Bayesian Vector Autoregressive Heterogeneous Modeling

KSS 2024, 04 July 2024

Young Geun Kim Changryong Baek

Department of Statistics, Sungkyunkwan University

Table of contents

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

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:

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

- 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):

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

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 likeihood for reduced form of VAR is:

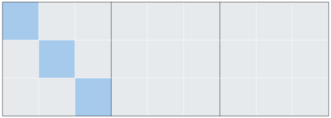
$$(\mathbb{Y}_0 \mid \Phi, \Sigma_\epsilon) \sim \mathcal{MN}(\mathbb{X}_0 \mathbb{A}, I_n, \Sigma_\epsilon), \quad n = T - p$$

Conjugate prior:

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

Minnesota prior by [Litterman, 1986]: based on the stylized facts of macroeconomic data of US: model using univariate random walk processes

- Assume diagonal $\Sigma_\epsilon := \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right)$
- Minnesota moment for coefficient
  - Shrink towards zero for longer lags
  - Shrink own-lag less than cross-lag
- Construct MNIW prior using Minnesota moment



Minnesota moment

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

- 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

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

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

Minnesota prior

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

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:

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

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

where

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

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

Minnesota Prior in VHAR

Apply Minnesota prior in BVAR the same: **BVHAR-S**

$$\mathbb{X}_* = \begin{bmatrix} \mathbb{X}_1 \\ \mathbb{X}_H \end{bmatrix}$$

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

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

Use dummy response as

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

Hyperparameter Selection

[Giannone et al., 2015] suggested prior selection method based on the analytical form of BVAR's marginal likelihood. For BVHAR:

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

Cholesky decomposition for $\Omega_0 = L_P L_P'$ and $\Psi_0^{-1} = L_U L_U'$ gives numerically stable form:

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

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 $L_U' \left\{ \hat{\mathbb{Z}}_H{}' \hat{\mathbb{Z}}_H + (\hat{\Phi} - M_0)' \Omega_0 (\hat{\Phi} - M_0) \right\} L_U$.

Posterior consistency

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.

Simulation Study

Generate true

$$(\Phi_0, \Sigma_{\epsilon,0}) \sim \mathscr{MNIW}(M_0, \Omega_0, \Psi_0, \nu_0)$$

where $\sigma_i^2 = 1$ and $\delta_i = 0.1$. 100 samples from the following data generating processes (DGPs):

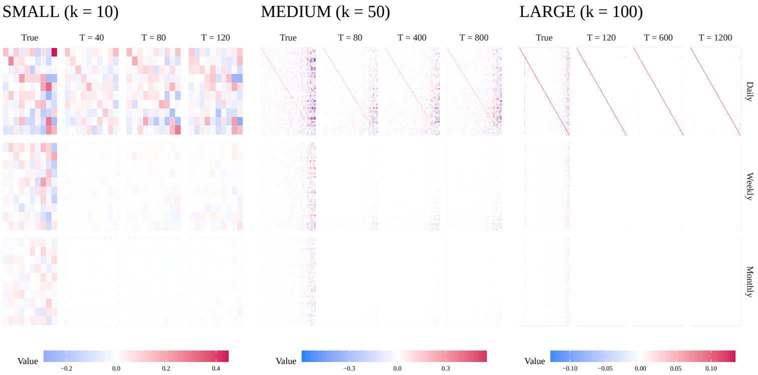
- **SMALL** ($k = 10$) BVHAR-S with $\lambda = 0.2$
- **MEDIUM** ($k = 50$) BVHAR-S with $\lambda = 0.1$
- **LARGE** ($k = 100$) BVHAR-S with $\lambda = 0.01$

with three different sample sizes with $T/k = 4, 8, 12$.

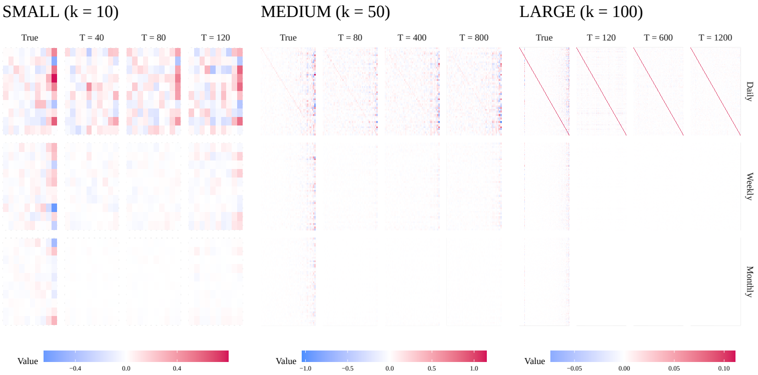
Two innovation distributions:

- $(\boldsymbol{\epsilon}_t) \sim \mathcal{N}(\mathbf{0}_k, \Sigma_{\boldsymbol{\epsilon},0})$
- $(\boldsymbol{\epsilon}_t) \sim t_\nu(\mathbf{0}_k, V_0^{-1/2} \Sigma_{\boldsymbol{\epsilon},0} V_0^{-1/2})$ where V_0 is the diagonal matrix whose elements are the diagonals of above $\Sigma_{\boldsymbol{\epsilon},0}$

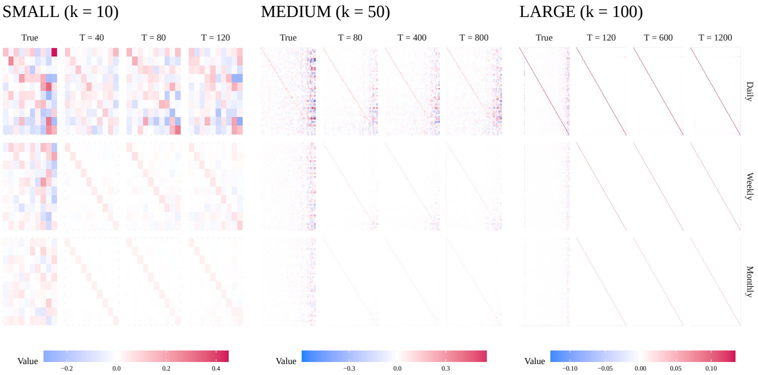
BVHAR-S


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

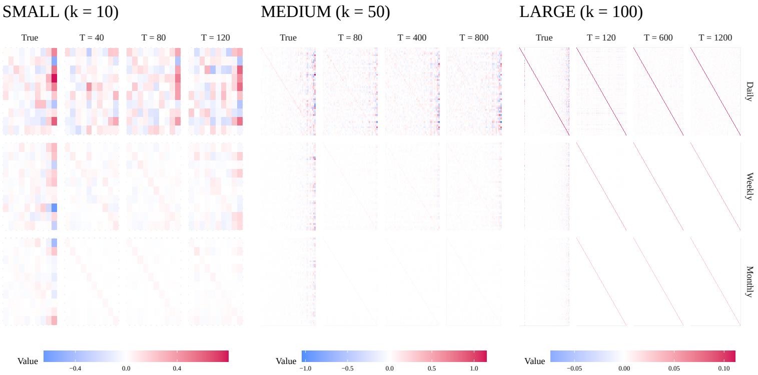
BVHAR-S in MVT



BVHAR-L


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

BVHAR-L in MVT



Relative Estimation Error $\frac{\|\widehat{\Phi}-\Phi_0\|}{\|\Phi_0\|}$ and Standard error of $\|\widehat{\Phi}\|$

		Normal		MVT(df=3)	
k	T	BVHAR-S	BVHAR-L	BVHAR-S	BVHAR-L
SMALL	40	0.936 (0.0705)	0.944 (0.0706)	0.866 (0.2927)	0.866 (0.2905)
	80	0.874 (0.0714)	0.882 (0.0705)	0.765 (0.1684)	0.764 (0.1689)
	120	0.839 (0.0783)	0.852 (0.0782)	0.707 (0.1520)	0.706 (0.1511)
MEDIUM	200	0.886 (0.1915)	0.886 (0.1903)	0.883 (0.2989)	0.885 (0.2981)
	400	0.846 (0.1467)	0.846 (0.1464)	0.855 (0.3211)	0.856 (0.3208)
	600	0.838 (0.1456)	0.837 (0.1455)	0.852 (0.2317)	0.853 (0.2315)
LARGE	400	0.978 (0.0126)	0.985 (0.0110)	0.966 (0.0276)	0.975 (0.0258)
	800	0.976 (0.0190)	0.981 (0.0168)	0.947 (0.0271)	0.956 (0.0256)
	1200	0.975 (0.0120)	0.980 (0.0106)	0.933 (0.0490)	0.942 (0.0481)

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

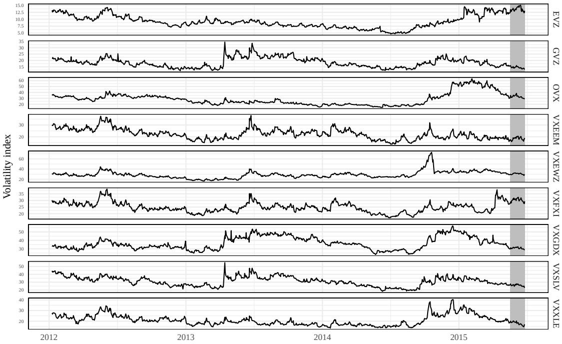
$$(\mathbf{Y}_{T+h} \mid \Sigma, \mathbb{Y}_0) \sim \mathcal{N}\left(\hat{\mathbf{Y}}_{T+h}, \Sigma \otimes \left(1 + \hat{\mathbf{R}}_{T+h-1}' (\mathbb{X}_*' \mathbb{X}_*)^{-1} \hat{\mathbf{R}}_{T+h-1} \right) \right).$$

where

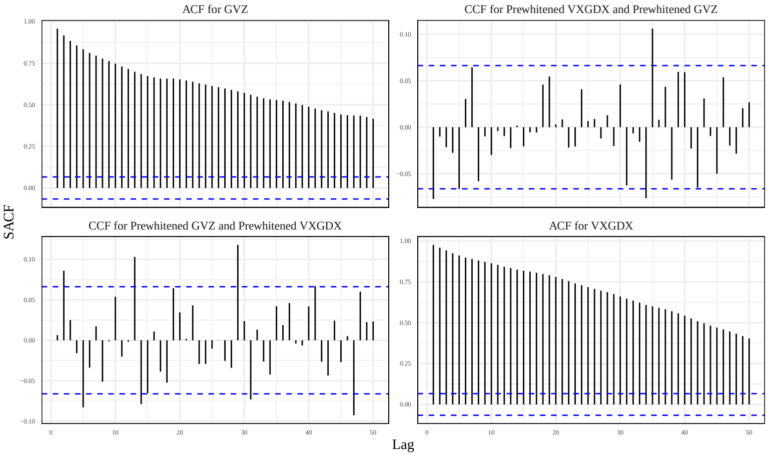
$$\hat{\mathbf{Y}}_{T+h} = (I_k \otimes \hat{\mathbf{R}}_{T+h-1}') \mathrm{vec}(\hat{\Phi}) = \mathrm{vec}\left(\hat{\mathbf{R}}_{T+h-1}' \hat{\Phi} \right). \tag{1}$$

Empirical Study

- Volatility index (VIX): Measures the market's fear level
- Calculate volatility using call and put options on S&P500
- Use ETFs (Euro, gold, oil, etc) instead of S&P500



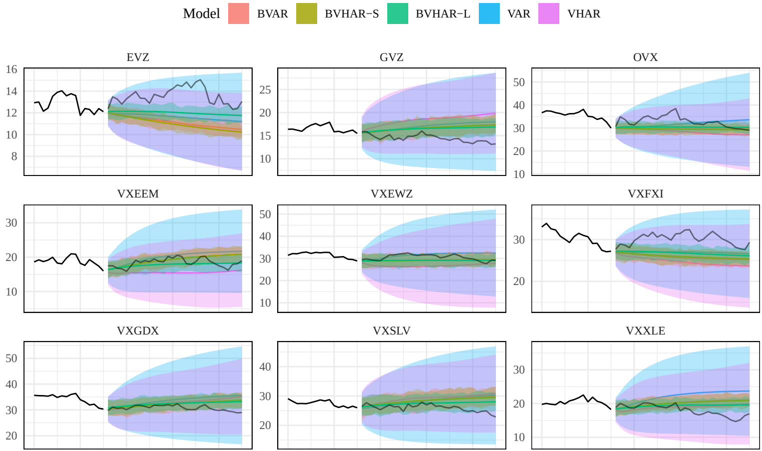
Long-Range Dependency



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

Forecasting Interval



Conclusion

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

Future studies

- Other shrinkage priors: Horseshoe prior, SSVS prior, (hieararchical) Minnesota prior
- $\Sigma^{-1} = L' D^{-1} L$: same shrinkage priors on contemporaneous coefficients (off-diagonal of L)

References I

[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.

[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.

[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.

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

[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.

[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.

Thanks for your attention!

Any questions?

- Based on Kim, Y. G. and Baek, C. (2024). Bayesian vector heterogeneous autoregressive modeling.
  Journal of Statistical Computation and Simulation, 94(6):1139–1157
- For codes,
  - R package available in CRAN: `bvhar`
  - Codes for this study: ygeunkim/paper-bvhar