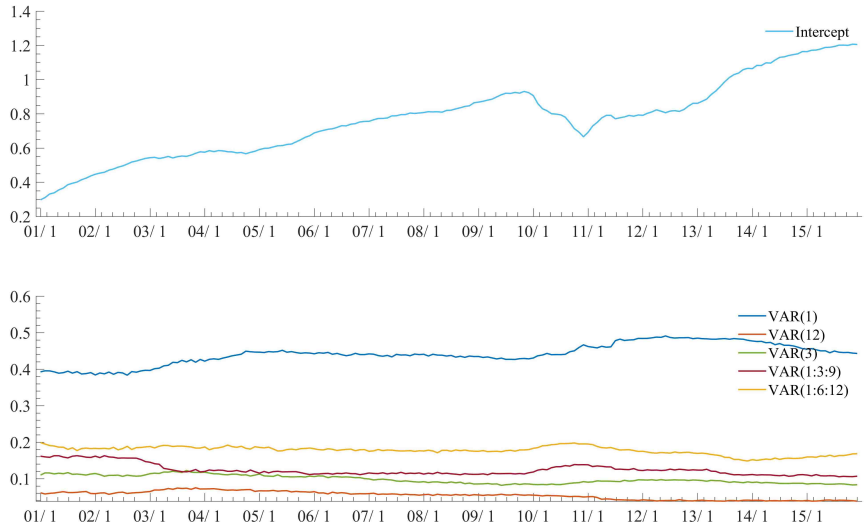

Figure C18: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for unemployment (u) sequentially computed at each of the $t = 1{:}180$ months.

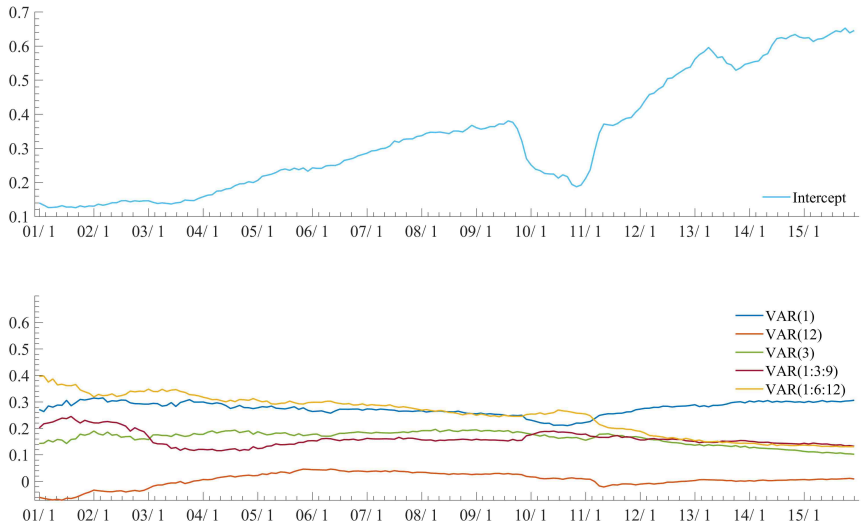

Figure C19: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for consumption (c) sequentially computed at each of the $t = 1{:}180$ months.

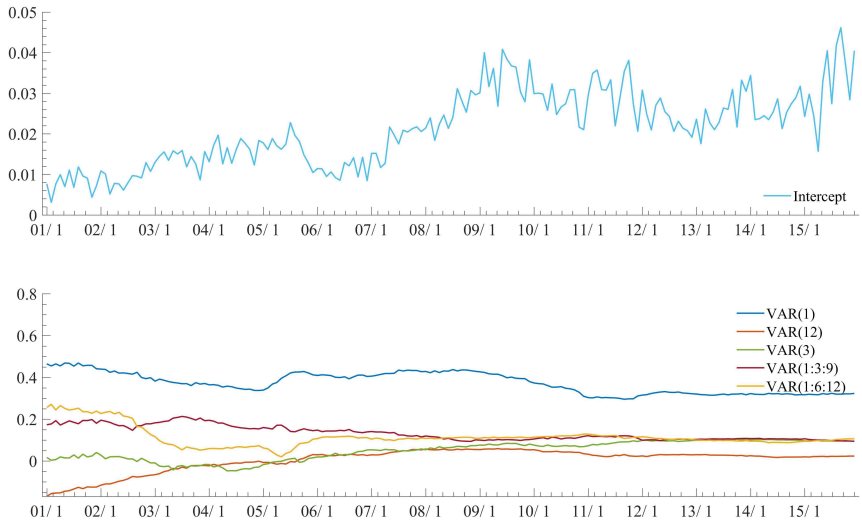

Figure C20: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for investment (i) sequentially computed at each of the $t = 1{:}180$ months.

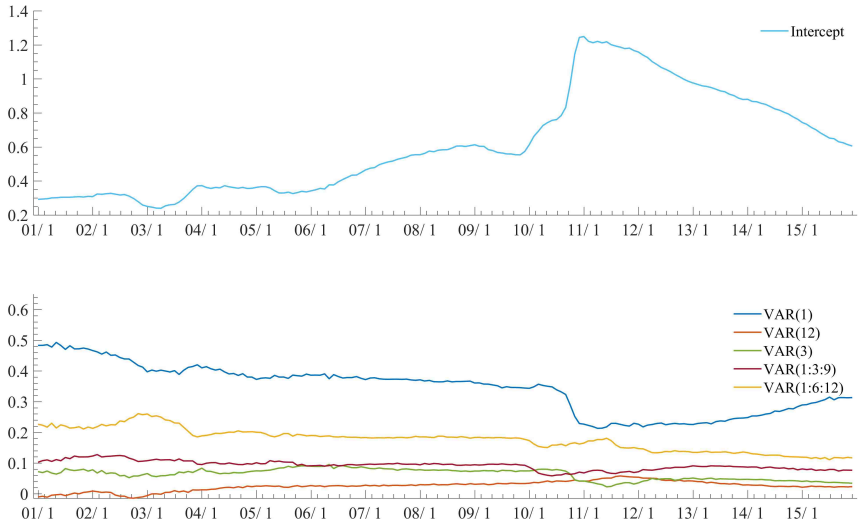

Figure C21: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(12) model coefficients for interest rate (r) sequentially computed at each of the $t = 1{:}180$ months.

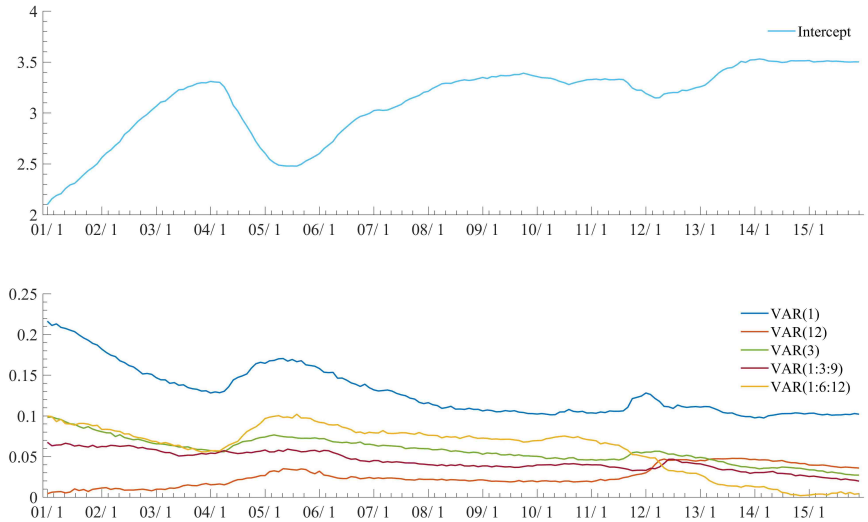

Figure C22: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for inflation (p) sequentially computed at each of the $t = 1{:}180$ months.

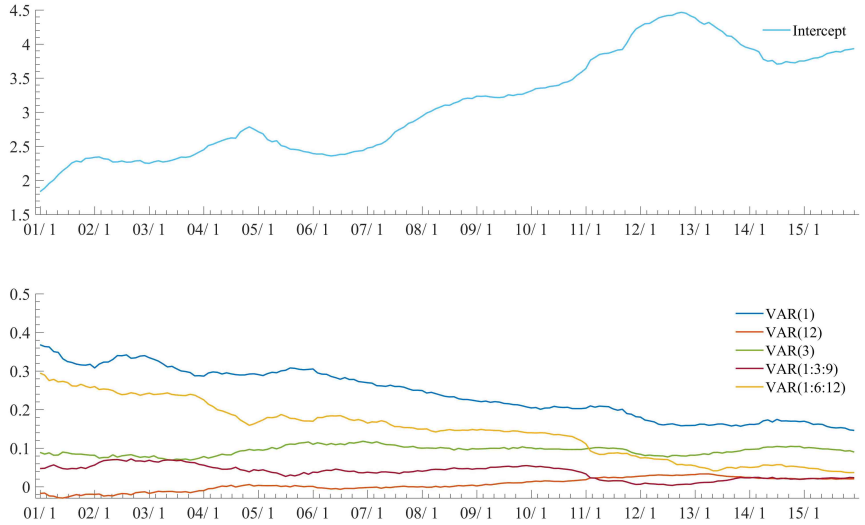

Figure C23: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for wage (w)sequentially computed at each of the $t = 1{:}180$ months.

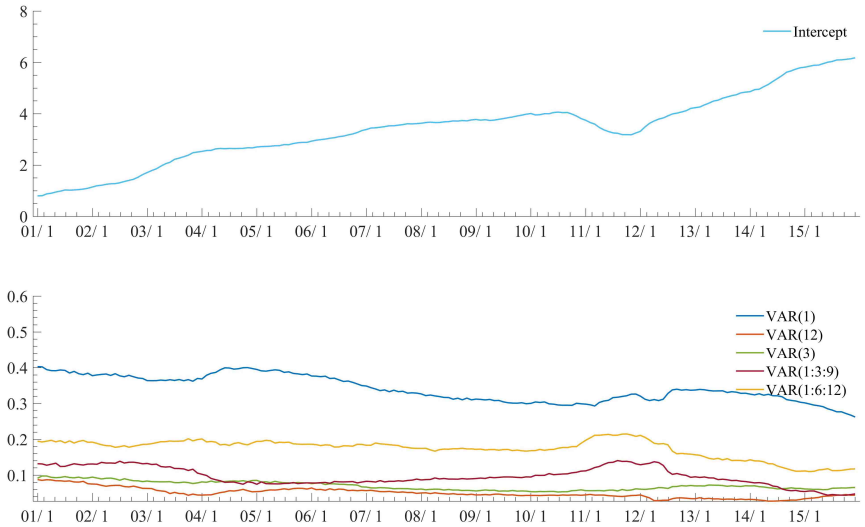

Figure C24: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for unemployment (u) sequentially computed at each of the $t = 1{:}180$ months.

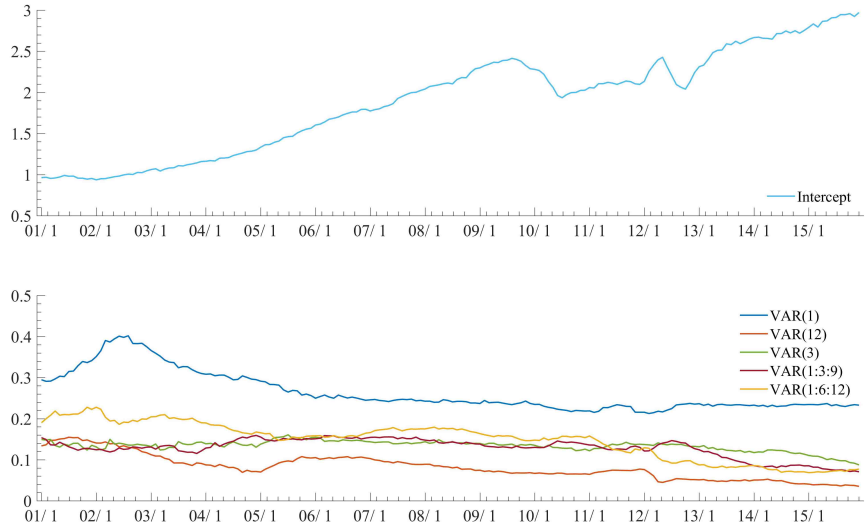

Figure C25: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for consumption (c) sequentially computed at each of the $t = 1{:}180$ months.

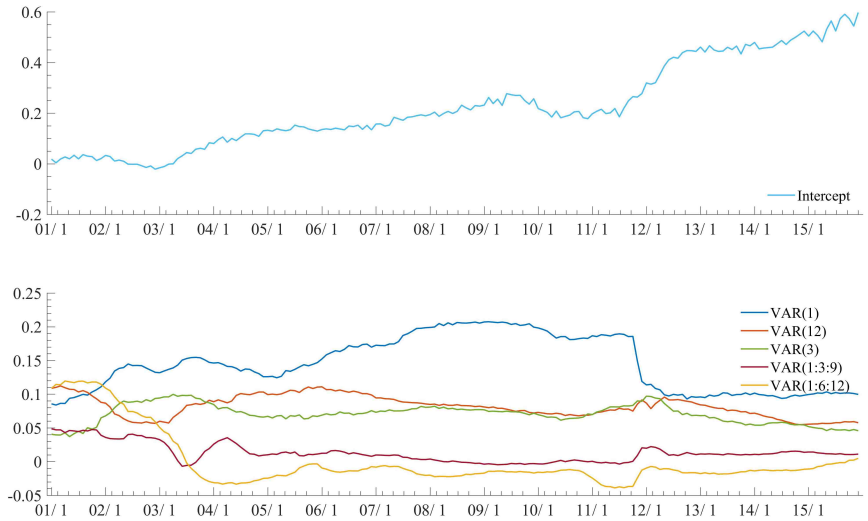

Figure C26: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for investment (i) sequentially computed at each of the $t = 1{:}180$ months.

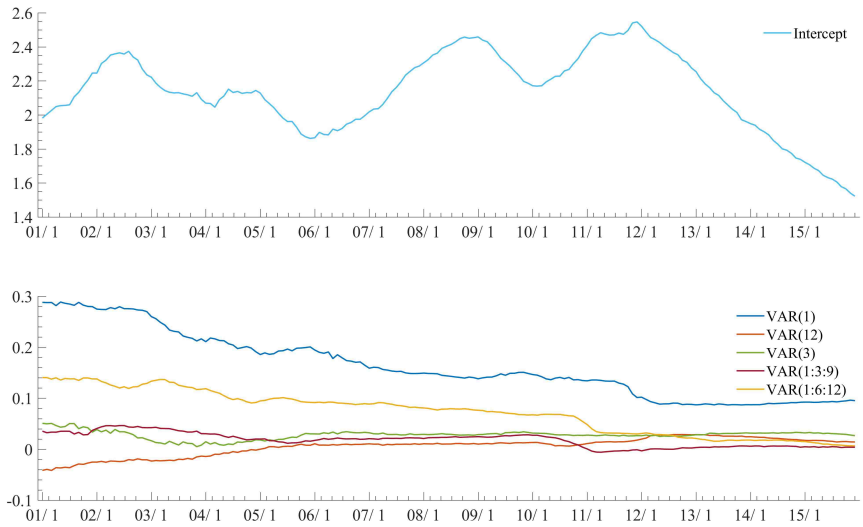

Figure C27: US macroeconomic forecasting 2001/1-2015/12: On-line posterior means of BPS(24) model coefficients for interest rate (r) sequentially computed at each of the $t = 1{:}180$ months.