Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting

Ken McAlinn[1] **Knut Are Aastveit**[2] Jouchi Nakajima[3] Mike West[4]

[1]Chicago Booth and Duke University

[2]Norges Bank and BI Norwegian Business School

[3]Bank for International Settlements

[4]Duke University

10th ECB Workshop on Forecasting Techniques:
Economic Forecasting with Large Datasets
Frankfurt am Main, June 18-19, 2018

Disclaimer: The views expressed herein are solely those of the authors and do not necessarily reflect the views of Norges Bank or the Bank for International Settlements.

Motivation

- In economic policy making, dependencies among macroeconomic time series provide fundamental insights into the state of economy.
  - Useful for improving forecasts over multiple horizons
  - Guiding policy decisions and understand their impact
- Central banks set national target interest rates based on (implicit or explicit) utility/loss considerations that weigh future outcomes of inflation and measures of the real economy.
  - Crucial to understand the (time-varying) dependencies of these measures
  - Researchers and policy makers therefore use multivariate models (e.g., VARs and DSGEs)

Motivation: Policymaking in practice

- To produce accurate and useful forecasts, policy makers routinely rely on multiple sources to produce forecasts.
  - Forecast combination (Bates and Granger (1969) and Timmermann (2006))
- To ensure appropriate normative decision making as well as reflecting increased uncertainty into the future, it has become popular, particularly for central banks, to provide probabilistic (density) forecasts.
  - See monetary policy reports of the Bank of England, Norges Bank, Swedish Riksbank, and recently also for the Federal Reserve Bank.

Motivation: Combining probabilistic forecasts

- Building on earlier work in statistics by West (1992) and West and Crosse (1992) the research interest in forecast combination has more recently focused on the construction of combinations of predictive densities
  - Combining predictive densities (Hall and Mitchell (2007), Jore et. al (2010) and Aastveit et al. (2014))
  - Optimal prediction pool: Geweke and Amisano (2011)
  - Time varying weights: Koop and Korobilis (2012)
  - Time varying weights with learning and model set incompleteness: Billio et al. (2013) and Casarin et al. (2015) Aastveit et al. (2017).
  - Bayesian predictive synthesis: McAlinn and West (2017)

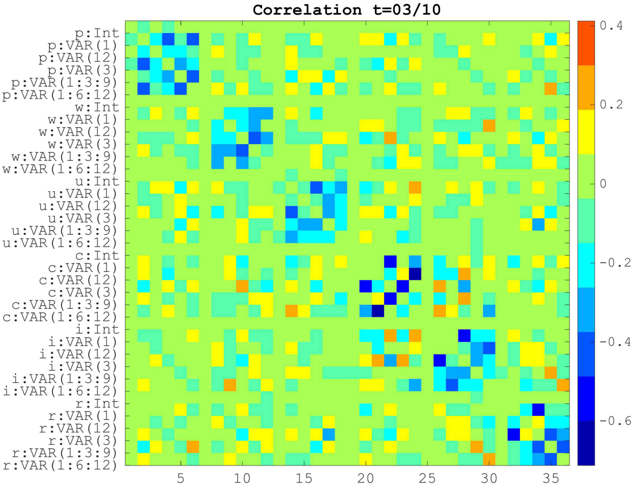
Motivation: Combining probabilistic forecasts

- But most of these papers focus on univariate forecasting...
  - Exceptions: Andersson and Karlsson (2008), Amendola and Sorti (2015) and Amisano and Geweke (2017)
    - But restrict attention to direct extensions of univariate methods, with models combined linearly using one metric for overall performance.
    - Limiting as: ignores inter-dependencies among series and that some models might be good at forecasting one series but poor in another (or poor overall).
- Economic policy makers use "ad hoc" strategies, which either rely on:
  - The policy maker's "favorite" model, or
  - Ignore inter-dependencies all together.
- Need a coherent methodology that gives policy makers flexibility in incorporating multivariate density forecasts from multiple sources.

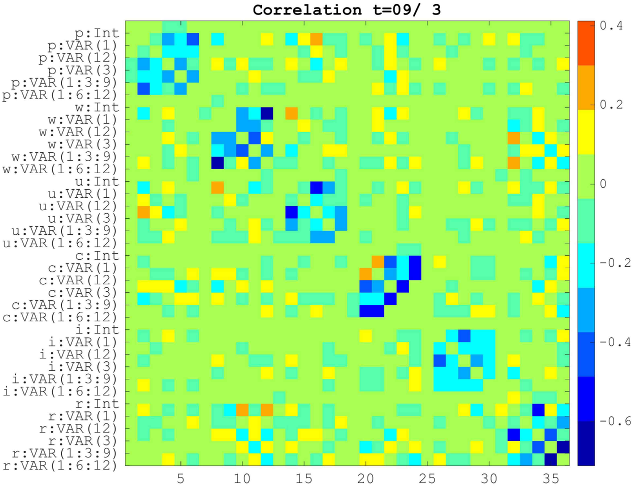
Contribution of this paper

- We develop the methodology of Bayesian predictive synthesis (BPS) models for multivariate time series forecasting.
  - Extend the recently introduced foundational framework of BPS in McAlinn and West (2017) to the multivariate setting
  - BPS is a coherent Bayesian framework for evaluation, calibration, comparison, and combination of multiple forecast densities.
  - As a multivariate extension we use a flexible dynamic latent factor model with seemingly-unrelated regression structure (DFSUR model)
- In an application using various TVP-VARs for forecasting six monthly US macroeconomic time series for 1-, 12-, and 24-month ahead we find that our multivariate BPS:
  - Improve forecast accuracy for each of several multiple macroeconomic series together at multiple horizons
    - Can adapt to time-varying biases and miscalibration of multiple models or forecasters
    - Adapt and account for patterns of time-varying relationships and dependencies among sets of models or forecasters

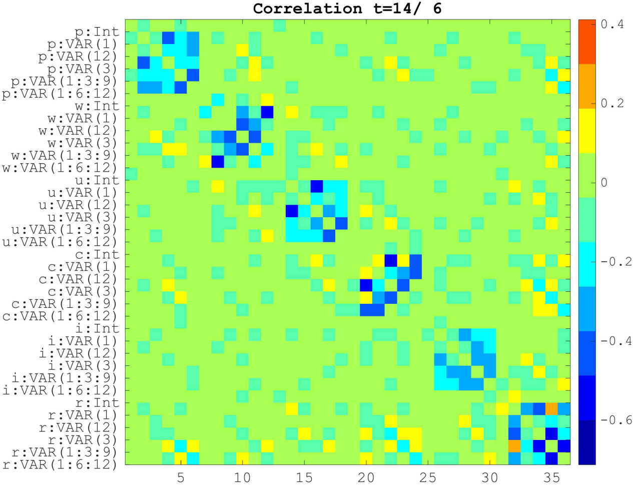
On-line posterior correlation of BPS model coefficients at 2003/10



On-line posterior correlation of BPS model coefficients at 2009/03



On-line posterior correlation of BPS model coefficients at 2014/06



BPS background

- A Bayesian decision maker $\mathcal{D}$ receive forecast distributions for $\mathbf{y}$ from each of J *agents*
- Agent $\mathcal{A}_j$ provides a probability density function $h_j(\mathbf{x}_j) = p(\mathbf{y}|\mathcal{A}_j)$.
- The information set $\mathcal{H} = \{h_1(\cdot), \ldots, h_J(\cdot)\}$ now available to $\mathcal{D}$.
- $\mathcal{D}$ will then use the information set $\mathcal{H}$ to predict $\mathbf{y}$ using the implied posterior $p(\mathbf{y}|\mathcal{H})$ from a full Bayesian prior-to-posterior analysis.
- West (1992) showed that for a subset of all Bayesian models $\mathcal{D}$'s posterior has the mathematical form

$$p(\mathbf{y}|\mathcal{H}) = \int_{\mathbf{X}} \alpha(\mathbf{y}|\mathbf{X}) \prod_{j=1:J} h_j(\mathbf{x}_j) d\mathbf{x}_j \tag{1}$$

where each $\mathbf{x}_j$ is a latent $q\times 1-$dimensional vector, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_J]'$ collects these latent vectors in a $J\times q-$dimensional matrix, and $\alpha(\mathbf{y}|\mathbf{X})$ is a conditional p.d.f. for $\mathbf{y}$ given $\mathbf{X}$.

BPS background - interpretation

- For $\mathcal{D}$ there *must* exist latent factors $\mathbf{x}_j$ potentially related to $\mathbf{y}$ and such that agent $\mathcal{A}_j$'s forecast density is that of $\mathbf{x}_j$.
  - Refer to $\mathbf{x}_j$ as *latent agent states*
- *Conditional on learning* $\mathcal{H}$, the $\mathcal{D}$ regards the latent factors as conditionally independent with $\mathbf{x}_j \sim h_j(\mathbf{x}_j)$.
  - This does not imply that $\mathcal{D}$ regards the forecasts as independent, since under her prior the $\mathbf{h}_j(\cdot)$ are uncertain and likely highly inter-dependent.
- $\alpha(\mathbf{y}|\mathbf{X})$ is $\mathcal{D}$'s regression model relating the $\mathbf{x}_j$ as a collective to the $\mathbf{y}$.
  - The key element $\alpha(\mathbf{y}|\mathbf{X})$ is how $\mathcal{D}$ expresses her views of dependencies.
  - We refer to $\alpha(\mathbf{y}|\mathbf{X})$ as the BPS *synthesis function*.

BPS: Dynamic Sequential Setting

- $\mathcal{D}$ receives forecast densities from each agent sequentially over time.
- At time $t-1$, $\mathcal{D}$ receives current forecast densities $\mathcal{H}_t = \{h_{t1}(\mathbf{x}_t), \ldots, h_{tJ}(\mathbf{x}_t)\}$ from the set of agents and aims to forecast $\mathbf{y}_t$.
  - The full information set used by $\mathcal{D}$ at time t is thus $\{\mathbf{y}_{1:t-1},\ \mathcal{H}_{1:t}\}$.
- As $\mathcal{D}$ observes more information, her views of the agent biases and calibration characteristics, as well as of inter-dependencies among agents are repeatedly updated.
- $\mathcal{D}$ has a time $t-1$ distribution for $\mathbf{y}_t$ as

$$p(\mathbf{y}_t|\mathbf{\Phi}_t, \mathbf{y}_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(\mathbf{y}_t|\mathbf{\Phi}_t, \mathcal{H}_t) = \int \alpha_t(\mathbf{y}_t|\mathbf{X}_t, \mathbf{\Phi}_t) \prod_{j=1:J} h_{tj}(\mathbf{x}_{tj}) d\mathbf{x}_{tj} \tag{2}$$

where $\mathbf{X}_t = [\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tJ}]'$ is a $J \times q-$dimensional matrix of latent agent states at time t, the conditional p.d.f. $\alpha_t(\mathbf{y}_t|\mathbf{X}_t, \mathbf{\Phi}_t)$ is $\mathcal{D}$'s synthesis p.d.f. for $\mathbf{y}_t$ given $\mathbf{X}_t$, and involves time-varying parameters $\mathbf{\Phi}_t$ for which $\mathcal{D}$ has current beliefs represented in terms of her (time $t-1$) posterior $p(\mathbf{\Phi}_t|\mathbf{y}_{1:t-1}, \mathcal{H}_{1:t-1})$.

BPS synthesis function: $\alpha_t(\mathbf{y}_t|\mathbf{X}_t, \mathbf{\Phi}_t)$

- $\mathbf{x}_{tj}$ are realizations of inherent dynamic latent factors – the *latent agent states* at time t
- Synthesis is achieved by relating these latent factor processes to the time series $\mathbf{y}_t$ via models of the time-varying synthesis function $\alpha_t(\mathbf{y}_t|\mathbf{X}_t, \mathbf{\Phi}_t)$.
- Would like flexibility for $\mathcal{D}$ to specify and incorporate information on:
  - Agent-specific biases, calibration
  - Relative expertise/accuracy
  - Agent inter-dependencies
  - Time-variation . . . in all the above
- Our choice:
  - Dynamic latent factor model with seemingly-unrelated regression structure (DFSUR model)

Multivariate Latent Factor Dynamic Models

Consider a dynamic multivariate BPS synthesis function

$$\alpha_t(\mathbf{y}_t|\mathbf{X}_t, \Phi_t) = N(\mathbf{y}_t|\mathbf{F}_t\theta_t, \mathbf{V}_t) \tag{3}$$

with

$$\mathbf{F}_t = \begin{pmatrix} 1 & \mathbf{f}'_{t1} & 0 & \mathbf{0}' & \cdots & \cdots & 0 & \mathbf{0}' \\ 0 & \mathbf{0}' & 1 & \mathbf{f}'_{t2} & & & & \vdots \\ \vdots & & & & \ddots & & & \vdots \\ 0 & \mathbf{0}' & \cdots & \cdots & \cdots & \cdots & 1 & \mathbf{f}'_{tq} \end{pmatrix} \quad \text{and} \quad \theta_t = \begin{pmatrix} \theta_{t1} \\ \theta_{t2} \\ \vdots \\ \theta_{tq} \end{pmatrix} \tag{4}$$

- For each series $r = 1{:}q$, the $J \times 1$–vector $\mathbf{f}_{tr} = (x_{tr1}, x_{tr2}, ..., x_{trJ})'$ is a realization of the set of J latent agents states for series r
- $\theta_{tr} = (1, \theta_{tr1}, \theta_{tr2}, ..., , \theta_{trJ})'$ contains an intercept and coefficients representing time-varying bias/calibration weights of the J latent agent states for series r

Multivariate Latent Factor Dynamic Models

Modeling time evolution of the parameter processes $\mathbf{\Phi}_t = (\theta_t, \mathbf{V}_t)$ is needed to complete model specification:

$$\mathbf{y}_t = \mathbf{F}_t \theta_t + \nu_t, \quad \nu_t \sim N(\mathbf{0}, \mathbf{V}_t), \tag{5}$$

$$\theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(\mathbf{0}, \mathbf{W}_t) \tag{6}$$

- θ_t evolves in time according to a linear/normal random walk with innovations variance matrix $\mathbf{W}_t$ at time t
  - $\mathbf{W}_t$ is defined via a standard single discount factor specification (see Prado and West (2010))
- $\mathbf{V}_t$ is the residual variance in predicting $\mathbf{y}_t$ based on past information and the set of agent forecast distributions.
  - $\mathbf{V}_t$ follows a standard inverse Wishart random walk volatility model (also based on discounting)

Multivariate Latent Factor Dynamic Models

- We now have a class of dynamic, multivariate latent factor models in which latent factors are realized as draws from the set of agent densities $h_{tj}(\cdot)$, becoming available to $\mathcal{D}$ at $t-1$ for forecasting $\mathbf{y}_t$.
- Coupled with eqns. (5,6), we have the time t *prior* for the latent states–conditional on $\mathcal{H}_{1:t}$, as

$$p(\mathbf{X}_t|\mathbf{\Phi}_t, \mathbf{Y}_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(\mathbf{X}_t|\mathcal{H}_t) = \prod_{j=1:J} h_{tj}(\mathbf{x}_{tj}) \quad (7)$$

  with $\mathbf{X}_t, \mathbf{X}_s$ conditionally independent for all $t \neq s$.
- The *conditional* independence of the $\mathbf{x}_{tj}$ given the $h_{tj}(\cdot)$ must not be confused with the $\mathcal{D}$'s modeling and estimation of the dependencies among agents.
- This dependence is central and integral, and is reflected through the effective dynamic parameters $\mathbf{\Phi}_t = (\theta_t, \mathbf{V}_t)$.

Posterior computations via MCMC

- Three-component block Gibbs sampler for the latent agent states, dynamic coefficient parameters, and dynamic volatility parameters.
  1. Conditional on the agent states and residual volatility, draw new dynamic coefficient parameters from $p(\theta_{1:t}|\mathbf{X}_{1:t}, \mathbf{V}_{1:t}, \mathbf{y}_{1:t})$.
     - Sampled using an extension of the traditional forward filtering, backward sampling (FFBS) algorithm (Prado and West 2010))
  2. Draw new dynamic volatility matrices $\mathbf{V}_t$ from the full joint conditional posterior $p(\mathbf{V}_{1:t}|\mathbf{X}_{1:t}, \theta_{1:t}, \mathbf{y}_{1:t})$– conditional on the agent states and dynamic coefficient parameters.
     - Employs the standard FFBS algorithm for inverse Wishart discount volatility models (Prado and West 2010))
  3. Conditional on values of dynamic parameters $\mathbf{\Phi}_{1:t} = (\theta_{1:t}, \mathbf{V}_{1:t})$, draw new agent states from $p(\mathbf{X}_{1:t}|\mathbf{\Phi}_{1:t}, \mathbf{y}_{1:t}, \mathcal{H}_{1:t})$.
     - The $\mathbf{X}_t$ are conditionally independent over time t in this conditional distribution, with time t conditionals $p(\mathbf{X}_t|\mathbf{\Phi}_t, \mathbf{y}_t, \mathcal{H}_t) \propto N(\mathbf{y}_t|\mathbf{F}_t\theta_t, \mathbf{V}_t)\prod_{j=1:J} h_{tj}(\mathbf{x}_{tj})$.

Computing the forecasts

At time t we forecast 1-step ahead by generating "synthetic futures" from the BPS model, as follows.

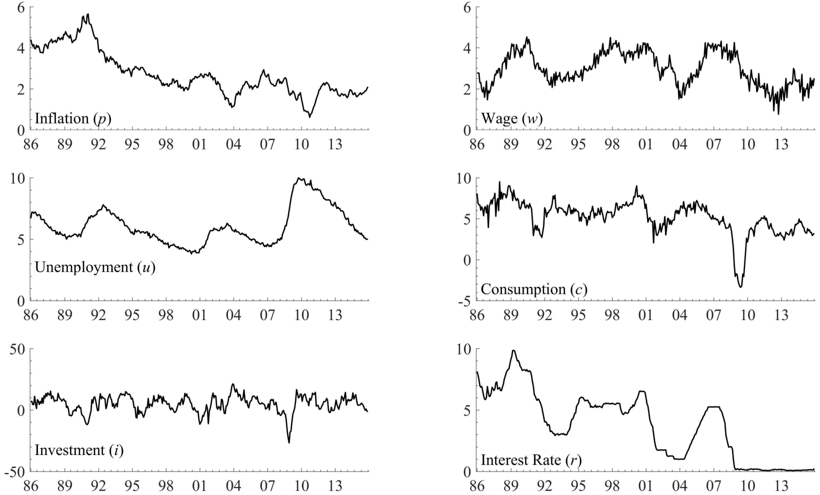
1. Draw $\mathbf{V}_{t+1}$ from its discount volatility evolution model, and then θ_{t+1} conditional on $\theta_t, \mathbf{V}_{t+1}$ from the evolution model eqn. (6)
   $\Rightarrow$ Gives us a draw $\mathbf{\Phi}_{t+1} = \{\theta_{t+1}, \mathbf{V}_{t+1}\}$ from $p(\mathbf{\Phi}_{t+1}|\mathbf{y}_{1:t}, \mathcal{H}_{1:t})$
2. Draw $\mathbf{X}_{t+1}$ via independent sampling of the $h_{t+1,j}(\mathbf{x}_{t+1,j})$, $(j = 1{:}J)$.
3. Bring these samples together and draw a synthetic 1-step outcome $\mathbf{y}_{t+1}$ from the conditional normal of eqn. (5) given these sampled parameters and agent states.

Repeating this generates a random Monte Carlo sample from the 1-step ahead forecast distribution for time $t+1$.

Empirical exercise

- Forecast 6 macroeconomic variables for the U.S.
  - Variables: annual inflation rate (p), wage (w), unemployment rate (u), consumption (c), investment (i), and short-term nominal interest rate (r)
  - Forecast evaluation: MSFE and LPDR
  - Data sample: 1986/1 to 2015/12
  - Training period VARs: 1986/1 to 1993/6, Training period BPS: 1993/7 to 2000/12
  - Evaluation period: 2001/1 to 2015/12,
  - Forecast horizons: $h = 1, 12, 24$
  - Consider forecast from $J = 5$ agents using the following TVP-VAR models:
    - M1- VAR(1); M2- VAR(12); M3- VAR(3); M4- VAR(1:3:9); M5- VAR(1:6:12)
  - Directly synthesize $k - step$ models

Data



Inflation (p)
Wage (w)
Unemployment (u)
Consumption (c)
Investment (i)
Interest Rate (r)
86
89
92
95
98
01
04
07
10
13

Forecasting results - 1 step ahead

1-step	Infl	%	Wage	%	Unemp	%
			MSFE$_{1:T}$			
VAR(1)	0.0141	−8.22	0.1444	−35.91	0.0206	0.74
VAR(12)	0.0160	−22.93	0.1110	−4.44	0.0230	−10.73
VAR(3)	0.0147	−13.24	0.1105	−3.96	0.0219	−5.67
VAR(1:3:9)	0.0135	−3.76	0.1198	−12.77	0.0222	−7.18
VAR(1:6:12)	0.0137	−5.14	0.1449	−36.40	0.0215	−3.70
BMA	0.0146	−12.20	0.1111	−4.53	0.0218	−5.26
BPS	0.0130	—	0.1063	—	0.0207	—

1-step	Cons	%	Invest	%	Interest	%
			MSFE$_{1:T}$			
VAR(1)	0.3908	−3.50	13.2183	−2.99	0.0275	−35.07
VAR(12)	0.4697	−24.41	15.3571	−19.65	0.0246	−20.92
VAR(3)	0.3982	−5.48	13.3210	−3.79	0.0211	−3.74
VAR(1:3:9)	0.4049	−7.25	13.8918	−8.24	0.0204	−0.55
VAR(1:6:12)	0.3889	−3.02	13.4301	−4.64	0.0228	−12.02
BMA	0.3971	−5.18	13.2145	−2.96	0.0215	−5.80
BPS	0.3775	—	12.8346	—	0.0203	—

1-step	LPDR$_{1:T}$
VAR(1)	−77.25
VAR(12)	−103.82
VAR(3)	−31.00
VAR(1:3:9)	−34.22
VAR(1:6:12)	−52.69
BMA	−32.48
BPS	—

Forecasting results - 12 step ahead

			MSFE$_{1:T}$			
12-step	Infl	%	Wage	%	Unemp	%
VAR(1)	0.5317	−143.15	0.4453	19.50	1.2028	−10.66
VAR(12)	0.4272	−95.35	0.7750	−40.12	1.6918	−55.65
VAR(3)	0.5789	−164.74	0.5215	5.72	1.1788	−8.45
VAR(1:3:9)	0.4541	−107.69	1.1207	−102.62	1.6353	−50.46
VAR(1:6:12)	0.5342	−144.30	0.8934	−61.52	1.3585	−24.99
BPS(12)	0.2187	—	0.5531	—	1.0869	—

			MSFE$_{1:T}$			
12-step	Cons	%	Invest	%	Interest	%
VAR(1)	7.2471	−23.21	7067.67	−65.55	5.5916	−68.74
VAR(12)	18.4145	−213.07	8824.02	−106.68	6.1707	−86.22
VAR(3)	7.3142	−24.35	6378.42	−49.40	4.8222	−45.52
VAR(1:3:9)	10.3823	−76.51	9111.99	−113.43	4.6622	−40.69
VAR(1:6:12)	10.1116	−71.91	10013.45	−134.54	7.4612	−125.16
BPS(12)	5.8818	—	4269.33	—	3.3137	—

12-step	LPDR$_{1:T}$
VAR(1)	−119.05
VAR(12)	−535.09
VAR(3)	−366.85
VAR(1:3:9)	−463.46
VAR(1:6:12)	−361.20
BPS(12)	—

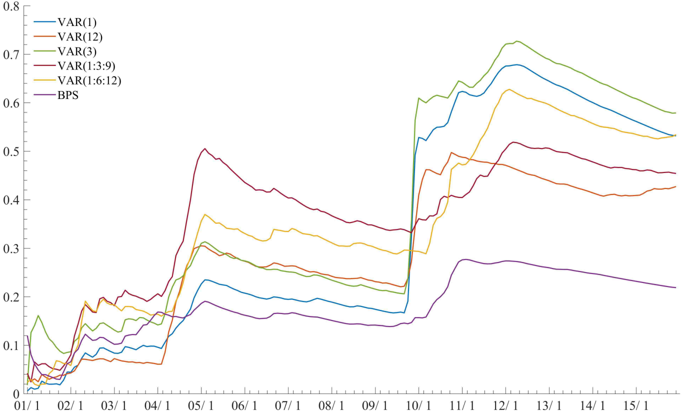
Forecasting results - 24 step ahead

			MSFE$_{1:T}$			
24-step	Infl	%	Wage	%	Unemp	%
VAR(1)	3.9536	−331.10	2.4117	7.71	16.46	−55.68
VAR(12)	2.7373	−198.47	4.5054	−72.41	18.32	−73.28
VAR(3)	3.8504	−319.85	3.1877	−21.98	13.78	−30.35
VAR(1:3:9)	4.8627	−430.23	8.8723	−239.52	21.06	−99.17
VAR(1:6:12)	4.4141	−381.32	8.4162	−222.06	16.99	−60.65
BPS(24)	0.9171	–	2.6132	–	10.58	–

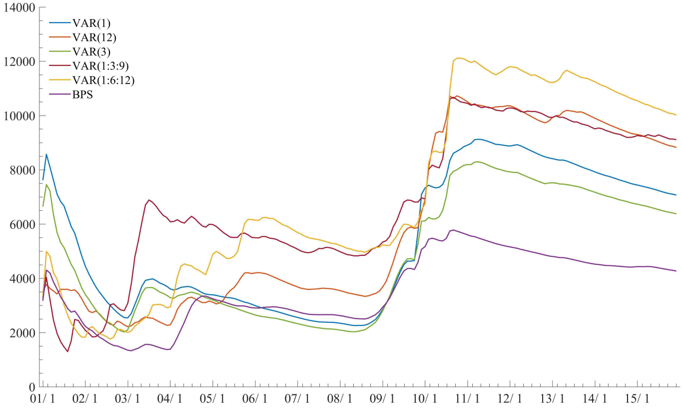
			MSFE$_{1:T}$			
24-step	Cons	%	Invest	%	Interest	%
VAR(1)	56.27	−104.54	51937	−776.23	31.68	−480.56
VAR(12)	118.09	−329.23	38151	−543.65	25.89	−374.58
VAR(3)	46.80	−70.09	39671	−569.30	21.84	−300.31
VAR(1:3:9)	78.73	−186.15	80278	−1254.37	25.41	−365.71
VAR(1:6:12)	72.54	−163.67	86671	−1362.23	62.16	−1039.32
BPS(24)	27.51	–	5927	–	5.46	–

24-step	LPDR$_{1:T}$
VAR(1)	−445.81
VAR(12)	−489.98
VAR(3)	−462.48
VAR(1:3:9)	−808.31
VAR(1:6:12)	−804.49
BPS(24)	–

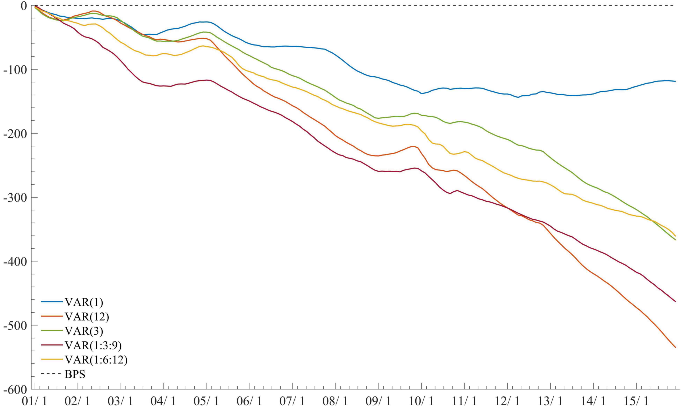
MSFE - 12 step ahead, inflation



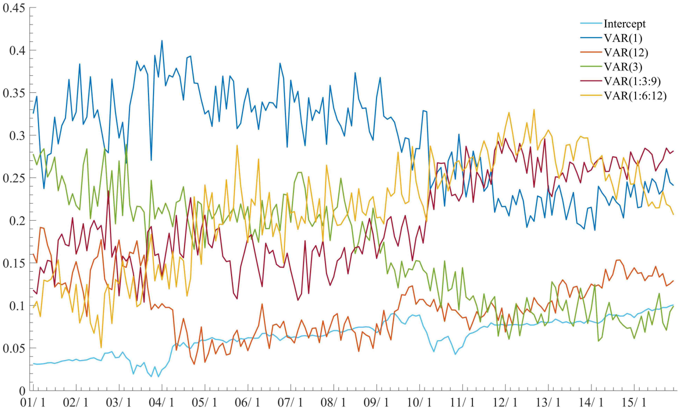
MSFE - 12 step ahead, investment



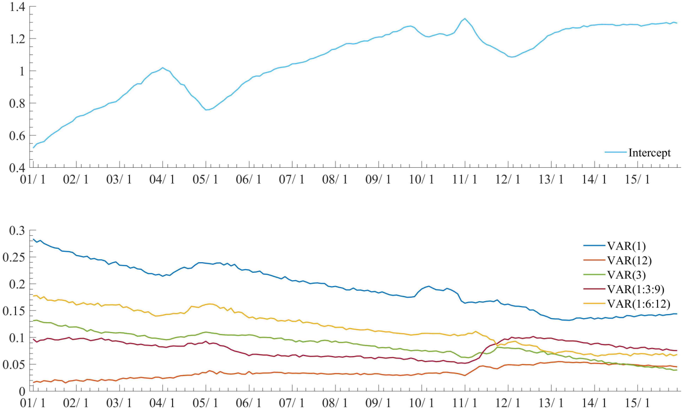
Log Predictive Density Ratios - 12 step ahead



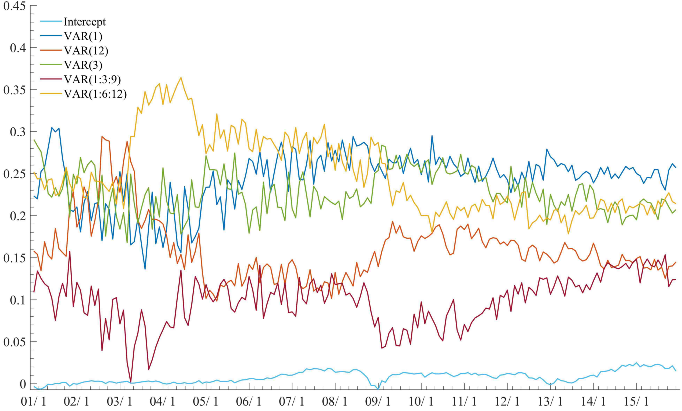
On-line posterior means of BPS(1) model, inflation



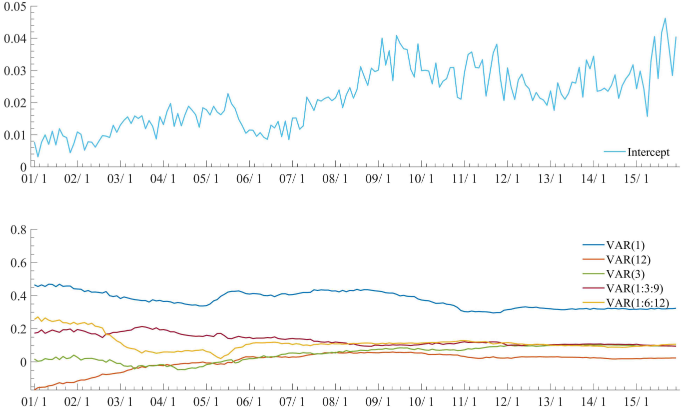
On-line posterior means of BPS(12) model, inflation



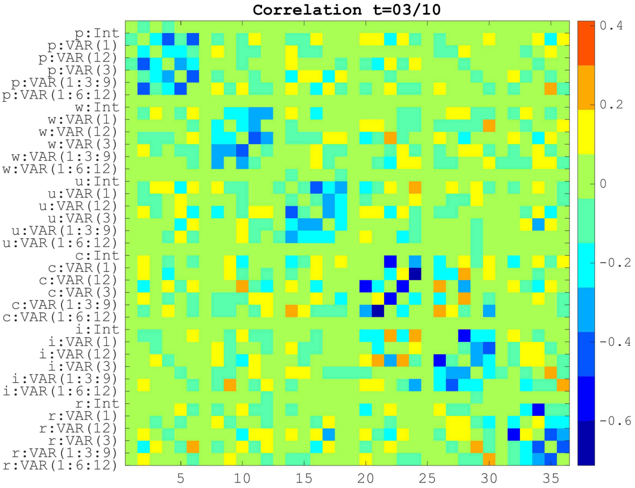
On-line posterior means of BPS(1) model, investment



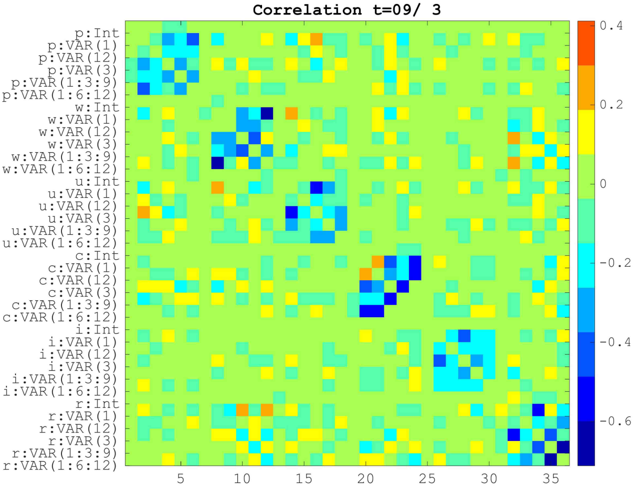
On-line posterior means of BPS(12) model, investment



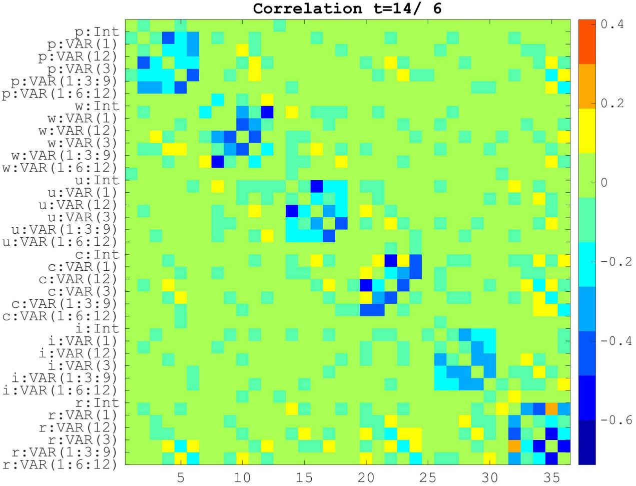
On-line posterior correlation of BPS model coefficients at 2003/10



On-line posterior correlation of BPS model coefficients at 2009/03



On-line posterior correlation of BPS model coefficients at 2014/06



Conclusion

- Our extensions and development of multivariate BPS define a theoretically and conceptually sound framework to compare and synthesize multivariate density forecasts in a dynamic context.
- The approach enables decision makers to dynamically calibrate, learn, and update predictions based on ranges of forecasts from sets of models, as well as from more subjective sources such as individual forecasters or agencies.
- In an application using various TVP-VARs for forecasting six monthly US macroeconomic time series for 1-, 12-, and 24-month ahead we show that our multivariate BPS:
  - Can adapt to time-varying biases and miscalibration of multiple models or forecasters
  - Adapt and account for patterns of time-varying relationships and dependencies among sets of models or forecasters,
  - Improve forecast accuracy– in some cases, most substantially– for each of several multiple macroeconomic series together, at multiple horizons.