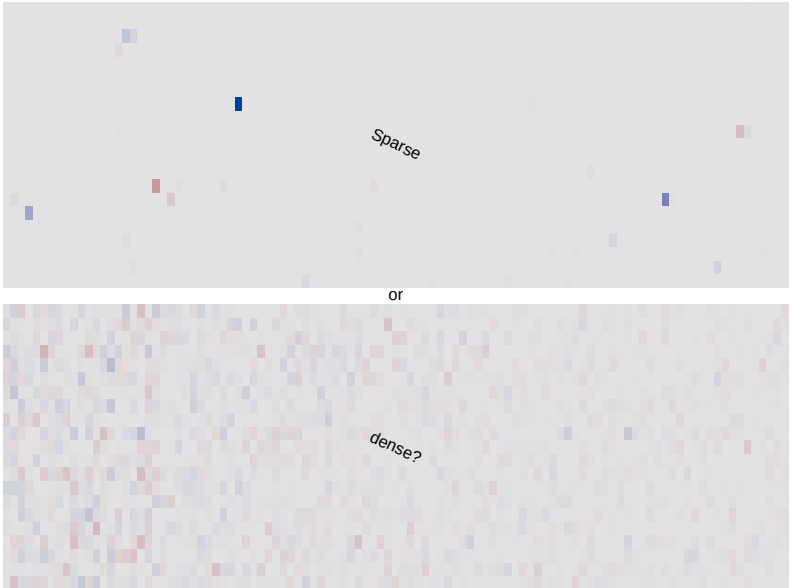
Forecasting macroeconomic data with Bayesian VARs: Sparse or dense? It depends!

Luis Gruber[1] Gregor Kastner[1]

[1]Department of Statistics, University of Klagenfurt

Introduction



Introduction

Sparse or dense?

- Economic predictions with Big Data: Illusion of sparsity (Giannone et al., 2021).
- The illusion of the illusion of sparsity (Fava and Lopes, 2021).

Sparsity of prior $\rightarrow$ sparsity of posterior $\rightarrow$ **predictive power?**

***Tailor-made* (problem specific) priors vs. *off the shelf* priors.**

- e.g. Minnesota prior vs. global-local priors.

Semi-global-(local) framework: combination of both worlds!

Outline

Econometric framework

Priors for the VAR coefficients

Semi-global-local shrinkage

Empirical application to data of the US economy

Econometric framework

Vector autoregression (VAR) with stochastic volatility for M time-series and p lags:

$$\boldsymbol{y}_t' = \sum_{k=1}^{p} \boldsymbol{y}_{t-k}' \boldsymbol{A}_k + \varepsilon_t', \quad \varepsilon_t \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{U'^{-1}} \boldsymbol{D_t} \boldsymbol{U^{-1}}).$$

- $K = M^2 p$ coefficients in $\phi = vec[(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)']$.
- $\boldsymbol{U}^{-1}$: $M \times M$ upper unitriangular matrix.
- $\boldsymbol{D}_t = \text{diag}(e^{h_{1t}}, \ldots, e^{h_{Mt}})$ with time-varying innovation variances featuring stochastic volatility.

Curse of dimensionality:

- T (number of observations) is typically small to moderate (e.g. quarterly data).
- K can be very large ($M = 21$ and $p = 1$ yields 441, $p = 5$ yields 2205 coefficients).
- Solution to prevent overfitting: **Bayesian shrinkage priors**.

Priors for the VAR coefficients

- Prior elicitation for VARs is a long-standing issue: e.g. Litterman (1986); Sims and Zha (1998); George et al. (2008); Bańbura et al. (2010); Koop (2013); Korobilis (2013); Giannone et al. (2015); Follett and Yu (2019); Huber and Feldkircher (2019); Cadonna et al. (2020); Cross et al. (2020); Kastner and Huber (2020); Chan (2021) etc. etc.

Our setting:

- The generic prior for ϕ is

$$\phi \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V}),$$
$$\boldsymbol{V} = diag(v_1, \ldots, v_n).$$

- The considered priors distinguish themselves in their treatment of $\boldsymbol{V}$.

Priors under scrutiny

- Global-local shrinkage priors (e.g., Polson and Scott, 2011):

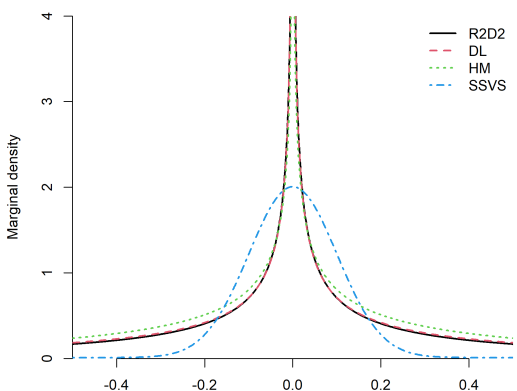
$$\phi_j|\vartheta_j\zeta \sim N(0, \vartheta_j\zeta), \quad \vartheta_j \sim f, \quad \zeta \sim g \quad \text{for } j = 1, \ldots, K.$$

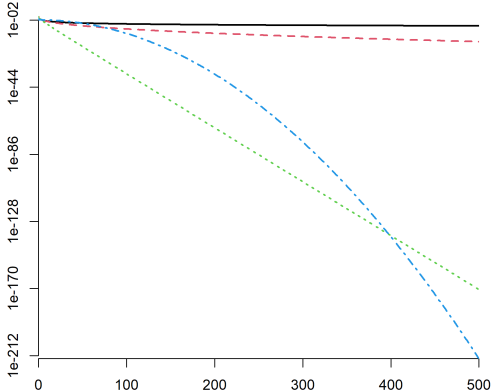
  - **R2D2** R^2-induced Dirichlet decomposition prior (Zhang et al., 2022). R2D2
  - **DL** Dirichlet-Laplace prior (Bhattacharya et al., 2015; Kastner and Huber, 2020). DL
  - **NG** Normal-gamma prior (Brown and Griffin, 2010).

- Other priors:

  - **HM** Hierarchical Minnesota prior (Huber and Feldkircher, 2019): Based on the prior in Litterman (1986) with independent hyperpriors on the shrinkage parameter for own- and the shrinkage parameter for cross-lags. HM
  - **SSVS** Stochastic search variable selection prior (George et al., 2008): Discrete mixture of two Gaussians, one with small variance (spike) and one with large variance (slab). SSVS

Concentration at zero and asymptotic tail-behavior (selection)





Univariate marginal prior densities of R2D2, DL, HM and SSVS. Hyperparameters s.t. interquartile range is 0.5.

Concentration at zero and asymptotic tail-behavior (selection)

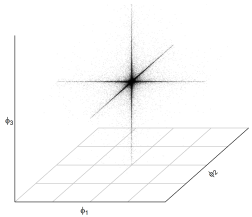
	Tail decay	Concentration at zero
R2D2	$O\left(\frac{1}{\lvert\phi\rvert^{1+2b}}\right)$	$O\left(\frac{1}{\lvert\phi\rvert^{1-2a_\pi}}\right)$
DL	$O\left(\frac{\lvert\phi\rvert^{a_\nu/2-3/4}}{\exp\left\{\sqrt{2\lvert\phi\rvert}\right\}}\right)$	$O\left(\frac{1}{\lvert\phi\rvert^{1-a_\nu}}\right)$
HM	$O\left(\frac{\lvert\phi\rvert^{c-1}}{\exp\left\{\sqrt{2d}\lvert\phi\rvert\right\}}\right)$	$O\left(\frac{1}{\lvert\phi\rvert^{1-2c}}\right)$
SSVS	$O\left(\frac{\tau_1 \exp\left\{\frac{\phi^2}{\tau_1^2}\right\}+\tau_0^2 \exp\left\{\frac{\phi^2}{\tau_0^2}\right\}}{\exp\left\{\frac{\phi^2}{\tau_0^2}\right\}\exp\left\{\frac{\phi^2}{\tau_1^2}\right\}}\right)$	no singularity

Limiting behavior and concentration at zero of the univariate marginal distributions of the R2D2 prior, the DL prior, the HM prior and the SSVS prior.

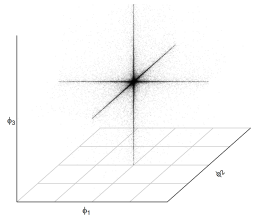
Sparseness of the priors

Hoyer measure

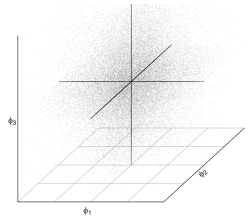
Sparseness by means of the **Hoyer sparseness measure** (Hoyer, 2004). Normalized measure in the range $[0, 1]$, where a value of 1 indicates maximum sparseness.



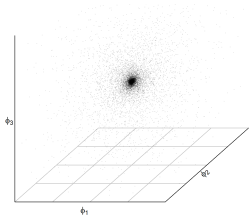
(a) R2D2: $(0.5, 1)$



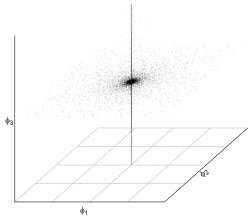
(b) DL: $(0.6, 1)$



(c) SSVS: ≈ 0.45



(d) HM (within): ≈ 0.2



(e) HM (across): $\approx 0.2+$

Shrink semi-globally!

- Global-local shrinkage (Horseshoe, Normal-Gamma, R2D2, etc.):

$$\phi_i|\vartheta_i\zeta \sim K(\vartheta_i\zeta), \quad \vartheta_i \sim f, \quad \zeta \sim g \quad \text{for } i = 1, \ldots, K.$$

Problem:

- *Shrink globally, act locally* à la Polson and Scott (2011) does not work well if the noise level is different in different parts of the parameter space.

Solution: Structured shrinkage.

- Semi-global-local shrinkage, i.e.

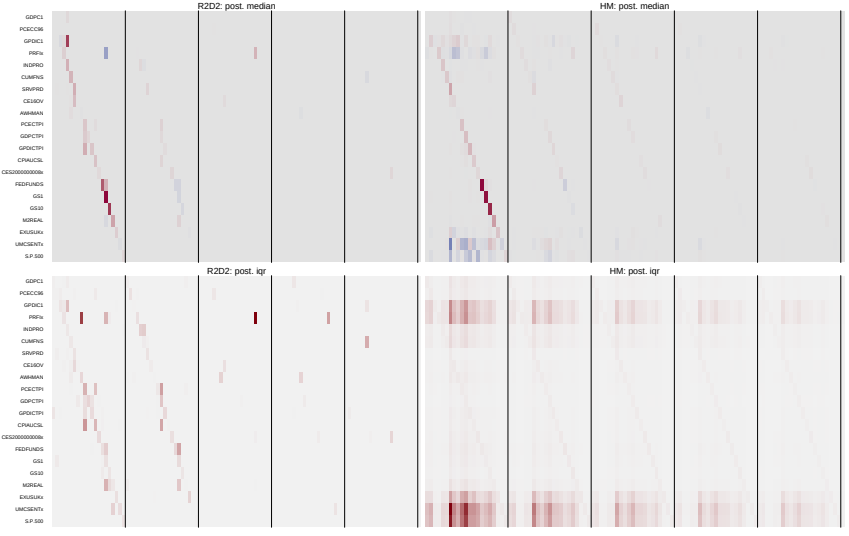
$$\phi_i \sim K(\vartheta_i\zeta_j), \quad \vartheta_i \sim f, \quad \zeta_j \sim g, \quad i \in \mathcal{A}_j, \quad j = 1, \ldots, s,$$

  where $\mathcal{A}_j$ denotes the generic index set that labels the coefficients of the jth group in ϕ.
- Semi-global shrinkage for VARs: Partition $\phi = vec[(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)']$ into $s = 2p$ groups: Distinction between diagonal and off-diagonal elements in $\boldsymbol{A}_k$, $k = 1, \ldots, p$.

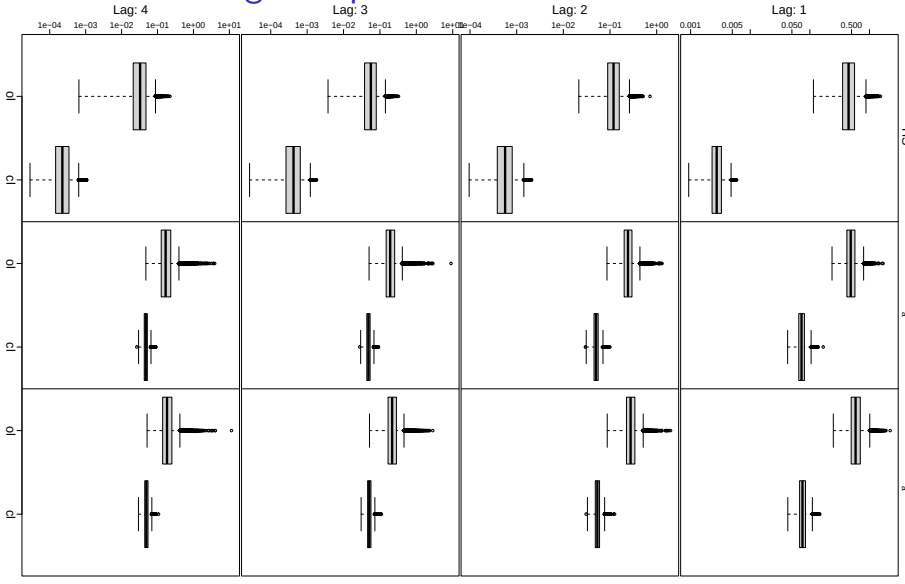
Empirical application to data of the US economy

- Data from Fred Quarterly (McCracken and Ng, 2021).
  - 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 2020:Q1.
- Transformations: $\Delta \log(x_t)$ (growth rates).

Empirical application: Full sample posterior estimates of VAR(5)s

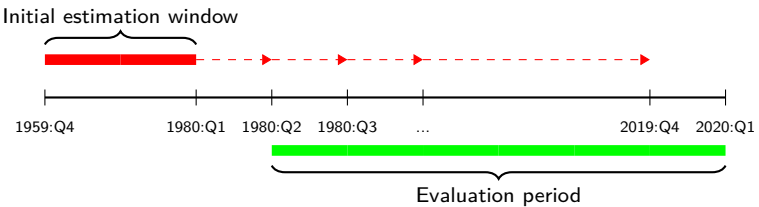


Posterior of some semi-global parameters



Forecasting design

- Forecasting with 21-dimensional VAR(p)s for $p = 1, \ldots, 5$.
- Expanding window:



Forecasting performance (selection)

Model	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
MP_LIT	-63.35	-18.84	-7.99	0.00	-10.48
DL	-39.61	10.89	24.49	2.70	-7.65
HS	-55.62	-18.83	-7.05	-39.38	-39.53
HS*	-25.96	23.59	**40.41**	**54.33**	**41.75**
NG_a	-43.22	8.09	5.38	-17.43	-58.55
NG_a^*	-24.68	37.59	28.14	8.69	-24.75
$R2D2_a$	-42.95	11.48	1.42	-24.70	-40.59
$R2D2_a^*$	**-17.67**	**44.37**	32.56	13.31	-7.36

Sum of one step ahead log predictive Bayes factors (BFs) relative to the traditional Minnesota prior with four lags (MP_LIT, $p = 4$). Asterisks denote the semi-global-local modifications.

Dynamic model averaging (Raftery et al., 2010)

- Equal weights for the first out-of-sample forecast.
- Update weights according to:

$$\omega_{t|t,i} = \frac{\omega_{t|t-1,i} PL_{t|t-1,i}}{\sum_{i \in \mathcal{M}} \omega_{t|t-1,i} PL_{t|t-1,i}},$$

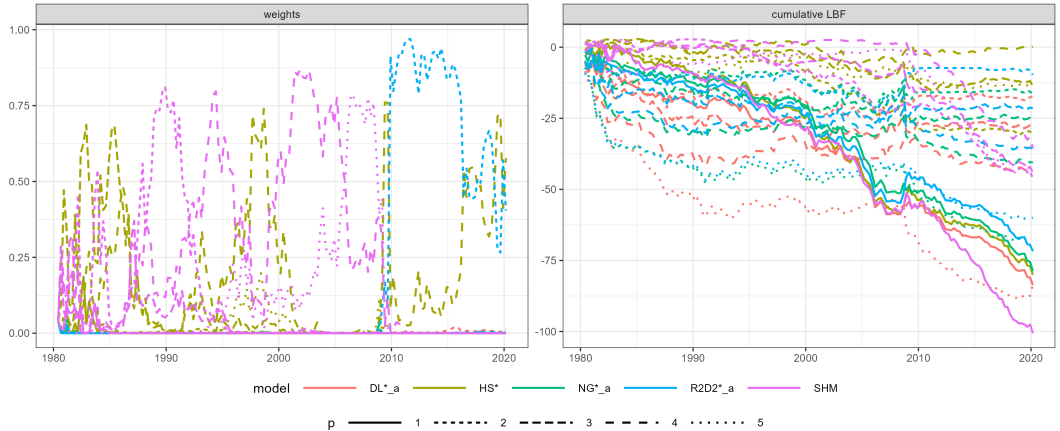
where

$$\omega_{t|t-1,i} = \frac{\omega_{t-1|t-1,i}^{\alpha}}{\sum_{i \in \mathcal{M}} \omega_{t-1|t-1,i}^{\alpha}},$$

and $0 \leq \alpha \leq 1$ denotes the discount factor.

- We set $\alpha = 0.99$: $PL_{t-1|t-2}$ receives 99% as much weight as $PL_{t|t-1}$!

Model weights



Model weights and cumulative one-step-ahead log predictive BFs relative to DMA.

MCMC estimation via bayesianVARs

- R package **bayesianVARs** available on CRAN and at https://github.com/luisgruber/bayesianVARs.
- Corrected triangular algorithm (Carriero et al., 2022) for Cholesky SV via stochvol (Kastner, 2016).
- Also allows for factor SV error structure (Kastner and Huber, 2020) via factorstochvol (Hosszejni and Kastner, 2021) instead of Cholesky SV, guaranteeing variable order independence.



Summary of contributions/results

- Comprehensive **overview of shrinkage priors** in VARs: **Minnesota** prior, **SSVS**-prior, **Horseshoe** prior, **Dirichlet-Laplace** prior, **Normal-Gamma** prior and **R2D2**-prior.
  - Comparison of **shrinking behavior**, i.e. tail behavior and concentration at zero.
  - Analysis of the **sparseness of the priors**: Applying the Hoyer sparseness measure we find that **prior sparseness translates to posterior sparseness**, where sparseness is best expressed by global-local priors and denseness by Minnesota-type priors.
  - Extensive simulation study confirms our considerations.
- **Semi-global-(local)** framework:
  - Problem specific fine-tuning of *off the shelf* priors improves forecasting performance.
- **Heterogeneous predictive evidence** in an application to data of the US economy.
  - **Forecasting performance varies** over **time** and across **evaluated variables**.
  - **Dynamic Model Averaging** combines merits of different modelling approaches.

References I

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

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.

Brown, P. J. and Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.

Cadonna, A., Frühwirth-Schnatter, S., and Knaus, P. (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics*, 8(2).

Carriero, A., Chan, J., Clark, T. E., and Marcellino, M. (2022). Corrigendum to "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors" [J. Econometrics 212 (1) (2019) 137–154]. *Journal of Econometrics*, 227(2):506–512.

Chan, J. C. C. (2021). Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting*, 37(3):1212–1226.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

References II

Cross, J. L., Hou, C., and Poon, A. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915.

Fava, B. and Lopes, H. F. (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics*, 35(4):699–720.

Follett, L. and Yu, C. (2019). Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics*, 11:130–144.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics*, 97(2):436–451.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

Hosszejni, D. and Kastner, G. (2021). Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *Journal of Statistical Software*, 100:1–34.

References III

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.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 12(4):371–389.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(1):1–30.

Kastner, G. and Huber, F. (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, 39(7):1142–1165.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230.

References IV

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

McCracken, M. W. and Ng, S. (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis Review*, 103(1):1–44.

Polson, N. G. and Scott, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting*, pages 501–538, Oxford, UK. Oxford University Press.

Raftery, A. E., Kárný, M., and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1):52–66. PMID: 20607102.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Zhang, Y. D., Naughton, B. P., Bondell, H. D., and Reich, B. J. (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117(538):862–874.

R^2-induced Dirichlet decomposition (R2D2) prior

- Zhang et al. (2022).
- R2D2 prior on ϕ_j for $j = 1, \ldots, n$:

$$
\begin{aligned}
\phi_j | \vartheta_j, \zeta &\sim \textit{DoubleExp}\left((\vartheta_j \zeta / 2)^{1/2} \right), \\
\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)' &\sim \textit{Dirichlet}(a_\pi, \ldots, a_\pi), \\
\zeta &\sim \textit{BetaPrime}(a, b), \\
b &\sim \textit{DiscreteUnif} \text{ on } [0.01, 0.02, \ldots, 1], \\
(\Rightarrow \textit{Var}(\phi_j | \vartheta_j, \zeta) &= \vartheta_j \zeta).
\end{aligned}
\tag{1}
$$

- Let $W := \sum_{j=1}^{n} \vartheta_j \zeta = \zeta$ be the **total prior variability**.
  - ζ: total prior variability.
  - ϑ_j: proportion allocated to ϕ_j.
- R2D2 prior is induced by a prior on the total prior variability:
  - $R^2 := \frac{W}{W+1} \sim \textit{Beta}(a, b) \quad \Leftrightarrow \quad W \sim \textit{BetaPrime}(a, b)$.

Dirichlet-Laplace Prior

◂ Priors under scrutiny

$$\phi_j \sim DE(\vartheta_j \zeta) \quad \Leftrightarrow \quad \phi_j | \psi_j, \vartheta_j, \zeta \sim N(0, \psi_j \vartheta_j^2 \zeta^2), \quad \psi_j \sim E(1/2),$$
$$\boldsymbol{\vartheta} \sim Dir(a_{DL}, \ldots, a_{DL}), \quad \zeta \sim G(n a_{DL}, 1/2).$$

Hierarchical Minnesota prior

◂ Priors under scrutiny

Denote $\underline{\mathbf{V}}_i$ the block of $\underline{\mathbf{V}}$ that corresponds to the K coefficients in the ith equation, and let $\underline{\mathbf{V}}_{i,jj}$ be its diagonal elements. The diagonal elements are set to

$$\underline{\mathbf{V}}_{i,jj} = \begin{cases} \frac{\lambda_1}{r^2} & \text{for coefficients on own lag } r \text{ for } r = 1, \ldots, p, \\ \frac{\lambda_2 \hat{\sigma}_i}{r^2 \hat{\sigma}_j} & \text{for coefficients on lag } r \text{ of variable } j \neq i, \end{cases}$$

$$\lambda_i \sim Gamma(c_i, d_i) \text{ for } i = 1, 2,$$

where $\hat{\sigma}_i$ is the OLS variance of a univariate AR(6) model of the ith variable.

Stochastic Search Variable Selection

◂ Priors under scrutiny

$$\phi_j|\gamma_j \sim N\left(0, (1-\gamma_j)\tau_{0j}^2 + \gamma_j\tau_{1j}^2\right),$$
$$\gamma_j|\underline{p}_j \sim Bernoulli(\underline{p}_j).$$

Hoyer sparseness measure

Sparseness of the priors

The Hoyer measure for a vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$,

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

is a normalized measure where $H = 1$ indicates maximum sparseness, i.e., only one single nonzero value, and $H = 0$ indicates maximum denseness, i.e., the absolute values of all components are equal.

Marginal prior distribution of the Hierarchical Minnesota prior

$$
\begin{aligned}
p_{HM}(\phi) &= \int_0^\infty \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{\phi^2}{2\lambda}\right\} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp\{-\beta\lambda\}\, d\lambda \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{(\alpha-\frac{1}{2})-1} \exp\left\{-\frac{1}{2}\left(\frac{\phi^2}{\lambda} + 2\beta\lambda\right)\right\} d\lambda \\
&= \frac{1}{\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{2K_{\alpha-\frac{1}{2}}(\sqrt{2\beta\phi^2})}{(\frac{2\beta}{\phi^2})^{\frac{\alpha-\frac{1}{2}}{2}}} \\
&\quad \times \overbrace{\int_0^\infty \frac{(\frac{2\beta}{\phi^2})^{\frac{\alpha-\frac{1}{2}}{2}}}{2K_{\alpha-\frac{1}{2}}(\sqrt{2\beta\phi^2})} \lambda^{(\alpha-\frac{1}{2})-1} \exp\left\{-\frac{1}{2}\left(\frac{\phi^2}{\lambda} + 2\beta\lambda\right)\right\} d\lambda}^{\text{Density of GIG distribution}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{2K_{\alpha-\frac{1}{2}}(\sqrt{2\beta\phi^2})}{(\frac{2\beta}{\phi^2})^{\frac{\alpha-\frac{1}{2}}{2}}} = \frac{\left(\frac{\phi^2}{2\beta}\right)^{\frac{\alpha-\frac{1}{2}}{2}} \beta^\alpha 2K_{\alpha-\frac{1}{2}}(\sqrt{2\beta}|\phi|)}{\Gamma(\alpha)\sqrt{2\pi}},
\end{aligned}
$$

where $K_\nu(z)$ is the modified Bessel function of the second kind.

Hierarchical Minnesota prior: Concentration at zero

According to 10.30.2 in `https://dlmf.nist.gov/10.30`, when $\nu > 0$, $z \to 0$ and z is real, $K_\nu(z) \approx \Gamma(\nu)(z/2)^{-\nu}/2$. Moreover, according to 10.27.3 in `https://dlmf.nist.gov/10.27` $K_\nu(z) = K_{-\nu}(z)$. Given $0 < \alpha < 0.5$, $|\phi| \to 0$,

$$
\begin{aligned}
p_{HM}(\phi) &= \frac{\left(\frac{\phi^2}{2\beta}\right)^{\frac{\alpha-\frac{1}{2}}{2}} \beta^\alpha 2 K_{\alpha-\frac{1}{2}}(\sqrt{2\beta}|\phi|)}{\Gamma(\alpha)\sqrt{2\pi}} \\
&\approx \frac{\left(\frac{\phi^2}{2\beta}\right)^{\frac{\alpha-1/2}{2}} \beta^\alpha 2\Gamma(1/2-\alpha)\left(\frac{\sqrt{2\beta}|\phi|}{2}\right)^{\alpha-1/2}}{2\Gamma(\alpha)\sqrt{2\pi}} \\
&= C\phi^{\alpha-\frac{1}{2}}\phi^{\alpha-\frac{1}{2}} = C\phi^{2\alpha-1},
\end{aligned}
$$

where $C = \frac{\left(\frac{1}{2\beta}\right)^{\frac{\alpha-\frac{1}{2}}{2}} \beta^\alpha \Gamma\left(\frac{1}{2}-\alpha\right)\left(\sqrt{\frac{\beta}{2}}\right)^{\alpha-\frac{1}{2}}}{\Gamma(\alpha)\sqrt{2\pi}}$ is a constant value.

Hierarchical Minnesota prior: Asymptotic tail behavior

According to 10.25.3 in `https://dlmf.nist.gov/10.25`, when both ν and z are real, if $z \to \infty$, then $K_\nu(z) \approx \pi^{1/2}(2z)^{-1/2}\exp\{-z\}$. Then as $|\phi| \to \infty$,

$$
\begin{aligned}
p_{HM}(\phi) &= \frac{\left(\frac{\phi^2}{2\beta}\right)^{\frac{\alpha-\frac{1}{2}}{2}} \beta^\alpha 2K_{\alpha-\frac{1}{2}}(\sqrt{2\beta}|\phi|)}{\Gamma(\alpha)\sqrt{2\pi}} \\
&\approx \frac{\left(\frac{\phi^2}{2\beta}\right)^{\frac{\alpha-1/2}{2}} \beta^\alpha 2\pi(2\sqrt{2\beta\phi^2})^{-1/2}\exp\left\{-\sqrt{2\beta\phi^2}\right\}}{\Gamma(\alpha)\sqrt{2\pi}} \\
&= C|\phi|^{\alpha-1/2}|\phi|^{-1/2}\exp\left\{-\sqrt{2\beta}|\phi|\right\} = C\frac{|\phi|^{\alpha-1}}{\exp\left\{\sqrt{2\beta}|\phi|\right\}},
\end{aligned}
$$

where $C = \frac{\left(\frac{1}{2\beta}\right)^{\frac{\alpha-1/2}{2}} \beta^\alpha 2\pi(2\sqrt{2\beta})^{-1/2}}{\Gamma(\alpha)\sqrt{2\pi}}$ is a constant value.

VAR(p) for M time series

$$\boldsymbol{Y}_t = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{Y}_{t-k} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{3}$$

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

Curse of dimensionality:

- Denote $\boldsymbol{\Phi} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$.
- $n = pM^2$ free parameters in $\boldsymbol{\Phi}$.
  - Solution: Regularization!

Var(p) for M time series continued

- Cogley and Sargent (2005):

$$\boldsymbol{\Sigma}_t = \boldsymbol{L}'^{-1} \boldsymbol{D}_t \boldsymbol{L}^{-1}. \tag{4}$$

- $\boldsymbol{L}^{-1}$ is an upper triangular with ones on the diagonal.
- $\boldsymbol{D}_t$ is a diagonal matrix.
- Following Jacquier et al. (1994) and Kim et al. (1998) the orthogonalized errors $\boldsymbol{\xi}_t = \boldsymbol{L}\boldsymbol{\varepsilon}_t$ are assumed to follow univariate SV-models: For $i = 1, \ldots, M$

$$\xi_{it} = \exp\left(\frac{h_{it}}{2}\right) \epsilon_{it}, \tag{5}$$

$$h_{it} = \mu_i + \rho_i (h_{i(t-1)} - \mu_i) + \sigma_i \eta_{it}, \tag{6}$$

- $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be i.i.d. $N(\mathbf{0}, \boldsymbol{I}_M)$.