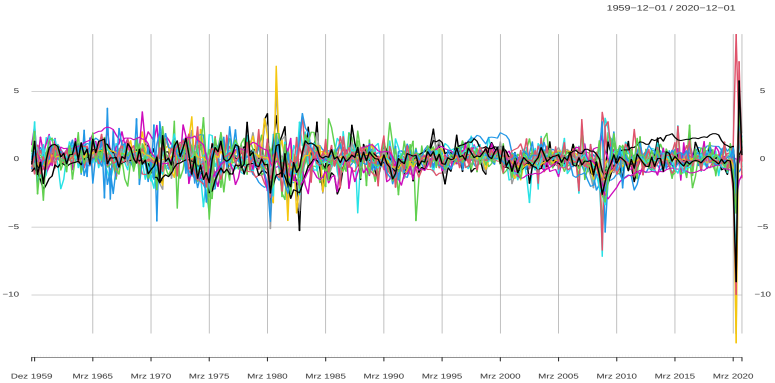
Forecasting with large Bayesian Vectorautoregressions

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ
FAKULTÄT FÜR TECHNISCHE WISSENSCHAFTEN
Institut für Statistik



zk35.org

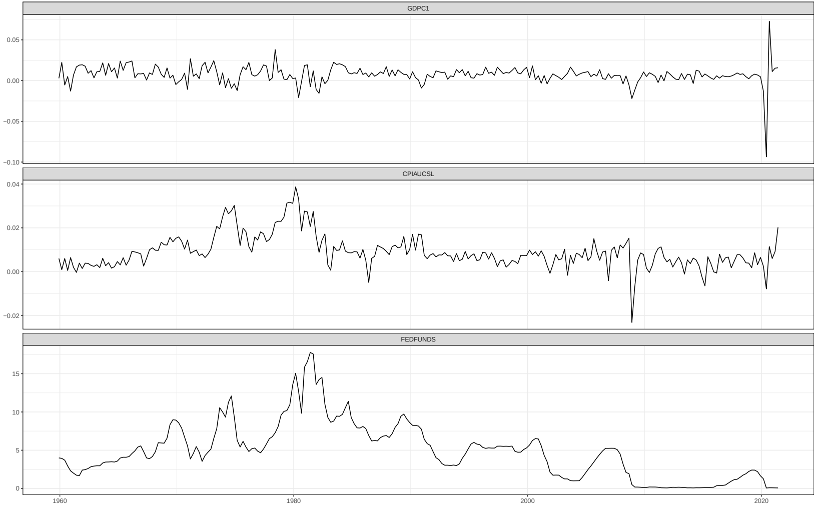
FWF Der Wissenschaftsfonds.
fwf.ac.at

Luis Gruber
Department of Statistics,
University of Klagenfurt

Gregor Kastner
Department of Statistics,
University of Klagenfurt

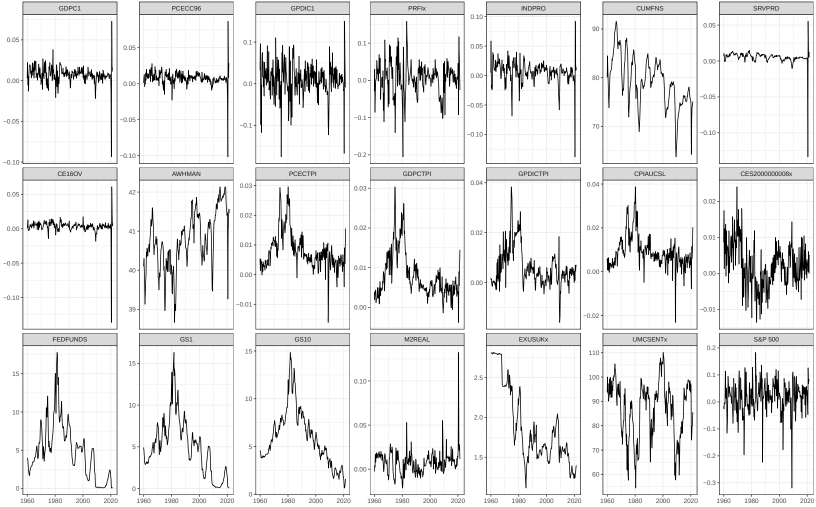
October 14, 2021

Motivation



GDPC1
0.05
0.00
−0.05
−0.10
CPIAUCSL
0.04
0.02
0.00
−0.02
FEDFUNDS
15
10
5
0
1960
1980
2000
2020

Motivation



GDPC1
PCECC96
GPDIC1
PRFIx
INDPRO
CUMFNS
SRVPRD
CE16OV
AWHMAN
PCECTPI
GDPCTPI
GPDICTPI
CPIAUCSL
CES2000000008x
FEDFUNDS
GS1
GS10
M2REAL
EXUSUKx
UMCSENTx
S&P 500

Outline

Econometric Framework & Bayesian Inference

Priors

Empirical Illustration

VAR(p): A concrete example . . .

. . . for a 3-dimensional VAR(p):

- GDP: Gross Domestic Product,
- CPI: Consumer Price Index,
- FFR: Federal Funds Rate.

$$
\begin{aligned}
GDP_t &= \phi_{11} GDP_{t-1} + \phi_{12} CPI_{t-1} + \phi_{13} FFR_{t-1} + \cdots + \phi_{1(p\times 1)} GDP_{t-p} + \phi_{1(p\times 2)} CPI_{t-p} + \phi_{1(p\times 3)} FFR_{t-p} + \epsilon_{1t} \\
CPI_t &= \phi_{21} GDP_{t-1} + \phi_{22} CPI_{t-1} + \phi_{23} FFR_{t-1} + \cdots + \phi_{2(p\times 1)} GDP_{t-p} + \phi_{2(p\times 2)} CPI_{t-p} + \phi_{2(p\times 3)} FFR_{t-p} + \epsilon_{2t} \\
FFR_t &= \phi_{31} GDP_{t-1} + \phi_{32} CPI_{t-1} + \phi_{33} FFR_{t-1} + \cdots + \phi_{3(p\times 1)} GDP_{t-p} + \phi_{3(p\times 2)} CPI_{t-p} + \phi_{3(p\times 3)} FFR_{t-p} + \epsilon_{3t}
\end{aligned}
$$

where blue represents *own-lags* and brown represents *cross-lags*.

VAR(p) for M time series

$$\boldsymbol{Y}_t = \sum_{k=1}^{p} \boldsymbol{\Phi}_k \boldsymbol{Y}_{t-k} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

where $\boldsymbol{Y}_t$ and $\boldsymbol{\epsilon}_t$ are M-dimensional vectors, and $\boldsymbol{\Phi}^k$ and $\boldsymbol{\Sigma}_t$ are $M \times M$ matrices.

In Matrix form: Let $K = Mp$. Define

- a $K \times 1$ vector of predictors $\boldsymbol{X}_t = (\boldsymbol{Y}'_{t-1}, \ldots, \boldsymbol{Y}'_{t-p})'$, and
- a $K \times M$ matrix of coefficients $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)'$. Then

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Phi} + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T)'$ and $\boldsymbol{E} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)'$ are $T \times M$ matrices, and $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)'$ is a $T \times K$ matrix.

Bayesian Approach

Let ...

- $\boldsymbol{\theta}$ be the unknown parameters and latent variables and
- $\boldsymbol{Y}$ the observed data.

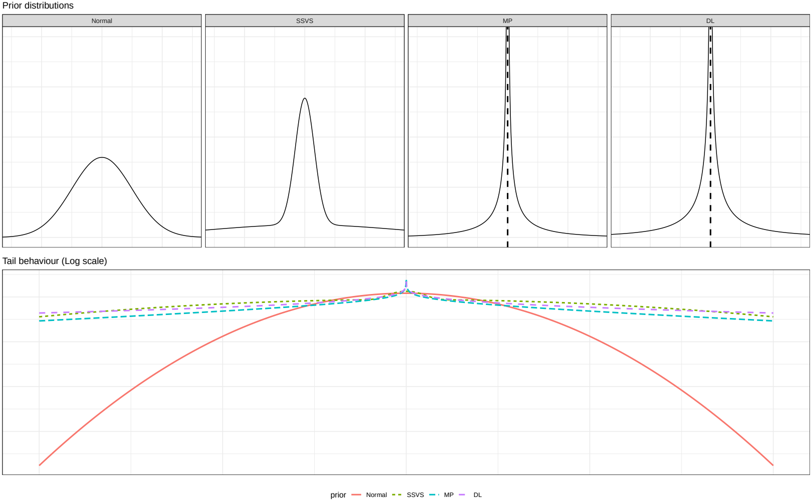
Bayes Theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y},\boldsymbol{\theta})}{p(\boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y})}$$
$$\propto p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{3}$$

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior}.$$

.

Prior distributions



Minnesota Prior (MP)

Let $\boldsymbol{\phi} = vec\{\boldsymbol{\Phi}\} = \{\phi_i, \ldots, \phi_n\}$. Then the Minnesota Prior (Litterman, 1986; Koop and Korobilis, 2010), is

$$\boldsymbol{\phi} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V}^{MP}), \quad \text{where } \boldsymbol{V}^{MP} \text{ is diagonal}^1. \tag{4}$$

Let $\boldsymbol{V}_m$ denote the block in $\boldsymbol{V}$ that corresponds to equation $m = 1, \ldots, M$ and $\boldsymbol{V}_{m,jj}^{MP}$ its diagonal elements. Then

$$\boldsymbol{V}_{m,jj}^{MP} = \begin{cases} \frac{\lambda_1}{r^2} & \text{for coefficients on own lag } r \text{ for } r = 1, \ldots, p, \\ \frac{\lambda_2 \hat{\sigma}_i}{r^2 \hat{\sigma}_j} & \text{for coefficients on cross lag } r. \end{cases} \tag{5}$$

It is common to fix λ_i for $i = 1, 2$ with $\lambda_2 \ll \lambda_1$. Instead of fixing these parameters Huber and Feldkircher (2019) propose

$$\lambda_i \quad \sim G(a_i, b_i) \quad \text{for } i = 1, 2. \tag{6}$$

[1] For data in levels (instead of growth rates), the prior mean of first own-lag coefficients would be 1.

MP: Contour plots

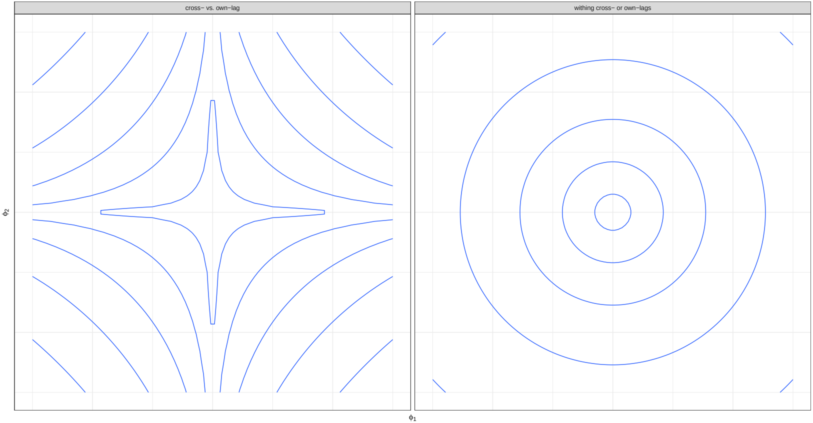


Figure: Contour plots of MP

Stochastic Search Variable Selection (SSVS)

For $i = 1, \ldots, n$:

$$\phi_i|\gamma_i \sim \gamma_i N(0, \tau_{1i}) + (1 - \gamma_i)N(0, \tau_{0i}), \qquad (7)$$

where $\tau_{1i} \gg \tau_{0i}$.

$$\gamma_j|\underline{p}_i \sim Bernoulli(\underline{p}_i), \qquad (8)$$
$$\underline{p}_i \sim Beta(s_1, s_2). \qquad (9)$$

Eq. 7 and
8 as in George et al. (2008).
Key hyperparameters:

- γ:
  auxiliary dummy variable,
- $\underline{p}$:
  prior inclusion probability.

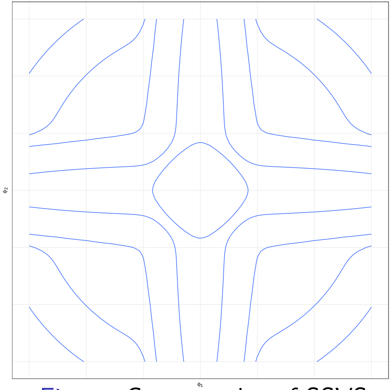


Figure: Contour plot of SSVS

Dirichlet Laplace Prior (DL)

As in Bhattacharya et al. (2015). For $i = 1, \ldots, n$:

$$\phi_i|\psi_i, \vartheta_i^2, \zeta^2 \sim N(0, \psi_i \vartheta_i^2 \zeta^2), \tag{10}$$

$$\psi_i \sim E(1/2), \tag{11}$$

$$\vartheta_i \sim Dir(a, \ldots, a), \tag{12}$$

$$\zeta \sim G(na, 1/2). \tag{13}$$

Key hyperparameters:

- ψ_i: auxiliary scaling parameter,
- ϑ_i: local shrinkage parameter,
- ζ: global shrinkage parameter.

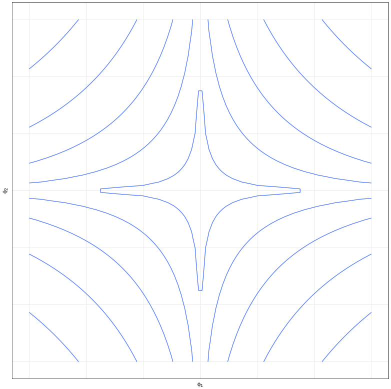


Figure: Contour plot of DL

Data

- Data from Fred Quarterly (McCracken and Ng, 2020).
  - Quarterly database for Macroeconomic Research: `https://research.stlouisfed.org/econ/mccracken/fred-databases/`
- Quarterly observations from 21 marcoeconomic and financial time series from 1959:Q4 to 2021:Q2.
- Transformations as in McCracken and Ng (2020).
- All series are *standardized* to have zero-mean and unit-variance.

Posterior Inference: Heatmaps of VAR(2) coefficients

Figure: Posterior Medians of different VAR(2) models

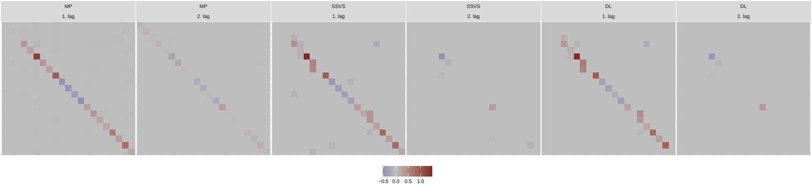
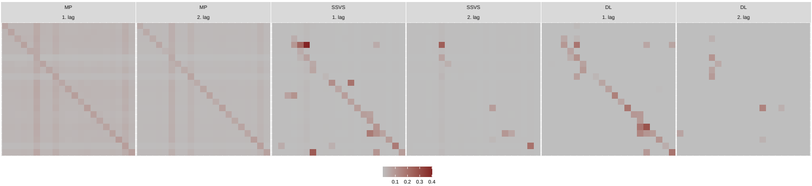


Figure: Posterior Interquartile ranges



Illusion of Sparsity

- Economic predictions with Big Data: Illusion of sparsity (Giannone et al., 2021, 2018, 2017).
  - Giannone et al. design a prior in order to detect whether the data is sparse or dense.
  - Conclusion: Data is dense!
- The illusion of the illusion of sparsity (Fava and Lopes, 2020).
  - Results are sensitive to prior choice!
  - With little arrangements of the prior used by Giannone et al. (2021) posterior estimates suggest data to be sparse.

Sparseness of coefficient matrices: Hoyer measure

The Hoyer measure (Hoyer, 2004) is

$$H = \frac{\sqrt{n} - \left(\sum_{i=1}^{n} |\phi_i|\right) / \sqrt{\sum_{i=1}^{n} \phi_i^2}}{\sqrt{n} - 1}. \tag{14}$$

- $H = 1$ indicates that there is only a single nonzero component in ϕ.
- $H = 0$ indicates that all components in ϕ are equal.

Sparseness of coefficient matrices: Results

		MP				
p	ol/cl	lag: 1	2	3	4	5
1	ol	0.235				
	cl	0.298				
2	ol	0.237	0.254			
	cl	0.300	0.315			
3	ol	0.239	0.227	0.232		
	cl	0.302	0.314	0.319		
4	ol	0.240	0.238	0.202	0.316	
	cl	0.304	0.314	0.319	0.323	
5	ol	0.242	0.252	0.206	0.291	0.301
	cl	0.306	0.315	0.319	0.322	0.322

		SSVS					DL				
p	ol/cl	lag: 1	2	3	4	5	lag: 1	2	3	4	5
1	ol	0.291					0.323				
	cl	0.875					0.852				
2	ol	0.326	0.777				0.383	0.829			
	cl	0.860	0.828				0.891	0.881			
3	ol	0.343	0.823	0.886			0.396	0.809	0.964		
	cl	0.864	0.755	0.760			0.909	0.902	0.905		
4	ol	0.287	0.867	0.729	0.904		0.396	0.915	0.917	0.973	
	cl	0.873	0.766	0.691	0.788		0.924	0.907	0.918	0.920	
5	ol	0.356	0.764	0.671	0.547	0.969	0.407	0.868	0.970	0.978	0.989
	cl	0.880	0.814	0.677	0.752	0.765	0.920	0.919	0.926	0.925	0.926

Table: Posterior means of the Hoyer measure for own-, cross-lag coefficients and lags separately.

Forecasting exercise & Predictive Likelihoods

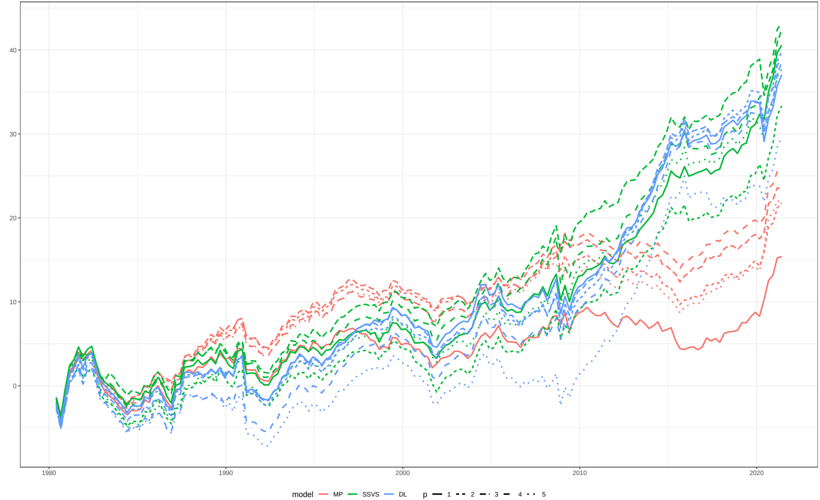
- Forecasting with 21-dimensional VAR(p)s for $p = 1, \ldots, 5$.
- Variables of Interest (VoI): GDP, CPI & FFR.
  - Benchmark model: a small VAR(4) containing only the VoI with a flat prior (normal prior with large variance).
- Initial training sample 1959:Q4 to 1980:Q1.
- Recursive forecasting design.
- Evaluation of the predictive density by means of *Predictive Likelihoods* (PL):

$$PL(t) = p(\mathbf{y}_{t+1}^{obs}|\mathbf{Y}_{[1:t]}^{obs}) = \int_{\boldsymbol{\Theta}} p(\mathbf{y}_{t+1}^{obs}|\mathbf{Y}_{[1:t]}^{obs}) \times p(\boldsymbol{\theta}|\mathbf{Y}_{[1:t]}^{obs}) d\boldsymbol{\theta}. \tag{15}$$

"Predictive density evaluated at the actual outcome".

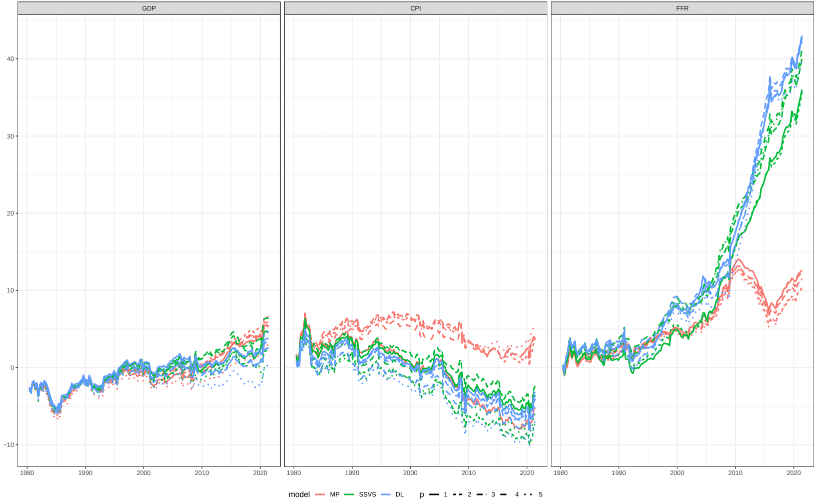
Evaluation of joint predictive density of GDP, CPI & FFR

Figure: Cumulative LPLs relative to LPLs of the benchmark model.



Evaluation of marginal predictive densities of GDP, CPI & FFR

Figure: Cumulative LPLs relative to LPLs of the benchmark model.



Dynamic Model Averaging (DMA)

- Can predictions be improved by combining weighted forecasts of all different models?
- Bayesian Model Averaging: Prior on model weights.
- Dynamic Averaging:
  - Equal weights for the first out of sample forecast.
  - Recursively update weights according to:

$$\omega_{t|t-1,i} = \frac{(\omega_{t-1|t-2,i} PL_{t-1|t-2,i})^{\alpha}}{\sum_{i \in \mathcal{M}} (\omega_{t-1|t-2,i} PL_{t-1|t-2,i})^{\alpha}}, \tag{16}$$

where α is the forgetting factor and $\mathcal{M}$ is the model space.

DMA: Evaluation of joint predictive density of GDP, CPI & FFR

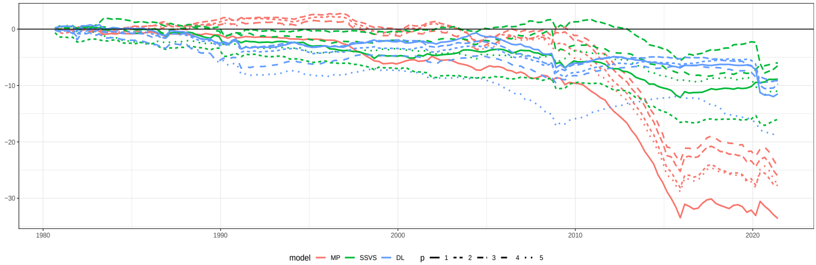


Figure: Cumulative LPLs relative to DMA ($\alpha = 0.9$).

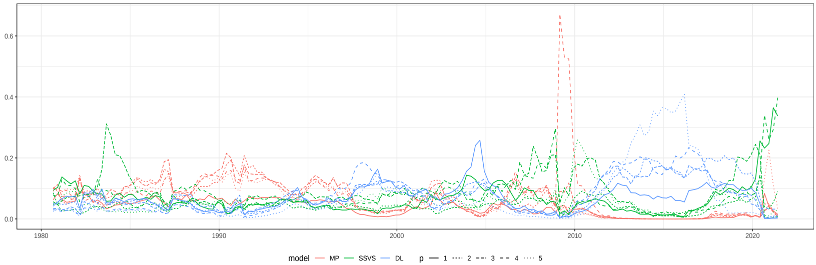


Figure: Model weights over time.

Summary

- Shrinkage priors are an effective tool against overfitting.
  - Predictions of specific variables can benefit from the use of large datasets.
- Illusion of sparsity debate: Predictive performances of different priors change over time and differ across evaluated variables.
- Dynamic Model Averaging combines merits of different modeling approaches.

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). Dirichlet–Laplace Priors for Optimal Shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490. Number: 512 https://doi.org/10.1080/01621459.2014.960967.

Fava, B. and Lopes, H. F. (2020). The Illusion of the Illusion of Sparsity: An exercise in prior sensitivity. *arXiv:2009.14296 [stat]*. arXiv: 2009.14296.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580. Number: 1 https://doi.org/10.1016/j.jeconom.2007.08.017.

Giannone, D., Lenza, M., and Primiceri, G. E. (2017). Economic Predictions with Big Data: The Illusion of Sparsity. SSRN Scholarly Paper ID 3031893, Social Science Research Network, Rochester, NY.

Giannone, D., Lenza, M., and Primiceri, G. E. (2018). Economic Predictions with Big Data: The Illusion of Sparsity. SSRN Scholarly Paper ID 3166281, Social Science Research Network, Rochester, NY.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic Predictions with Big Data: The Illusion of Sparsity. SSRN Scholarly Paper ID 3835164, Social Science Research Network, Rochester, NY. Issue: ID 3835164.

Hoyer, P. O. (2004). Non-negative Matrix Factorization with Sparseness Constraints. *The Journal of Machine Learning Research*, 5:1457–1469.

Huber, F. and Feldkircher, M. (2019). Adaptive Shrinkage in Bayesian Vector Autoregressive Models. *Journal of Business & Economic Statistics*, 37(1):27–39. Number: 1 https://doi.org/10.1080/07350015.2016.1256217.

Koop, G. and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends® in Econometrics*, 3(4):267–358. Number: 4 https://doi.org/10.1561/0800000013.

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics*, 4(1):25–38. Number: 1 https://doi.org/10.2307/1391384.

McCracken, M. W. and Ng, S. (2020). FRED-QD: A Quarterly Database for Macroeconomic Research. Technical report. https://doi.org/10.20955/wp.2020.005.