Bayesian Modeling and Forecasting of High Dimensional Long Range Dependent Time Series

Bayesian Vector Autoregressive Heterogeneous Modeling[1]

Young Geun Kim

Department of Statistics, Sungkyunkwan University

March 12, 2025

ygeunkimstat@gmail.com ygeunkim.github.io github.com/ygeunkim

[1] with Changryong Baek.

Vector autogressive model

$\{\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{kt})' \in \mathbb{R}^k; t = 1, \ldots, T\}$ follows VAR(p):

$$\mathbf{Y}_t = A_1\mathbf{Y}_{t-1} + A_2\mathbf{Y}_{t-2} + \cdots + A_p\mathbf{Y}_{t-p} + \mathbf{c} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_k, \Sigma_\epsilon)$$

Becomes seemingly unrelated multivariate regression:

$$\underbrace{\begin{pmatrix} \mathbf{Y}'_{p+1} \\ \mathbf{Y}'_{p+2} \\ \vdots \\ \mathbf{Y}'_T \end{pmatrix}}_{\mathbb{Y}_0} = \underbrace{\begin{pmatrix} \mathbf{Y}'_p & \cdots & \mathbf{Y}'_1 & 1 \\ \mathbf{Y}'_{p+1} & \cdots & \mathbf{Y}'_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{Y}'_{T-1} & \cdots & \mathbf{Y}'_{T-p} & 1 \end{pmatrix}}_{\mathbb{X}_0} \underbrace{\begin{pmatrix} A'_1 \\ A'_2 \\ \vdots \\ A'_p \\ \mathbf{c}' \end{pmatrix}}_{\mathbb{A}} + \underbrace{\begin{pmatrix} \boldsymbol{\epsilon}'_{p+1} \\ \boldsymbol{\epsilon}'_{p+2} \\ \vdots \\ \boldsymbol{\epsilon}'_T \end{pmatrix}}_{\mathbb{Z}_0}$$

$$\sim \mathcal{MN}(\mathbb{X}_0\mathbb{A}, I_{T-p}, \Sigma_\epsilon)$$

Gives OLS: $\hat{\mathbb{A}}^{LS} = (\mathbb{X}'_0\mathbb{X}_0)^{-1}\mathbb{X}_0\mathbb{Y}_0$.

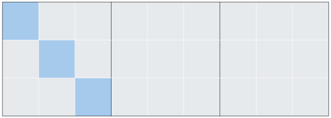
Minnesota prior beliefs [Litterman, 1986]

- Minnesota prior by [Litterman, 1986]: based on the stylized facts of macroeconomic data of US: model using unit root processes
- Minnesota moment:

$$\mathbb{E}\left[(A_\ell)_{ij}\right] = \begin{cases} 1 & j = i, \ell = 1 \\ 0 & \text{otherwise} \end{cases}, \quad \mathrm{Var}\left[(A_\ell)_{ij}\right] = \begin{cases} \frac{\lambda^2}{\ell^2} & j = i \\ \nu \frac{\lambda^2}{\ell^2} \frac{\sigma_i^2}{\sigma_j^2} & \text{otherwise,} \end{cases}$$

  - λ: Overall tightness
  - $\nu \leqslant 1$: Relative tightness - $\lambda > \nu\lambda$ for cross-variable shrinkage
  - $1/l^2$: Lag decay
  - σ_i^2/σ_j^2: Scale factor with $\Sigma_\epsilon = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right)$

- Shrink towards zero for longer lags
- Shrink own-lag less than cross-lag
- [Litterman, 1986]: Ridge regression in the frequentist view
- [Bańbura et al., 2010]: Independent normal-Wishart prior



Themes for the talk

Vector heterogeneous autoregressive (VHAR) models for long-range dependent time series

1. Independent normal-Wishart prior
   - Closed form: MCMC not required
   - Empirical Bayes: hyperparameter selection
2. Adaptive hierarchical priors
   - Cholesky parameterization for contemporaneous effects
   - Heavy tails with large mass at 0: shrink weak signals and keep strong signals
   - Continuous shrinkage priors in the spirit of Minnesota prior: Horseshoe, Normal-gamma, Dirichlet-Laplace, SSVS, hierarchical Minnesota, and Generalized Double Pareto shrinkage

Vector heterogeneous autoregressive model

- Heterogeneous autoregressive (HAR) model for long-range dependent volatility: [Corsi, 2008]
- Multivariate extension: [Bubák et al., 2011]
- Penalization using adaptive lasso: [Baek and Park, 2021]

Daily $\{\mathbf{Y}_t\}$ follows VHAR(5, 22):

$$\mathbf{Y}_t = \Phi^{(d)}\mathbf{Y}_{t-1} + \Phi^{(w)}\mathbf{Y}_{t-1}^{(w)} + \Phi^{(m)}\mathbf{Y}_{t-1}^{(m)} + \mathbf{c} + \epsilon_t,$$

- Weekly aggregation: $\mathbf{Y}_{t-1}^{(w)} := \frac{1}{5}\sum_{i=1}^{5}\mathbf{Y}_{t-i}$
- Monthly aggregation: $\mathbf{Y}_{t-1}^{(m)} := \frac{1}{22}\sum_{i=1}^{22}\mathbf{Y}_{t-i}$

Relation to VAR

VHAR is constrained VAR(22):

$$\mathbf{Y}_t = (\Phi^{(d)} + 5^{-1}\Phi^{(w)} + 22^{-1}\Phi^{(m)})\mathbf{Y}_{t-1} + (5^{-1}\Phi^{(w)} + 22^{-1}\Phi^{(m)})Y_{t-2} + \cdots$$
$$+ (5^{-1}\Phi^{(w)} + 22^{-1}\Phi^{(m)})\mathbf{Y}_{t-5} + 22^{-1}\Phi^{(m)}\mathbf{Y}_{t-6} + \cdots + 22^{-1}\Phi^{(m)}\mathbf{Y}_{t-22} + \mathbf{c} + \epsilon_t,$$

giving:

$$\begin{pmatrix} A_1' \\ A_2' \\ \vdots \\ A_p' \end{pmatrix} = \underbrace{\left[\begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 1/5 & 1/5 & 1/5 & 0 & \ldots & 0 \\ 1/22 & 1/22 & 1/22 & 1/22 & \ldots & 1/22 \end{pmatrix}' \otimes I_k \right]}_{\mathbb{C}_0'} \begin{pmatrix} \Phi^{(d)\prime} \\ \Phi^{(w)\prime} \\ \Phi^{(m)\prime} \end{pmatrix}$$

Multivariate regression:

$$\mathbb{Y}_0 = \mathbb{X}_0 \begin{pmatrix} \mathbb{C}_0' & 0 \\ 0 & 1 \end{pmatrix} \Phi + \mathbb{Z}_0 =: \mathbb{X}_1\Phi + \mathbb{Z}_0 \sim \mathcal{MN}(\mathbb{X}_1\Phi, I_{T-p}, \Sigma_\epsilon)$$

Gives OLS: $\widehat{\Phi}^{LS} = (\mathbb{X}_1'\mathbb{X}_1)^{-1}\mathbb{X}_1'\mathbb{Y}_0$.

Independent normal-Wishart prior

Denote likelihood for reduced form of VHAR is:

$$(\mathbb{Y}_0 \mid \Phi, \Sigma_\epsilon) \sim \mathcal{MN}(\mathbb{X}_1 \Phi, I_n, \Sigma_\epsilon), \quad n = T - p$$

Conjugate prior:

$$(\Phi \mid \Sigma_\epsilon) \sim \mathcal{MN}\left(M_0, \Omega_0, \Sigma_\epsilon\right), \quad (\Sigma_\epsilon) \sim \mathcal{IW}\left(\Psi_0, \nu_0\right)$$

[Bańbura et al., 2010] suggests adding dummy observations to compute Minnesota moments easily:

$$\mathbb{Y}_H := \begin{bmatrix} \operatorname{diag}\left(\delta_1 \sigma_1, \ldots, \delta_k \sigma_k\right)/\lambda \\ 0_{k(p-1) \times k} \\ \operatorname{diag}\left(\sigma_1, \ldots, \sigma_k\right) \\ \mathbf{0}_k^\prime \end{bmatrix}, \quad \mathbb{X}_H := \begin{bmatrix} \operatorname{diag}(1,2,3) \otimes \operatorname{diag}\left(\sigma_1, \ldots, \sigma_k\right)/\lambda & \mathbf{0}_{kp} \\ 0_{k \times kp} & \mathbf{0}_k \\ \mathbf{0}_{kp}^\prime & \epsilon \end{bmatrix}$$

Since this makes $\nu = 1$, use hyperparameter δ_j for own-lag prior mean instead of 1.

Minnesota prior

$$(\Phi \mid \Sigma_\epsilon) \sim \mathcal{MN}\left(M_0, \Omega_0, \Sigma_\epsilon\right), \quad (\Sigma_\epsilon) \sim \mathcal{IW}\left(\Psi_0, \nu_0\right)$$

where

$$\begin{cases} M_0 := \left(\mathbb{X}_H'\mathbb{X}_H\right)^{-1}\mathbb{X}_H'\mathbb{Y}_H, \quad \Omega_0 := \left(\mathbb{X}_H'\mathbb{X}_H\right)^{-1} \\ \mathbb{Z}_H := \mathbb{Y}_H - \mathbb{X}_H M_0, \quad \Psi_0 := \mathbb{Z}_H'\mathbb{Z}_H, \quad \nu_0 := k+2 \end{cases}$$

Closed form of posterior distribution based on augmented matrix:

$$\mathbb{Y}_* = \begin{bmatrix} \mathbb{Y}_0 \\ \mathbb{Y}_H \end{bmatrix}, \quad \mathbb{X}_* = \begin{bmatrix} \mathbb{X}_1 \\ \mathbb{X}_H \end{bmatrix}, \quad \mathbb{Z}_* = \begin{bmatrix} \mathbb{Z}_0 \\ \mathbb{Z}_H \end{bmatrix},$$

$$(\mathbb{A} \mid \Sigma_\epsilon, \mathbb{Y}_0) \sim \mathcal{MN}\left(\widehat{\mathbb{A}}, \left(\mathbb{X}_*'\mathbb{X}_*\right)^{-1}, \Sigma_\epsilon\right), \quad (\Sigma_\epsilon \mid \mathbb{Y}_0) \sim \mathcal{IW}\left(\widehat{\Sigma}, \nu_0 + n\right)$$

where

$$\widehat{\mathbb{A}} = \left(\mathbb{X}_*'\mathbb{X}_*\right)^{-1}\mathbb{X}_*'\mathbb{Y}_*, \quad \widehat{\Sigma} = \left(\mathbb{Y}_* - \mathbb{X}_*\widehat{\Phi}\right)'\left(\mathbb{Y}_* - \mathbb{X}_*\widehat{\Phi}\right).$$

Posterior mean is OLS of augmented regression: $\mathbb{Y}_* = \mathbb{X}_*\Phi + \mathbb{Z}_*$

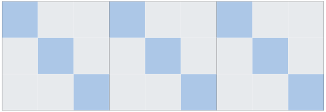
Minnesota Prior in VHAR: [Kim and Baek, 2024]

- Original Minnesota prior above: **BVHAR-S**
- In VHAR: $\Phi^{(w)}$ and $\Phi^{(m)}$ are related to long memory: **BVHAR-L**

$$\mathbb{E}\left(\Phi_{ij}^{(\ell)}\right) = \begin{cases} d_j & j = i, \ell = 1 \\ w_j & j = i, \ell = 2 \\ m_j & j = i, \ell = 3 \\ 0 & \text{otherwise} \end{cases}, \quad \mathrm{Var}\left(\Phi_{ij}^{(\ell)}\right) = \begin{cases} \frac{\lambda^2}{\ell^2} & j = i \\ \frac{\lambda^2}{\ell^2} \frac{\sigma_i^2}{\sigma_j^2} & \text{otherwise.} \end{cases}$$

Use dummy response as

$$\mathbb{Y}_L = \begin{bmatrix} \mathrm{diag}\left(d_1\sigma_1, \ldots, d_k\sigma_k\right)/\lambda \\ \mathrm{diag}\left(w_1\sigma_1, \ldots, w_k\sigma_k\right)/\lambda \\ \mathrm{diag}\left(m_1\sigma_1, \ldots, m_k\sigma_k\right)/\lambda \\ \mathrm{diag}\left(\sigma_1, \ldots, \sigma_k\right) \\ \mathbf{0}_k' \end{bmatrix}.$$



Hyperparameter Selection: [Giannone et al., 2015]

- Choose hyperparameters $\gamma = (d_j, w_j, m_j, \lambda, \sigma_j)'$ in a data-based manner
- Analytical form of BVHAR's marginal likelihood:

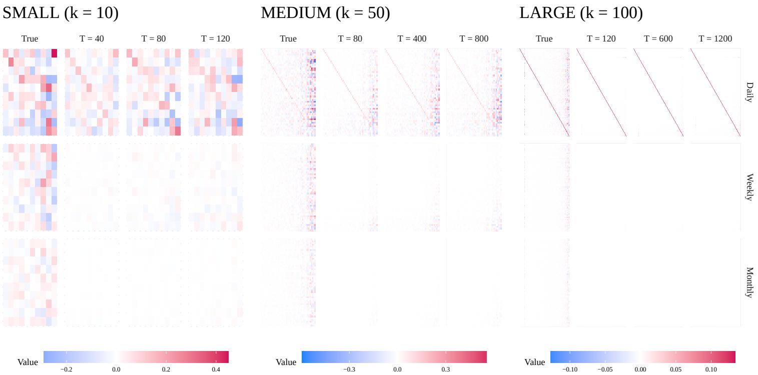
$$[\mathbb{Y}_0] = \pi^{-kn/2} \frac{\Gamma_k\left((\nu_0 + n)/2\right)}{\Gamma_k\left(\nu_0/2\right)} \det(\Omega_0)^{-k/2} \det(\Psi_0)^{\nu_0/2} \det(\mathbb{X}_*'\mathbb{X}_*)^{-k/2} \det(\hat{\Sigma})^{-(\nu_0+n)/2}$$
$$\propto \det(\Omega_0)^{-k/2} \det(\Psi_0)^{\nu_0/2} \det(\mathbb{X}_*'\mathbb{X}_*)^{-k/2} \det(\hat{\Sigma})^{-(\nu_0+n)/2},$$

- $\Omega_0 = L_P L_P'$ and $\Psi_0^{-1} = L_U L_U'$ for numerical stability

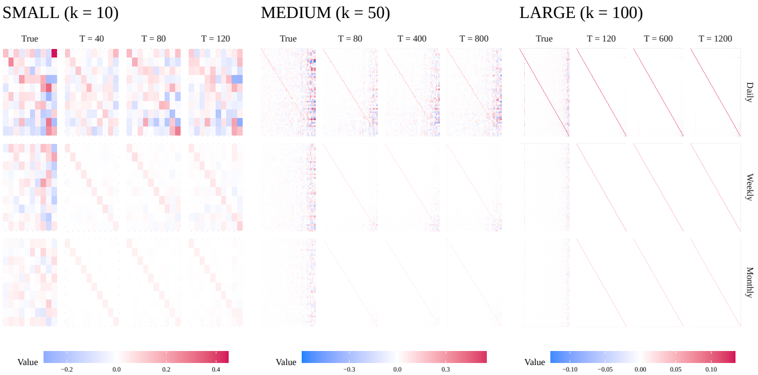
$$\hat{\gamma} = \arg\min_{\gamma} \left\{ \frac{n}{2} \log\det(\Psi_0) + \frac{k}{2} \sum_{i=1}^{3k} \log\left(\tau_i(\gamma) + 1\right) + \frac{\nu_0 + n}{2} \sum_{j=1}^{k} \log\left(\kappa_j(\gamma) + 1\right) \right\}$$

where $\tau_i(\gamma) \in \mathbb{R}$, $i = 1, \ldots, 3k$ are the eigenvalues of $L_P'\mathbb{X}_1'\mathbb{X}_1 L_P$ and $\kappa_j(\gamma) \in \mathbb{R}$, $j = 1, \ldots, k$ are the eigenvalues of $L_U' \left\{ \hat{\mathbb{Z}}_H{}' \hat{\mathbb{Z}}_H + (\hat{\Phi} - M_0)'\Omega_0(\hat{\Phi} - M_0) \right\} L_U$.

BVHAR-S


SMALL (k = 10)
MEDIUM (k = 50)
LARGE (k = 100)
True
T = 40
T = 80
T = 120
True
T = 80
T = 400
T = 800
True
T = 120
T = 600
T = 1200
Daily
Weekly
Monthly
Value
−0.2
0.0
0.2
0.4
Value
−0.3
0.0
0.3
Value
−0.10
−0.05
0.00
0.05
0.10

BVHAR-L


SMALL (k = 10)
MEDIUM (k = 50)
LARGE (k = 100)
True
T = 40
T = 80
T = 120
True
T = 80
T = 400
T = 800
True
T = 120
T = 600
T = 1200
Daily
Weekly
Monthly
Value
−0.2
0.0
0.2
0.4
Value
−0.3
0.0
0.3
Value
−0.10
−0.05
0.00
0.05
0.10

Posterior consistency: [Ghosh et al., 2018]

BVHAR satisfies posterior consistency with MNIW prior:

$$\mathbb{E}_0 \left[\Pi_n \left(\| \Phi - \Phi_0 \| > \epsilon \mid \mathbb{Y}_0 \right) \right] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

under the conditions of [Ghosh et al., 2018], which showed Bayesian VAR's posterior consistency.

Since VHAR's Φ is linear upon VAR's $\mathbb{A}$, the theory can be easily shown in BVHAR.

h-step-ahead Forecasting

Iteratively apply 1-step-ahead forecasts:

$$\mathbf{R}'_T := \begin{pmatrix} \mathbf{Y}'_T & \mathbf{Y}'_{T-1} & \cdots & \mathbf{Y}'_{T-21} & 1 \end{pmatrix} \begin{pmatrix} \mathbb{C}'_0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\hat{\mathbf{R}}'_{T+h-1} := \begin{pmatrix} \hat{\mathbf{Y}}'_{T+h-1} & \cdots & \hat{\mathbf{Y}}'_{T+1} & \mathbf{Y}'_T & \cdots \mathbf{Y}'_{T-h+20} & 1 \end{pmatrix} \begin{pmatrix} \mathbb{C}'_0 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

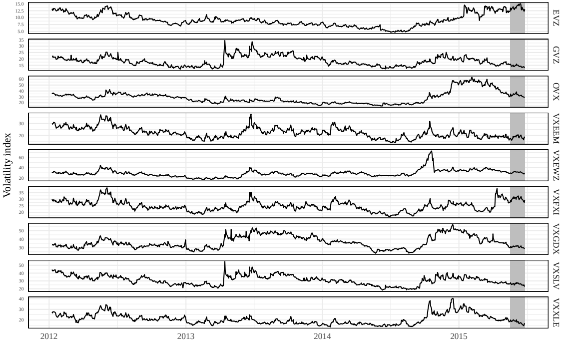
$$(\mathbf{Y}_{T+h} \mid \Sigma, \mathbb{Y}_0) \sim \mathcal{N}\left(\hat{\mathbf{Y}}_{T+h}, \Sigma \otimes \left(1 + \hat{\mathbf{R}}'_{T+h-1}(\mathbb{X}'_*\mathbb{X}_*)^{-1}\hat{\mathbf{R}}_{T+h-1}\right)\right).$$

where

$$\hat{\mathbf{Y}}_{T+h} = (I_k \otimes \hat{\mathbf{R}}'_{T+h-1})\mathrm{vec}(\hat{\Phi}) = \mathrm{vec}\left(\hat{\mathbf{R}}'_{T+h-1}\hat{\Phi}\right). \tag{1}$$

Empirical Study

- Volatility index (VIX): Measures the market's fear level
- Calculate volatility using call and put options on an asset
- Calculate based on individual nine assets (Euro, gold, oil, etc)



Pseudo out-of-sample forecasting performance measures with VAR(3) as benchmark

	RMAFE			RMSFE			RMAPE			RMASE		
	$h = 1$	$h = 5$	$h = 20$	$h = 1$	$h = 5$	$h = 20$	$h = 1$	$h = 5$	$h = 20$	$h = 1$	$h = 5$	$h = 20$
VHAR	0.964	0.895	0.734	0.943	0.799	0.552	0.970	0.891	0.744	0.958	0.875	0.737
BVAR(3)	0.943	0.830	0.703	0.916	0.737	0.494	0.945	0.811	0.718	0.932	0.806	0.710
BVHAR-S	0.945	0.828	0.681	0.915	0.731	0.457	0.947	0.812	0.701	0.934	0.806	0.688
BVHAR-L	0.937	0.798	0.538	0.880	0.679	0.300	0.935	0.773	0.531	0.918	0.787	0.540

Conclusion

- Minnesota prior and VHAR's long memory structure
- Improves forecasting
- Diagonal structure of Σ_ϵ is still possible for structural analysis such as impulse response [Bańbura et al., 2010]
- But nowadays, use a practical setting that has contemporaneous effects in the model, also with heteroskedastic covariance Σ_t

VHAR with Cholesky Parameterization

From [Pourahmadi, 1999, Smith and Kohn, 2002, Cogley and Sargent, 2005],

$$\Sigma^{-1} = L' D^{-1} L$$

Let $\mathbf{X}_t := (\mathbf{Y}'_{t-1}, \ldots, \mathbf{Y}'_{t-22}, 1)' \in \mathbb{R}^{22k+1}$:

$$\mathbf{Y}_t = \Phi' \mathbb{C}_0 \mathbf{X}_t + \boldsymbol{\epsilon}_t = \Phi' \tilde{\mathbf{X}}_t + L^{-1} D^{1/2} \mathbf{u}_t, \quad (\mathbf{u}_t) \sim \mathcal{N}(\mathbf{0}_k, I_k)$$

where

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k-1} & 1 \end{bmatrix}, \quad D = \operatorname{diag}(d_1, d_2, \ldots, d_k).$$

Impose prior:

$$(\phi) \sim \mathcal{N}(\phi_0, S_\phi), \quad (\mathbf{a}) \sim \mathcal{N}(\mathbf{a}_0, S_a), \quad (d_i) \sim \mathcal{G}^{-1}(s_0, s_1), \quad i = 1, \ldots, k$$

Corrected Triangular Algorithm: [Carriero et al., 2022]

- Per-equation: $\mathbf{Y}_j = X_j\phi_j + \mathbf{u}_j, \quad (\mathbf{u}_j) \sim \mathcal{N}(\mathbf{0}_K, I_K)$, where

$$Y_j = \mathrm{vec}\left((\mathbb{Y}_0 - \mathbb{X}_1\Phi^{(j=0)})L'_{j:\ k,:} \oslash D^*\right), \quad X_j := (L_{j:k,j} \otimes \mathbb{X}_1) \oslash D^*,$$

$$D^* := \left(\mathbf{1}_n \otimes (\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_k})\right)_{:\ ,j:k},$$

$$\Phi^{(j=0)} = (\phi_1, \dots, \phi_{j-1}, \mathbf{0}_K, \phi_{j+1}, \dots, \phi_k).$$

- Gives full posterior $(\phi_j \mid -) \sim \mathcal{N}\left(\bar{\mu}, \bar{V}\right)$ where

$$\bar{V}^{-1} = S_{\phi_j}^{-1} + X_j'X_j, \quad \bar{\mu} = \bar{V}(S_{\phi_j}^{-1}\phi_{0,j} + X_j'Y_j).$$

- Precision sampler of [Rue, 2002]
- Orthogonalized innovation regression for $\mathbf{a}$: $D_j^{-1/2}\tilde{\epsilon}_j = D_j^{-1/2}E_j\mathbf{a}_j + \mathbf{v}_j$, $j = 2, \dots, k$ where $\tilde{\epsilon}_j$ is j-th column of $\mathbb{Z}_0$, $\mathbf{a}_j = (a_{j1}, \dots, a_{j,j-1})'$, $E_j = (\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_{j-1}) \in \mathbb{R}^{n\times(j-1)}$, $D_j = \mathrm{diag}(d_j, \dots, d_j) \in \mathbb{R}^{n\times n}$, and $(\mathbf{v}_j) \sim \mathcal{N}(\mathbf{0}_n, I_n)$

Continuous Shrinkage Priors

Recall:

$$(\phi) \sim \mathcal{N}(\mathbf{0}_{K_\phi}, S_\phi), \quad (\mathbf{a}) \sim \mathcal{N}(\mathbf{0}_{K_a}, S_a)$$

- Same prior on both $\phi, \mathbf{a}$
  - Grouped global-local (GGL) priors: Horseshoe, Normal-Gamma (NG), and Dirichlet-Laplace (DL)
  - Stochastic search variable selection (SSVS) prior
  - Minnesota prior
  - Generalized Double Pareto shrinkage
- Normal prior on the constant term [George et al., 2008]

$$(\mathbf{c}) \sim \mathcal{N}(\mathbf{0}_k, cI_k)$$

Grouped Global-Local Priors: [Xu et al., 2017]

$$(\phi_i \mid \tau, \eta_j, \lambda_i) \sim \mathcal{N}(0, \tau^2 \eta_j^2 \lambda_i^2)$$

- τ: global
- η_j: group ($\mathscr{P}_j$) $j = 1, \ldots, 6$
- λ_i: local $i = 1, \ldots, K_\phi$
- Normal-Gamma, Dirichlet-Laplace, and Horseshoe

$$\left[\begin{array}{ccccc} 2 & 1 & \cdots & 1 & 1 \\ 1 & 2 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 2 \\ \hline 4 & 3 & \cdots & 3 & 3 \\ 3 & 4 & \cdots & 3 & 3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 3 & 3 & \cdots & 3 & 4 \\ \hline 6 & 5 & \cdots & 5 & 5 \\ 5 & 6 & \cdots & 5 & 5 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 5 & 5 & \cdots & 5 & 6 \end{array}\right] \sim \left[\begin{array}{cccc} \eta_2 & \eta_1 & \cdots & \eta_1 \\ \eta_1 & \eta_2 & \cdots & \eta_1 \\ \vdots & \vdots & \vdots & \vdots \\ \eta_1 & \cdots & \eta_1 & \eta_2 \\ \hline \eta_4 & \eta_3 & \cdots & \eta_3 \\ \eta_3 & \eta_4 & \cdots & \eta_3 \\ \vdots & \vdots & \vdots & \vdots \\ \eta_3 & \cdots & \eta_3 & \eta_4 \\ \hline \eta_6 & \eta_5 & \cdots & \eta_5 \\ \eta_5 & \eta_6 & \cdots & \eta_5 \\ \vdots & \vdots & \vdots & \vdots \\ \eta_5 & \cdots & \eta_5 & \eta_6 \end{array}\right]$$

Normal-Gamma Prior: [Caron and Doucet, 2008, Griffin and Brown, 2010, Huber and Feldkircher, 2019]

For $\tilde{\lambda}_i^2 = \tau^2 \eta_j^2 \lambda_i^2$ $(i \in \mathscr{P}_j)$,

$$\left(\tilde{\lambda}_i^2 \mid \tau^2, \eta_j^2\right) \sim \mathscr{G}\left(\nu_j, \frac{\nu_j}{\tau^2 \eta_j^2}\right), \quad \left(\eta_j^2\right) \sim \mathscr{G}^{-1}(b_0, b_1), \quad \left(\tau^2\right) \sim \mathscr{G}^{-1}(c_0, c_1)$$

$\Rightarrow$ **Gibbs sampler**

$(\nu_j) \sim \mathscr{E}xp(1) \Rightarrow$ **MH**

- Tails of ϕ_i decreases in

Dirichlet-Laplace Prior: [Bhattacharya et al., 2015, Zhang and Bondell, 2017, Kastner and Huber, 2020]

$$(\phi_i \mid \lambda_i, \rho_j, \tau) \sim \mathscr{DE}(\tau \lambda_i \eta_j)$$

$$\Leftrightarrow \quad (\phi_i \mid \tau, \eta_j, \lambda_i) \sim \mathscr{N}(0, \nu_i \tau^2 \eta_j^2 \lambda_i^2), \quad i = 1, \dots, K_\phi, \quad (\nu_i) \sim \mathscr{Exp}(\frac{1}{2})$$

$$\left((\lambda_1, \dots, \lambda_{K_\phi})'\right) \sim \mathscr{Dir}(\alpha, \dots, \alpha), \quad \left(\eta_j^2\right) \sim \mathscr{G}(c_0, c_1), \quad (\tau) \sim \mathscr{G}(K_\phi \alpha, \frac{1}{2}), \quad i \in \mathscr{P}_j$$

Griddy Gibbs [Ritter and Tanner, 1992] for α on $[K_\phi^{-1}, 0.5]$

Horseshoe Prior: [Carvalho et al., 2009, Carvalho et al., 2010]

$$(\lambda_i) \sim \mathscr{C}^+(0,1), \quad (\eta_j) \sim \mathscr{C}^+(0,1), \quad (\tau) \sim \mathscr{C}^+(0,1), \quad i \in \mathscr{P}_j$$

$\Leftrightarrow$ [Makalic and Schmidt, 2016]:

$$\left(\lambda_i^2 \mid \nu_i\right) \sim \mathscr{G}^{-1}(1/2, 1/\nu_i), \quad \left(\eta_j^2 \mid \rho_j\right) \sim \mathscr{G}^{-1}(1/2, 1/\rho_j), \quad \left(\tau^2 \mid \xi\right) \sim \mathscr{G}^{-1}(1/2, 1/\xi)$$

$$(\nu_i), (\rho_j), (\xi) \sim \mathscr{G}^{-1}(1/2, 1)$$

SSVS Prior: [George and McCulloch, 1993, George et al., 2008, Ishwaran and Rao, 2005]

$$(\phi_i \mid \gamma_i, \tau_{1i}) \sim \underbrace{\mathcal{N}(0, \tau_{1i}^2)}_{\text{Slab}} \gamma_i + \underbrace{\mathcal{N}(0, (c_\tau \tau_{1i})^2)}_{\text{Spike}} (1 - \gamma_i), \quad c_\tau \in (0,1)$$

$$(\gamma_i \mid p_j) \sim \mathcal{B}er(p_j), \quad (p_j) \sim \mathcal{B}eta(s_{1j}, s_{2j}), \quad i \in \mathcal{P}_j, j = 1, \ldots, 6$$

- Full Bayesian approach to choose τ_{1i}^2: $\left(\tau_{1i}^2\right) \sim \mathcal{G}^{-1}(a_0, a_1)$
- Griddy Gibbs for c_τ on $(0, 1)$

Hierarchical Minnesota Prior: [Cross et al., 2020, Chan, 2021, Chan, 2022, Gruber and Kastner, 2022]

$$\mathrm{Var}\left(\Phi_{ij}^{(\ell)}\right) = \begin{cases} \frac{\lambda^2}{\ell^2} & j = i \\ \nu \frac{\lambda^2}{\ell^2} \frac{\sigma_i^2}{\sigma_j^2} & \text{otherwise.} \end{cases}, \quad \mathrm{Var}(a_{ij}) = \lambda^{*2}$$

- Cross-variable shrinkage: $\lambda > \nu\lambda$ to control own and other lags
- Gamma prior on λ and λ^*: $(\lambda) \sim \mathscr{G}(\alpha, \beta)$
- Griddy Gibbs [Ritter and Tanner, 1992] for ν on $(0, 1)$
- **Semi**-hierarchical: no prior for σ_i^2

Generalized Double Pareto Shrinkage Prior: [Armagan et al., 2013]

$$
\begin{aligned}
&(\phi_i \mid \lambda_i) \sim \mathscr{GDP}(\delta/r, r) \\
&\Leftrightarrow [\phi_i \mid r, \delta] = \frac{1}{\delta/r}\left(1 + \frac{1}{r}\frac{|\phi_i|}{\delta/r}\right)^{-(r+1)} \\
&\Leftrightarrow (\phi_i \mid \lambda_i) \sim \mathscr{N}(0, \lambda_i^2), \quad \left(\lambda_i^2 \mid \eta_j\right) \sim \mathscr{E}xp(\eta_j^2/2), \quad (\eta_j) \sim \mathscr{G}(r, \delta), \quad i \in \mathscr{P}_j
\end{aligned}
$$

- Add group structure for η_j
- Griddy Gibbs for $1/(1+r)$ and $1/(1+\delta)$ on $(0, 1)$.

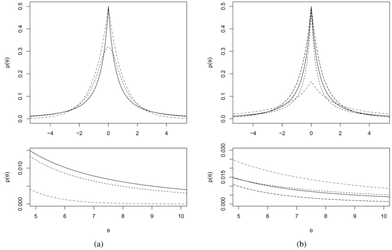


Figure 2.1.
(a) Probability density functions for standard double Pareto (solid line), standard Cauchy (dashed line) and Laplace (dot-dash line) ($\lambda = 1$) distributions. (b) Probability density functions for the generalized double Pareto with (ξ, α) values of (1, 1) (solid line), (0.5, 1) (dashed line), (1, 3) (long-dashed line), and (3, 1) (dot-dash line).

Signal Adaptive Variable Selector: [Ray and Bhattacharya, 2018]

Sparsification via posterior summaries:

$$\phi_j^* = \arg\min_{\phi_j} \left\{ \|\mathbb{X}_1 \hat{\phi}_j - \mathbb{X}_1 \phi_j\|_2^2 + \sum_i \mu_{ij} |\phi_{ij}| \right\},$$

- $\mu_{ij} = 1/|\hat{\phi}_{ij}|^2$: more penalties to the small posterior mean
- Single iteration of coordinate descent
- Sparse draws by applying to each draw [Huber et al., 2021, Hauzenberger et al., 2021]

$$\phi_{ij}^{*(b)} = sign(\phi_{ij}^{(b)}) \|\mathbb{X}_{1i}\|^{-2} \left(|\phi_{ij}^{(b)}| \|\mathbb{X}_{1i}\|^2 - \mu_{ij} \right)_+,$$

- Set zero-penalty to prevent over-shrinkage in own lag:

$$\mu_{ij} = \begin{cases} 0 & \text{own-lag} \\ 1/|\hat{\phi}_{ij}|^2 & \text{otherwise,} \end{cases}$$

Simulation Study

$$(\phi_i) \sim (1-\gamma_i)\mathcal{N}(0, s_{1i}^2) + \gamma_i \mathcal{N}(0, s_{0i}^2), \quad (\gamma_i) \sim \mathscr{Ber}(p_i)$$
$$(a_{ij}) \sim (1-\omega_{ij})\mathcal{N}(0, \kappa_{1ij}^2) + \omega_{ij} \mathcal{N}(0, \kappa_{0ij}^2), \quad (\omega_{ij}) \sim \mathscr{Ber}(q_{ij})$$
$$(d_i) \sim \mathscr{G}^{-1}(3, 1)$$

			Sparse		Intermediate		Dense	
			Non-zero	sd	Non-zero	sd	Non-zero	sd
ϕ	Daily	Own-lag	0.8	0.3	0.8	0.3	0.8	0.25
		Cross-lag	0.02	0.25	0.2	0.1	0.8	0.1
	Weekly	Own-lag	0.8	0.25	0.8	0.2	0.8	0.2
		Cross-lag	0.02	0.2	0.2	0.05	0.8	0.05
	Monthly	Own-lag	0.8	0.2	0.8	0.15	0.8	0.15
		Cross-lag	0.02	0.15	0.2	0.01	0.8	0.01
	a		0.02	0.2	0.2	0.1	0.8	0.01

- $k = 10$, $k = 20$, and $k = 40$
- High-dimension ($3k = K > n = T - 22$), approximately high-dimension, and low dimension

$k = 10$

	DGP		Sparse				Intermediate				Dense			
			Φ		Σ^{-1}		Φ		Σ^{-1}		Φ		Σ^{-1}	
Model	k	T	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse
Minnesota	10	40	6.609	0.603	6.609	0.603	8.129	0.598	8.129	0.598	2.518	**0.299**	2.518	**0.298**
		50	3.608	0.409	3.608	0.408	4.460	0.410	4.460	0.410	1.707	**0.269**	1.707	**0.267**
		80	2.355	0.270	2.355	0.267	2.707	**0.264**	2.707	0.259	1.237	**0.186**	1.237	**0.183**
SSVS		40	3.320	0.734	1.610	0.733	4.148	0.724	1.920	0.723	2.702	0.603	1.326	0.602
		50	3.402	0.611	1.631	0.610	4.235	0.603	1.880	0.601	3.049	0.464	1.469	0.460
		80	3.755	0.370	1.828	0.363	4.576	0.367	2.090	0.355	3.325	0.231	1.600	0.214
Horseshoe		40	4.024	0.479	3.738	0.473	4.624	0.506	4.286	0.496	2.096	0.400	1.857	0.386
		50	1.109	0.431	1.062	0.426	1.090	0.448	1.026	0.438	1.276	0.335	1.179	0.318
		80	1.000	0.345	0.975	0.334	1.029	0.345	0.983	0.319	0.888	0.244	0.850	0.223
NG		40	1.548	0.522	1.362	0.511	1.753	0.603	1.529	0.594	2.352	0.902	1.893	0.895
		50	1.036	0.357	1.018	0.344	1.173	0.442	1.083	0.427	1.669	0.462	1.458	0.453
		80	1.008	**0.253**	0.988	**0.237**	**0.965**	0.278	**0.948**	**0.247**	0.902	0.226	**0.840**	0.207
DL		40	3.411	0.578	3.015	0.578	2.124	0.578	1.793	0.579	2.336	0.572	1.958	0.551
		50	1.017	0.488	0.988	0.491	1.140	0.508	1.030	0.499	1.348	0.470	1.158	0.438
		80	0.992	0.374	0.987	0.364	1.032	0.391	1.007	0.358	0.935	0.330	0.851	0.291
GDP		40	**0.989**	**0.369**	**0.984**	**0.369**	**0.999**	**0.400**	**0.989**	**0.400**	**0.947**	0.307	**0.955**	0.307
		50	**0.982**	**0.318**	**0.981**	**0.318**	**0.984**	**0.351**	**0.981**	**0.351**	**0.933**	0.289	**0.949**	0.289
		80	**0.961**	0.267	**0.966**	0.267	0.965	0.287	0.963	0.287	**0.843**	0.219	0.883	0.219

$k = 20$

	DGP		Sparse				Intermediate				Dense			
			Φ		Σ^{-1}		Φ		Σ^{-1}		Φ		Σ^{-1}	
Model	k	T	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse
Minnesota	20	50	4.894	0.442	4.894	0.442	4.419	0.430	4.419	0.430	1.701	**0.352**	1.701	**0.352**
		80	2.654	0.298	2.654	0.297	2.502	**0.302**	2.502	**0.301**	1.139	**0.267**	1.139	**0.267**
		140	2.205	0.236	2.205	0.235	1.707	0.283	1.707	0.275	0.906	**0.187**	**0.906**	**0.187**
SSVS		50	2.838	0.749	1.184	0.749	2.985	0.751	1.173	0.751	1.829	0.791	1.039	0.791
		80	4.352	0.533	1.764	0.530	4.517	0.538	1.765	0.533	2.849	0.596	1.473	0.594
		140	5.357	0.320	2.466	0.306	5.517	0.372	2.445	0.347	3.418	0.299	1.746	0.289
Horseshoe		50	1.314	0.458	1.139	0.437	1.424	0.497	1.234	0.472	1.136	0.576	1.049	0.572
		80	1.633	0.409	1.522	0.387	1.403	0.446	1.265	0.416	1.034	0.444	1.055	0.438
		140	1.599	0.332	1.398	0.302	1.311	0.391	1.123	0.351	0.860	0.303	0.937	0.293
NG		50	1.117	0.423	1.002	0.408	1.596	0.550	1.403	0.535	1.130	0.699	1.075	0.696
		80	1.392	0.314	1.301	0.299	1.677	0.359	1.584	0.331	0.874	0.287	0.954	0.280
		140	1.047	0.262	0.975	0.243	1.012	0.319	0.964	0.286	0.772	0.211	0.930	0.203
DL		50	3.406	0.787	2.696	0.691	3.811	0.801	3.092	0.705	2.235	0.961	1.836	0.863
		80	1.833	0.527	1.609	0.496	1.806	0.554	1.528	0.515	1.209	0.607	1.126	0.597
		140	1.650	0.411	1.392	0.367	1.345	0.464	1.085	0.409	0.865	0.405	0.932	0.389
GDP		50	**0.971**	**0.294**	**0.931**	**0.294**	**0.967**	**0.335**	**0.954**	**0.335**	**0.905**	0.409	**0.967**	0.409
		80	**0.963**	**0.278**	**0.900**	**0.278**	**0.939**	0.308	**0.921**	0.308	**0.835**	0.296	**0.946**	0.296
		140	**0.947**	**0.227**	**0.871**	**0.227**	**0.900**	**0.267**	**0.884**	**0.267**	**0.759**	0.209	0.924	0.209

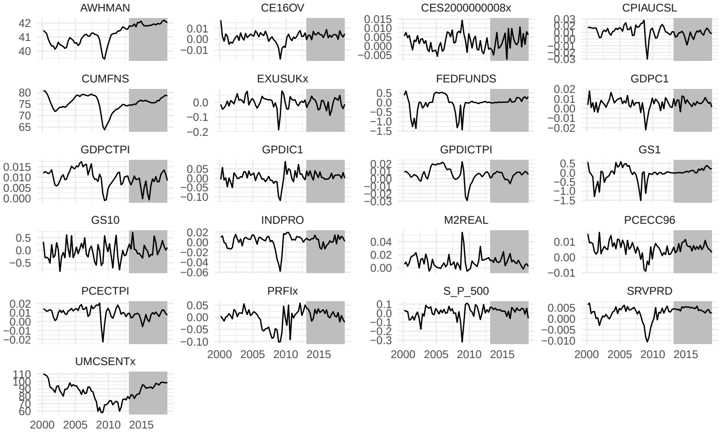
$k = 40$

	DGP		Sparse				Intermediate				Dense			
			Φ		Σ^{-1}		Φ		Σ^{-1}		Φ		Σ^{-1}	
Model	k	T	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse
Minnesota	40	80	3.271	0.321	3.271	0.321	3.141	0.345	3.141	0.344	0.982	**0.289**	0.982	**0.289**
		140	1.875	0.196	1.875	0.195	1.794	0.240	1.794	0.237	0.806	**0.212**	**0.806**	**0.212**
		260	1.351	0.168	1.351	0.165	1.301	0.234	1.301	0.222	0.695	**0.197**	**0.695**	**0.197**
SSVS		80	2.702	0.772	1.183	0.772	2.728	0.777	1.204	0.777	1.508	0.882	0.963	0.882
		140	4.753	0.525	1.990	0.523	4.878	0.542	2.004	0.539	2.699	0.637	1.429	0.636
		260	6.379	0.260	2.984	0.241	6.448	0.317	2.946	0.287	3.416	0.296	1.810	0.285
Horseshoe		80	1.413	0.415	1.346	0.393	1.155	0.452	1.094	0.428	0.965	0.619	1.009	0.617
		140	0.984	0.335	0.973	0.307	1.019	0.387	1.005	0.353	0.825	0.441	0.971	0.436
		260	0.865	0.281	0.883	0.248	0.835	0.346	0.899	0.297	0.684	0.328	0.927	0.319
NG		80	1.141	0.434	1.056	0.428	1.202	0.436	1.116	0.427	0.847	0.335	0.964	0.333
		140	0.914	0.244	0.890	0.234	0.887	0.288	0.895	0.271	0.738	0.243	0.951	0.239
		260	0.905	0.206	0.898	0.193	0.817	0.266	0.875	0.241	**0.644**	0.228	0.933	0.222
DL		80	4.047	1.458	3.574	1.066	4.423	1.452	3.842	1.136	3.636	2.047	3.175	1.882
		140	1.221	0.498	1.166	0.439	1.159	0.529	1.097	0.476	0.931	0.663	1.009	0.659
		260	0.882	0.369	0.879	0.311	0.879	0.428	0.897	0.359	0.716	0.469	0.930	0.457
GDP		80	**0.911**	**0.238**	**0.890**	**0.238**	**0.932**	**0.278**	**0.933**	**0.278**	**0.813**	0.309	**0.950**	0.309
		140	**0.856**	**0.186**	**0.844**	**0.186**	**0.868**	**0.232**	**0.885**	**0.232**	**0.733**	0.228	0.937	0.228
		260	**0.819**	**0.159**	**0.814**	**0.159**	**0.810**	**0.215**	**0.864**	**0.215**	0.653	0.204	0.922	0.204

Empirical Study

Quarterly macroeconomic time series [McCracken and Ng, 2020] with

- Train: 2000 Q1 to 2012 Q4 (52 points)
- Validation: 2013 Q1 to 2018 Q4 (24 points, shaded)
- 21 variables selected by [Gruber and Kastner, 2022]



Pseudo out-of-sample forecasting performance with dense BVHAR-DL as benchmark

			RMAFE			RMSFE		
			$h=1$	$h=5$	$h=20$	$h=1$	$h=5$	$h=20$
BVAR(3)	DL	dense	1.002×10^{-6}	2.682×10^{-6}	0.000 215 9	1.343×10^{-12}	5.906×10^{-12}	3.501×10^{-8}
		sparse	9.051×10^{-7}	2.614×10^{-7}	3.407×10^{-7}	1.269×10^{-12}	2.199×10^{-13}	2.932×10^{-13}
	GDP	dense	3.475×10^{-30}	3.536×10^{-31}	1.18×10^{-32}	1.074×10^{-60}	6.912×10^{-62}	9.663×10^{-65}
		sparse	3.474×10^{-30}	3.497×10^{-31}	1.126×10^{-32}	1.066×10^{-60}	6.702×10^{-62}	8.814×10^{-65}
	Horseshoe	dense	5.495×10^{-30}	3.57×10^{-29}	9.97×10^{-31}	1.014×10^{-59}	4.472×10^{-57}	2.412×10^{-60}
		sparse	4.957×10^{-30}	1.66×10^{-29}	9.23×10^{-31}	7.004×10^{-60}	8.768×10^{-58}	3.808×10^{-60}
	Minnesota	dense	3.396×10^{-30}	3.486×10^{-31}	3.57×10^{-30}	1.024×10^{-60}	6.698×10^{-62}	1.554×10^{-59}
		sparse	3.382×10^{-30}	3.572×10^{-31}	1.272×10^{-32}	1.002×10^{-60}	6.972×10^{-62}	1.095×10^{-64}
	NG	dense	3.476×10^{-30}	3.534×10^{-31}	1.166×10^{-32}	1.075×10^{-60}	6.91×10^{-62}	9.49×10^{-65}
		sparse	3.478×10^{-30}	3.528×10^{-31}	1.16×10^{-32}	1.073×10^{-60}	6.846×10^{-62}	9.322×10^{-65}
	SSVS	dense	3.482×10^{-30}	3.569×10^{-31}	1.427×10^{-25}	1.078×10^{-60}	7.013×10^{-62}	6.135×10^{-51}
		sparse	3.483×10^{-30}	3.533×10^{-31}	1.667×10^{-32}	1.076×10^{-60}	6.868×10^{-62}	1.938×10^{-64}
BVHAR	DL	sparse	1.009	0.3047	0.026 56	0.8846	0.167	0.001 107
	GDP	dense	3.476×10^{-30}	3.528×10^{-31}	1.168×10^{-32}	1.074×10^{-60}	6.87×10^{-62}	9.483×10^{-65}
		sparse	3.479×10^{-30}	3.496×10^{-31}	$1.123\,624 \times 10^{-32}$	1.064×10^{-60}	6.657×10^{-62}	$8.703\,186 \times 10^{-65}$
	Horseshoe	dense	3.019×10^{-25}	5.823×10^{-26}	1.199×10^{-32}	1.06×10^{-49}	2.064×10^{-50}	1.077×10^{-64}
		sparse	3.769×10^{-25}	1.937×10^{-26}	1.192×10^{-32}	2.269×10^{-49}	4.264×10^{-51}	9.766×10^{-65}
	Minnesota	dense	3.369×10^{-30}	3.461×10^{-31}	6.773×10^{-29}	1.007×10^{-60}	6.594×10^{-62}	6.991×10^{-57}
		sparse	$3.355\,609 \times 10^{-30}$	$3.344\,033 \times 10^{-31}$	1.236×10^{-32}	$9.926\,776 \times 10^{-61}$	$6.073\,716 \times 10^{-62}$	1.088×10^{-64}
	NG	dense	3.485×10^{-30}	3.551×10^{-31}	1.151×10^{-32}	1.081×10^{-60}	6.983×10^{-62}	9.257×10^{-65}
		sparse	3.491×10^{-30}	3.558×10^{-31}	1.146×10^{-32}	1.08×10^{-60}	6.944×10^{-62}	9.058×10^{-65}
	SSVS	dense	3.477×10^{-30}	3.579×10^{-31}	3.141×10^{-24}	1.075×10^{-60}	6.986×10^{-62}	3.866×10^{-48}
		sparse	3.48×10^{-30}	3.554×10^{-31}	2.715×10^{-32}	1.075×10^{-60}	6.97×10^{-62}	7.392×10^{-64}

- BVHAR(4, 20) is better than BVAR(3)
- Minnesota prior is good at short-term forecasting (Quarter-on-quarter and Year-on-year)
- Sparsified GDP prior is the best in long-term forecasting (5 YoYs)

Conclusion

- Implements shrinkage priors in BVHAR framework
- Minnesota-type GDP prior works well with time series forecasting
- SAVS improved macroeconomic forecasts
- Ongoing
  - In macroeconomic forecasting, the sample size was too small to capture long memory: need a longer set
  - Financial data: 15 CBOE VIX series

Summary

- Long memory adaptive Minnesota prior with VHAR
  - Posterior consistency
  - Enhance forecasting
- Group structure usage for VHAR in shrinkage priors
- Sparsification of individual draws using SAVS

Future works

- Software package: R and Python packages based on C++ codes (R package `bvhar` available in CRAN)
- Volatility spillovers [Diebold and Yilmaz, 2012, Diebold and Yılmaz, 2014, Deng and Matteson, 2022]
- Stationarity: Force stable MCMC draws to get good forecast values [Heaps, 2023]
- Time series variable selection: Choosing exogenous variables using row-wise sparsity [Bai and Ghosh, 2018]
- Time-varying coefficients and volatilities: $\Sigma_t^{-1} = L' D_t^{-1} L$
- Matrix-valued time series [Chan and Qi, 2024]

References I

[Armagan et al., 2013] Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double pareto shrinkage. *Statistica Sinica*, 23(1):119–143.

[Baek and Park, 2021] Baek, C. and Park, M. (2021). Sparse vector heterogeneous autoregressive modeling for realized volatility. *Journal of the Korean Statistical Society*, 50(2):495–510.

[Bai and Ghosh, 2018] Bai, R. and Ghosh, M. (2018). High-dimensional multivariate posterior consistency under global–local shrinkage priors. *Journal of Multivariate Analysis*, 167:157–170.

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

[Bhattacharya et al., 2015] 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.

[Bubák et al., 2011] Bubák, V., Kočenda, E., and Žikeš, F. (2011). Volatility transmission in emerging european foreign exchange markets. *Journal of Banking & Finance*, 35(11):2829–2841.

[Caron and Doucet, 2008] Caron, F. and Doucet, A. (2008). Sparse bayesian nonparametric regression. *Proceedings of the 25th international conference on Machine learning - ICML '08*, page 88–95.

[Carriero et al., 2022] 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.

[Carvalho et al., 2009] Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial intelligence and statistics*, pages 73–80. PMLR.

[Carvalho et al., 2010] Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

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

[Chan, 2022] Chan, J. C. (2022). Comparing stochastic volatility specifications for large bayesian vars. *Journal of Econometrics*.

[Chan and Qi, 2024] Chan, J. C. and Qi, Y. (2024). Large bayesian matrix autoregressions. *Available at SSRN 4855762*.

References II

[Cogley and Sargent, 2005] 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.

[Corsi, 2008] Corsi, F. (2008). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.

[Cross et al., 2020] 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.

[Deng and Matteson, 2022] Deng, G. and Matteson, D. S. (2022). Bayesian spillover graphs for dynamic networks. *arXiv*.

[Diebold and Yilmaz, 2012] Diebold, F. X. and Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.

[Diebold and Yılmaz, 2014] Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

[George and McCulloch, 1993] George, E. I. and McCulloch, R. E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423):881–889.

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

[Ghosh et al., 2018] Ghosh, S., Khare, K., and Michailidis, G. (2018). High-dimensional posterior consistency in bayesian vector autoregressive models. *Journal of the American Statistical Association*.

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

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

[Gruber and Kastner, 2022] Gruber, L. and Kastner, G. (2022). Forecasting macroeconomic data with bayesian vars: Sparse or dense? it depends! *arXiv*.

References III

[Hauzenberger et al., 2021] Hauzenberger, N., Huber, F., and Onorante, L. (2021). Combining shrinkage and sparsity in conjugate vector autoregressive models. *Journal of Applied Econometrics*, 36(3):304–327.

[Heaps, 2023] Heaps, S. E. (2023). Enforcing stationarity through the prior in vector autoregressions. *Journal of Computational and Graphical Statistics*, 32(1):74–83.

[Huber and Feldkircher, 2019] Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, 37(1):27–39.

[Huber et al., 2021] Huber, F., Koop, G., and Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 39(3):669–683.

[Ishwaran and Rao, 2005] Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and bayesian strategies. *The Annals of Statistics*, 33(2).

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

[Kim and Baek, 2024] Kim, Y. G. and Baek, C. (2024). Bayesian vector heterogeneous autoregressive modeling. *Journal of Statistical Computation and Simulation*, 94(6):1139–1157.

[Litterman, 1986] Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

[Makalic and Schmidt, 2016] Makalic, E. and Schmidt, D. F. (2016). A Simple Sampler for the Horseshoe Estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

[McCracken and Ng, 2020] McCracken, M. and Ng, S. (2020). Fred-qd: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

[Pourahmadi, 1999] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86(3):677–690.

[Ray and Bhattacharya, 2018] Ray, P. and Bhattacharya, A. (2018). Signal adaptive variable selector for the horseshoe prior. *arXiv*.

References IV

[Ritter and Tanner, 1992] Ritter, C. and Tanner, M. A. (1992). Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868.

[Rue, 2002] Rue, H. (2002). Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2):325–338.

[Smith and Kohn, 2002] Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97(460):1141–1153.

[Xu et al., 2017] Xu, Z., Schmidt, D. F., Makalic, E., Qian, G., and Hopper, J. L. (2017). Bayesian sparse global-local shrinkage regression for selection of grouped variables. *arXiv*.

[Zhang and Bondell, 2017] Zhang, Y. and Bondell, H. D. (2017). Variable selection via penalized credible regions with dirichlet–laplace global-local shrinkage priors. *Bayesian Analysis*, 13(3).

Thanks for your attention!

Any questions?