

Shrinkage Priors for Bayesian Vectorautoregressions featuring Stochastic Volatility Using the **R** Package **bayesianVARs**

Luis Gruber
University of Klagenfurt
luis.gruber@aau.at

Introduction

In multivariate time series analysis, vectorautoregressions (VARs) are widely applied in fields such as brain connectivity modeling [e.g., Goebel et al., 2003, Chiang et al., 2017, Wegmann et al., 2024] and the modeling of macroeconomic and financial time series [e.g., Sims, 1980, Karlsson, 2013, Crump et al., 2021]. Especially in macroeconomic applications, VARs have probably become the workhorse model for forecasting. The VAR model of order p , $\text{VAR}(p)$, can be formulated as follows:¹

$$\mathbf{y}'_t = \sum_{l=1}^p \mathbf{y}'_{t-l} \mathbf{A}_l + \boldsymbol{\varepsilon}'_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_t), \quad t = 1, \dots, T, \quad (1)$$

where \mathbf{y}_t is the M -dimensional vector of interest, \mathbf{A}_l , for $l = 1, \dots, p$, an unknown $M \times M$ matrix of regression coefficients, $\boldsymbol{\varepsilon}_t$ an M -dimensional vector of errors, and $\boldsymbol{\Sigma}_t$ the corresponding $M \times M$ variance-covariance matrix. For ease of notation, let $\boldsymbol{\Phi} := (\mathbf{A}_1, \dots, \mathbf{A}_p)'$ denote the $(K = pM) \times M$ matrix containing all VAR coefficients and let $\boldsymbol{\phi} := \text{vec}(\boldsymbol{\Phi})$ denote the vectorization thereof with length $n = pM^2$.

To facilitate efficient and reliable estimation when M gets large, we consider two different decompositions of the variance-covariance matrix $\boldsymbol{\Sigma}_t$ explained in the following paragraphs.

VAR with factor stochastic volatility Assuming that the errors feature a factor stochastic volatility structure, following Kastner and Huber [2020], we decompose the variance-covariance matrix into

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda} \mathbf{V}_t \boldsymbol{\Lambda}' + \mathbf{Q}_t. \quad (2)$$

Both $\mathbf{Q}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{Mt}})$ and $\mathbf{V}_t = \text{diag}(e^{h_{M+1,t}}, \dots, e^{h_{M+r,t}})$ are diagonal matrices of dimension M and r , respectively, and $\boldsymbol{\Lambda}$ is the $M \times r$ matrix of factor loadings. This is obviously equivalent to introducing r conditionally independent latent factors $\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{V}_t)$ and rewriting the error term in (1) as

$$\boldsymbol{\varepsilon}'_t = \mathbf{f}'_t \boldsymbol{\Lambda}' + \boldsymbol{\eta}'_t, \quad (3)$$

where $\boldsymbol{\eta}_t \sim \mathcal{N}_M(\mathbf{0}, \mathbf{Q}_t)$. The matrix \mathbf{Q}_t contains the idiosyncratic, series specific, variances. The matrix \mathbf{V}_t contains the factor specific variances governing the contemporaneous dependencies. The logarithms of

¹For simplicity of exposition we omit the intercept in the following (which nonetheless **bayesianVARs** implements by default).

the elements in \mathbf{Q}_t and \mathbf{V}_t follow a priori independent autoregressive processes of order one (AR(1)). More specifically, the evolution of the idiosyncratic log-variance $h_{it} \sim \mathcal{N}(\mu_i + \phi_i(h_{i,t-1} - \mu_i), \sigma_i^2)$, for $i = 1, \dots, M$, is described by the parameters μ_i , the level, ϕ_i , the persistence and σ_i^2 , the variance. The factor-specific log-variance $h_{jt} \sim \mathcal{N}(\phi_j h_{j,t-1}, \sigma_j^2)$, for $j = M+1, \dots, M+r$, is assumed to have mean zero to identify the scaling of the elements of $\mathbf{\Lambda}$. Without imposing restrictions on the factor loadings, the VAR with factor stochastic volatility is invariant to the way the variables are ordered.

VAR with Cholesky stochastic volatility Assuming that the errors feature a Cholesky stochastic volatility structure, following Cogley and Sargent [2005], we decompose the variance-covariance matrix into

$$\mathbf{\Sigma}_t = \mathbf{U}'^{-1} \mathbf{D}_t \mathbf{U}^{-1}, \quad (4)$$

where \mathbf{U} is an $M \times M$ upper triangular matrix with ones on the diagonal. The logarithms of the elements of the M -dimensional diagonal matrix $\mathbf{D}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{Mt}})$ are assumed to follow a priori independent AR(1) processes, i.e. $h_{it} \sim \mathcal{N}(\mu_i + \phi_i(h_{i,t-1} - \mu_i), \sigma_i^2)$, for $i = 1, \dots, M$. Since \mathbf{U}_t is a triangular matrix, the VAR with Cholesky stochastic volatility depends on the way the variables are ordered.

Prior Distributions

While flexible, VARs are known to be overparameterized: In macroeconomic applications the number of available observations T can be relatively small compared to the number of VAR coefficients n , since the data is usually reported on a quarterly or yearly basis. Bayesian shrinkage priors can be used to alleviate this issue. In the following paragraphs, we discuss several prior options for the VAR coefficients before briefly discussing prior choices for the variance-covariance matrix. In general, we assume that the joint prior distribution has a product form $p(\phi, \mathbf{\Sigma}_t) = p(\phi)p(\mathbf{\Sigma}_t)$, i.e. we assume that a priori ϕ and $\mathbf{\Sigma}_t$ are independent. The generic prior for the VAR coefficients is conditionally normal $\phi | \mathbf{V} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{V})$, where $\mathbf{V} = \text{diag}(v_1, \dots, v_n)$ is an n -dimensional diagonal matrix. The priors distinguish themselves in their treatment of \mathbf{V} .

Hierarchical Minnesota prior The *original* Minnesota prior proposed in Litterman [1986] is mainly characterized by two assumptions: First, the own past of a given variable is more important in predicting its current value than the past of other variables. Second, the most recent past is assumed to be more important in predicting current values than the more distant past. Hence, \mathbf{V} is structured in a way, such that the sub-diagonal elements of $\mathbf{\Phi}$ (the own-lag coefficients) are shrunk less than the off-diagonal elements (the cross-lag coefficients). And, coefficients associated with more recent lags are shrunk less than the ones associated with more distant lags. Denote \mathbf{V}_i the block of \mathbf{V} that corresponds to the K coefficients in the i th equation, and let $\mathbf{V}_{i,jj}$ be its diagonal elements. The diagonal elements are set to

$$\mathbf{V}_{i,jj} = \begin{cases} \frac{\lambda_1}{l^2} & \text{for coefficients on own lag } l \text{ for } l = 1, \dots, p, \\ \frac{\lambda_2 \hat{\sigma}_i^2}{l^2 \hat{\sigma}_j^2} & \text{for coefficients on lag } l \text{ of variable } j \neq i, \end{cases} \quad (5)$$

where $\hat{\sigma}_i^2$ is the OLS variance of a univariate AR(6) model of the i th variable. The term l^2 in the denominator automatically imposes more shrinkage on the coefficients towards their prior mean as lag length increases.

The term $\frac{\sigma_1^2}{\sigma_j^2}$ adjusts not only for different scales in the data, it is also intended to account for different scales of the responses of one economic variable to another. To shrink own-lag coefficients less than cross-lag coefficients, one could set $\lambda_1 > \lambda_2$. The hierarchical Minnesota prior, however, treats both shrinkage parameters as unknown. Following Huber and Feldkircher [2019], we place independent gamma priors on λ_1 and λ_2 ,

$$\lambda_i \sim \mathcal{G}(c_i, d_i), \quad \text{for } i = 1, 2. \quad (6)$$

Semi-global local shrinkage Global local shrinkage priors in the fashion of Polson and Scott [2011] are also used in the VAR literature [e.g., Follett and Yu, 2019, Huber and Feldkircher, 2019, Kastner and Huber, 2020]. In order to combine the merits of tailor-made priors, such as the Minnesota prior, with the flexibility of off-the-shelf global local shrinkage priors, Gruber and Kastner [2025] propose the class of semi-global local priors. Other than global local priors, which shrink globally, semi-global local priors shrink semi-globally, meaning that semi-global shrinkage is imposed on k pre-specified subgroups of the parameter space. Let \mathcal{A}_j , for $j = 1, \dots, k$, denote the generic index set that labels the coefficients of the j th group in ϕ (e.g., the first group could be the own-lag coefficients associated with the first lag, the second group could be the cross-lag coefficients associated with the first lag, etc.). Then, a semi-global local prior with k groups has the following hierarchical representation:

$$\phi_i \sim K(\vartheta_i \zeta_j), \quad \vartheta_i \sim f, \quad \zeta_j \sim g, \quad i \in \mathcal{A}_j, \quad j = 1, \dots, k, \quad (7)$$

where $K(\delta)$ denotes a symmetric unimodal density with variance δ , ζ_j represents the semi-global shrinkage, and ϑ_i the local shrinkage. The only additional input required is the partitioning of ϕ into k subgroups. Several options for grouping the coefficients are ready-made in **bayesianVARs**, though any custom grouping could be specified as well. The *equation-specific* grouping indicates that the covariates of each equation form M separate groups (column-wise shrinkage w.r.t. Φ). The *covariate-specific* partitioning implies that the K covariates across all equations form separate groups (row-wise shrinkage w.r.t. Φ). The *own-lag-cross-lag-lagwise* (olcl-lagwise) partitioning mimics some features of the Minnesota prior: In each lag, the diagonal elements (the own-lags) and the off-diagonal elements (the cross-lags) constitute separate groups, which makes $2p$ groups in total. The following list of hierarchical shrinkage priors, which can be cast in the form of semi-global (local) priors, are implemented in **bayesianVARs** (in alphabetical order): Dirichlet-Laplace (DL) prior [Bhattacharya et al., 2015], Horseshoe prior [Carvalho et al., 2010], normal-gamma (NG) prior [Brown and Griffin, 2010], R^2 -induced Dirichlet decomposition (R2D2) prior [Zhang et al., 2022] and stochastic-search-variable-selection (SSVS) prior [George et al., 2008]. For more detailed characteristics and comparisons of those priors we refer to Gruber and Kastner [2025].

Priors for the variance-covariance matrix In the case that the variance-covariance is modeled via the factor decomposition, the priors from Kastner et al. [2017] and Kastner [2019] are used. In the case that the errors are assumed to feature the Cholesky stochastic volatility structure, **bayesianVARs** implements the DL prior, the HS prior, the NG prior, the R2D2 prior, and the SSVS prior for the free off-diagonal elements in U . Concerning the latent variables and their associated parameters in D_t , the priors from Kastner and Frühwirth-Schnatter [2014] are used.

Homoscedastic VARs It should be noted that `bayesianVARs` also implements homoskedastic VARs where $\Sigma_t = \Sigma$ for all t . In case of the VAR with factor structure on the errors it holds that $V_t = V = I_r$ and $Q_t = Q = \text{diag}(q_1, \dots, q_M)$ for all t . A priori, the i th diagonal element $q_i \sim \mathcal{IG}(a_f, b_f)$ is assumed to follow an inverse gamma distribution for $i = 1, \dots, M$, independently. In case of the VAR with Cholesky structure on the errors, it holds that $D_t = D = \text{diag}(d_1, \dots, d_M)$ for all t . The prior distribution of the i th diagonal element is inverse gamma, i.e. $d_i \sim \mathcal{IG}(a_c, b_c)$ for $i = 1, \dots, M$, independently.

Algorithm

In a nutshell, `bayesianVARs` implements a Markov chain Monte Carlo (MCMC) algorithm which alternately samples from the full conditional posterior distribution of the VAR coefficients $p(\phi|\bullet)$ and from the full conditional posterior distribution of the paths of the variance-covariance matrix $p(\Sigma_t|\bullet)$ for $t = 1, \dots, T$, with \bullet indicating that we condition on the remaining parameters and latent quantities of the model. To render computation of the necessary steps required for sampling from $p(\phi|\bullet)$ feasible, `bayesianVARs` implements the equation-per-equation algorithm proposed in Kastner and Huber [2020] for the VAR with factor stochastic volatility and the correct triangular algorithm from Carriero et al. [2022] for the VAR with Cholesky stochastic volatility. The hyperparameters of the hierarchical shrinkage priors are sampled from the respective full conditional posterior distributions outlined in Gruber and Kastner [2025]. To sample from $p(\Sigma_t|\bullet)$ for $t = 1, \dots, T$, for the VAR with factor stochastic volatility, `bayesianVARs` accesses the package `factorstochvol` [Hosszejni and Kastner, 2021]. For the VAR with Cholesky stochastic volatility, the latent variables and associated parameters in D_t are sampled using the package `stochvol` [Kastner, 2016]. The free off-diagonal elements in U are sampled equation-per-equation as proposed in Cogley and Sargent [2005]. Last but not least, all computationally intensive tasks are written in C++ and interfaced with R via `Rcpp` [Eddelbuettel and François, 2011] and `RcppArmadillo` [Eddelbuettel and Sanderson, 2014] for increased computational efficiency.

Case study

We demonstrate the main functionality of `bayesianVARs` using the `usmacro_growth` dataset included in the package. The dataset – obtained from FRED-QD, a quarterly database for macroeconomic research [McCracken and Ng, 2021] – contains the time-series of 21 variables transformed to growth rates through taking log-differences (except for interest rates).

```
set.seed(123)
library(bayesianVARs)
variables <- c("GDPC1", "GDPCTPI", "FEDFUNDS", "EXUSUKx", "S&P 500")
train_data <- 100 * usmacro_growth[1:230, variables]
test_data <- 100 * usmacro_growth[231:234, variables]
```

The workhorse function of `bayesianVARs` for conducting MCMC inference is the function `bvar`. Though it offers a low barrier to entry for users (in case only data is supplied without any further sampler and or prior configurations, default values are used), we encourage the user to specify the model to be estimated in more

detail using the helper functions `specify_prior_phi` (prior configuration concerning the VAR coefficients) and `specify_prior_sigma` (prior configuration concerning the variance-covariance of the VAR). In our demonstration, we specify a VAR with $p = 2$ lags with factor stochastic volatility and $r = 4$ factors and a semi-global local HS prior with `olcl-lagwise` partitioning for the VAR coefficients. It is possible to impose standard global local priors by specifying `specify_prior_phi`'s argument `global_grouping = "global"`. An arbitrary grouping for semi-global local priors can be achieved by supplying an indicator matrix to `global_grouping`.

```
prior_phi <- specify_prior_phi(data = train_data,
                              lags = 2L,
                              prior = "HS",
                              global_grouping = "olcl-lagwise")
prior_sigma <- specify_prior_sigma(data = train_data,
                                   type = "factor",
                                   factor_factors = 4L)

mod <- bvar(train_data, lags = 2L, draws = 10000, burnin = 2000,
            prior_phi = prior_phi, prior_sigma = prior_sigma,
            sv_keep = "all")
```

The `plot` method shows the model fit via 90% in-sample prediction intervals by default, see Figure 1.

```
plot(mod, quantiles = c(0.05,0.5,0.95), dates = rownames(mod$Yraw))
```

The object output by `bvar` contains the posterior draws. The extractors `coef` and `vcov` come in handy to access the posterior draws of Φ and Σ_t , respectively. The function `posterior_heatmap` visualizes posterior summaries, such as the posterior median or posterior interquartile-range, as heatmaps, see Figure 2.

```
phi <- coef(mod)
posterior_heatmap(phi, median, detect_lags = TRUE,
                  border_color = rgb(0,0,0, alpha = 0.2))
posterior_heatmap(phi, IQR, detect_lags = TRUE,
                  border_color = rgb(0,0,0, alpha = 0.2))
```

The `predict` method simulates from the posterior predictive distribution. Log-predictive likelihoods will be computed if the ex-post observed data is supplied.

```
pred <- predict(mod, ahead = 1:4, LPL = TRUE, Y_obs = test_data)
```

The `plot` method for draws of the posterior predictive distribution defaults to displaying fan-charts by joining line charts for the observed data of the estimation sample with credible intervals of the posterior predictive distribution, see Figure 3.

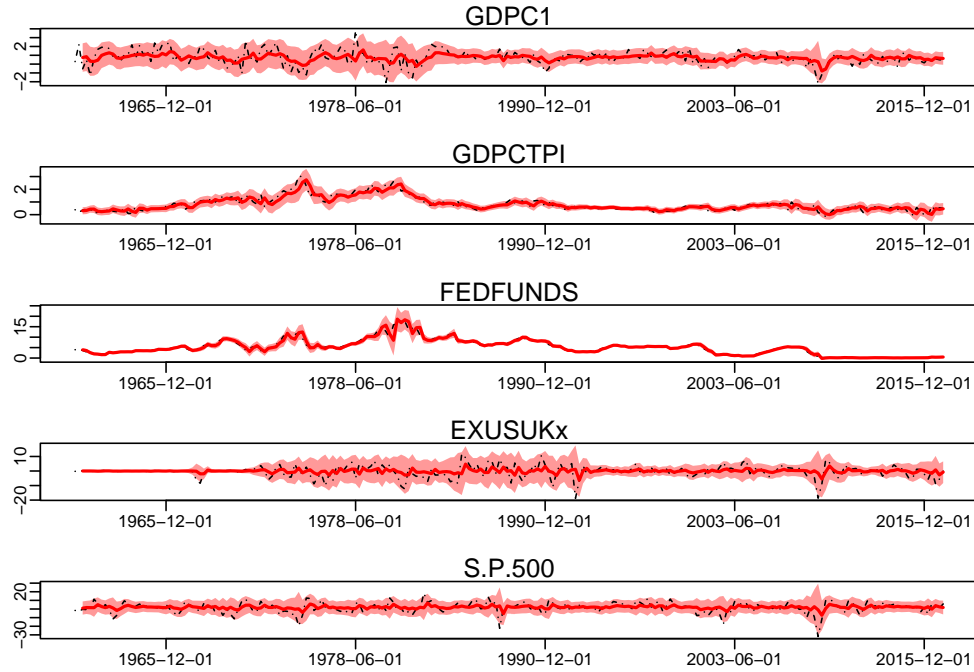


Figure 1: Visualization of estimated in-sample prediction intervals. The red solid line depicts the median, the red shaded region the 90% credible interval and the black dotted line the observed data used for estimation.

```
plot(pred, first_obs = 216,
      dates = c(rownames(train_data[-c(1:215),]), rownames(test_data)))
```

The calculated log-predictive likelihoods could be used to comparing forecasting performances of different models.

```
pred$LPL
#>      t+1      t+2      t+3      t+4
#> -4.603537 -4.962248 -5.588917 -7.235832
```

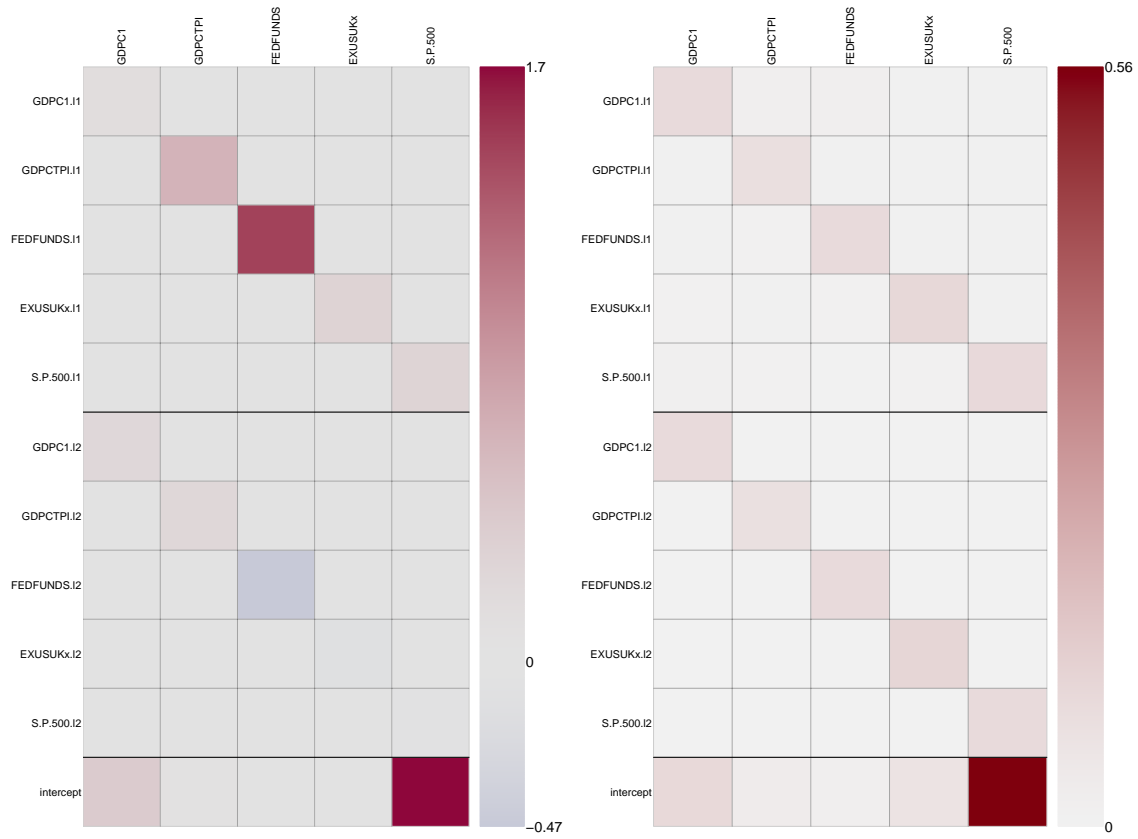


Figure 2: Posterior summary of the VAR coefficients. Left: Heatmap of the posterior median. Right: Heatmap of the posterior interquartile range.

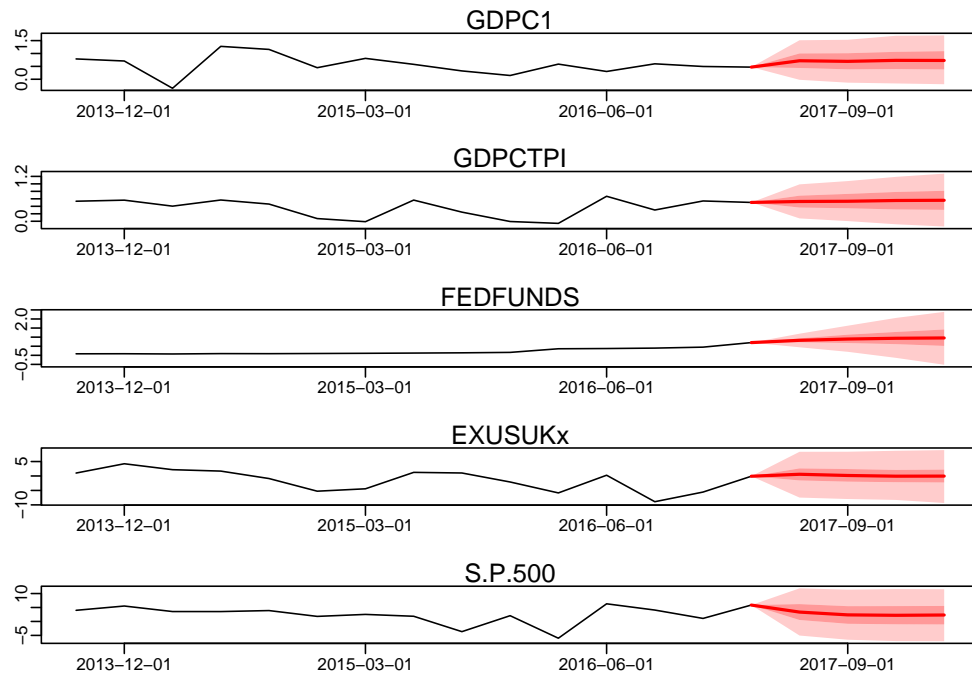


Figure 3: Fan-charts visualizing the last 15 out of 230 observations used for estimation through black solid lines, the median of the h -step ahead predictive distribution through red solid lines and the 50%/90% credible intervals of the h -step ahead predictive distribution through red shaded regions for $h = 1, \dots, 4$.

References

- Anirban Bhattacharya, Debdeep Pati, Natesh S. Pillai, and David B. Dunson. Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490, October 2015. doi: 10.1080/01621459.2014.960967.
- Philip J. Brown and Jim E. Griffin. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010. doi: 10.1214/10-BA507.
- Andrea Carriero, Joshua Chan, Todd E. Clark, and Massimiliano Marcellino. 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, 2022. ISSN 0304-4076. doi: 10.1016/j.jeconom.2021.11.010.
- Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. doi: 10.1093/biomet/asq017.
- Sharon Chiang, Michele Guindani, Hsiang J. Yeh, Zulfi Haneef, John M. Stern, and Marina Vannucci. Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping*, 38(3):1311–1332, 2017. doi: 10.1002/hbm.23456.
- Timothy Cogley and Thomas J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005. doi: 10.1016/j.red.2004.10.009.
- Richard K Crump, Stefano Eusepi, Domenico Giannone, Eric Qian, and Argia M Sbordone. A large Bayesian VAR of the United States economy. *FRB of New York Staff Report*, No. 976, 2021. doi: 10.2139/ssrn.3908154.
- Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011. doi: 10.18637/jss.v040.i08.
- Dirk Eddelbuettel and Conrad Sanderson. Rcpparmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, 2014. doi: 10.1016/j.csda.2013.02.005.
- Lendie Follett and Cindy Yu. Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics*, 11:130–144, July 2019. doi: 10.1016/j.ecosta.2018.12.004.
- Edward I. George, Dongchu Sun, and Shawn Ni. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580, January 2008. doi: 10.1016/j.jeconom.2007.08.017.
- Rainer Goebel, Alard Roebroeck, Dae-Shik Kim, and Elia Formisano. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging*, 21(10):1251–1261, 2003. doi: 10.1016/j.mri.2003.08.026.
- Luis Gruber and Gregor Kastner. Forecasting macroeconomic data with Bayesian VARs: Sparse or dense? It depends! *International Journal of Forecasting*, 41(4):1589–1619, 2025. doi: 10.1016/j.ijforecast.2025.02.001.
- Darjus Hosszejni and Gregor Kastner. Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *Journal of Statistical Software*, 100:1–34, November 2021. doi: 10.18637/jss.v100.i12.

- Florian Huber and Martin Feldkircher. Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, 37(1):27–39, January 2019. doi: 10.1080/07350015.2016.1256217.
- Sune Karlsson. Chapter 15 - forecasting with Bayesian vector autoregression. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 791–897. Elsevier, 2013. doi: 10.1016/B978-0-444-62731-5.00015-4.
- Gregor Kastner. Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(5):1–30, 2016. doi: 10.18637/jss.v069.i05.
- Gregor Kastner. Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics*, 210(1):98–115, 2019. doi: 10.1016/j.jeconom.2018.11.007.
- Gregor Kastner and Sylvia Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76: 408–423, August 2014. doi: 10.1016/j.csda.2013.01.002.
- Gregor Kastner and Florian Huber. Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, 39(7):1142—1165, March 2020. doi: 10.1002/for.2680.
- Gregor Kastner, Sylvia Frühwirth-Schnatter, and Hedibert Freitas Lopes. Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, 26(4): 905–917, 2017. doi: 10.1080/10618600.2017.1322091.
- Robert B. Litterman. Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38, 1986. doi: 10.2307/1391384.
- Michael W. McCracken and Serena Ng. FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis Review*, 103(1):1–44, 2021. doi: 10.20955/r.103.1-44.
- Nicholas G. Polson and James G. Scott. Shrink globally, act locally: Sparse Bayesian regularization and prediction. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting*, pages 501–538, Oxford, UK, 2011. Oxford University Press. doi: 10.1093/acprof:oso/9780199694587.003.0017.
- Christopher A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.
- Bertil Wegmann, Anders Lundquist, Anders Eklund, and Mattias Villani. Bayesian modelling of effective and functional brain connectivity using hierarchical vector autoregressions. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 73(4):835–856, 2024. doi: 10.1093/jrsssc/qlae014.
- Yan Dora Zhang, Brian P. Naughton, Howard D. Bondell, and Brian J. Reich. Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 117 (538):862–874, 2022. doi: 10.1080/01621459.2020.1825449.