High-Dimensional Inference for Sparse Vector Autoregressions (Sparse VAR)

Dananjani Liyanage* & Mihai Giurcanu† & Alex Trindade*

(†) *Dept. of Public Health Sciences, University of Chicago*

(∗) *Dept. of Mathematics & Statistics, Texas Tech University*

October 2024

The data, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, is postulated to come from an appropriate model class, $\mathcal{M} = \{f(\boldsymbol{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$.

- **Model Selection***: apply sound principles to select an appropriate element from $\mathcal{M}$; must resolve the bias-variance tradeoff problem...
  - use an information criterion (AIC/BIC); etc.
- **Model Fitting:** straightforward minimization of an appropriate criterion to estimate $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\theta}}$:
  - least-squares (LSE); maximum likelihood (MLE); etc.
- **Model Checking:** residuals from the fitted ("true") model have predictable properties:
  - independence; homoscedasticity; (asymptotically) normal; etc.
- **Inference***: Confidence intervals & hypothesis tests on functions of $\boldsymbol{\theta}$; usually based on an asymptotic normality result:

$$\sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

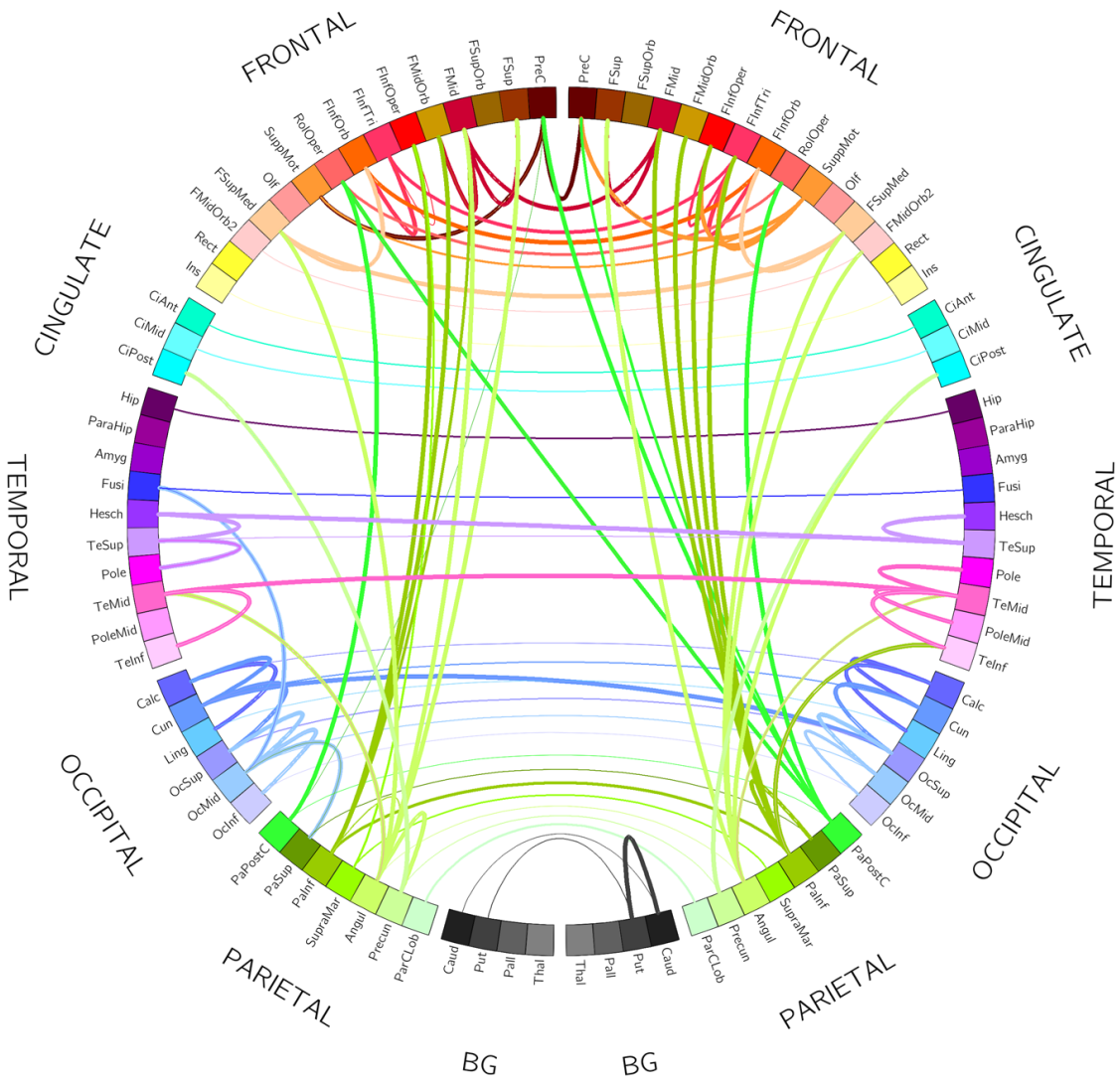
Outline

- **VAR Models:** Ubiquitous in (stationary) time series modeling.
- **Goal:** Model MANY series simultaneously **(high dimensional)**

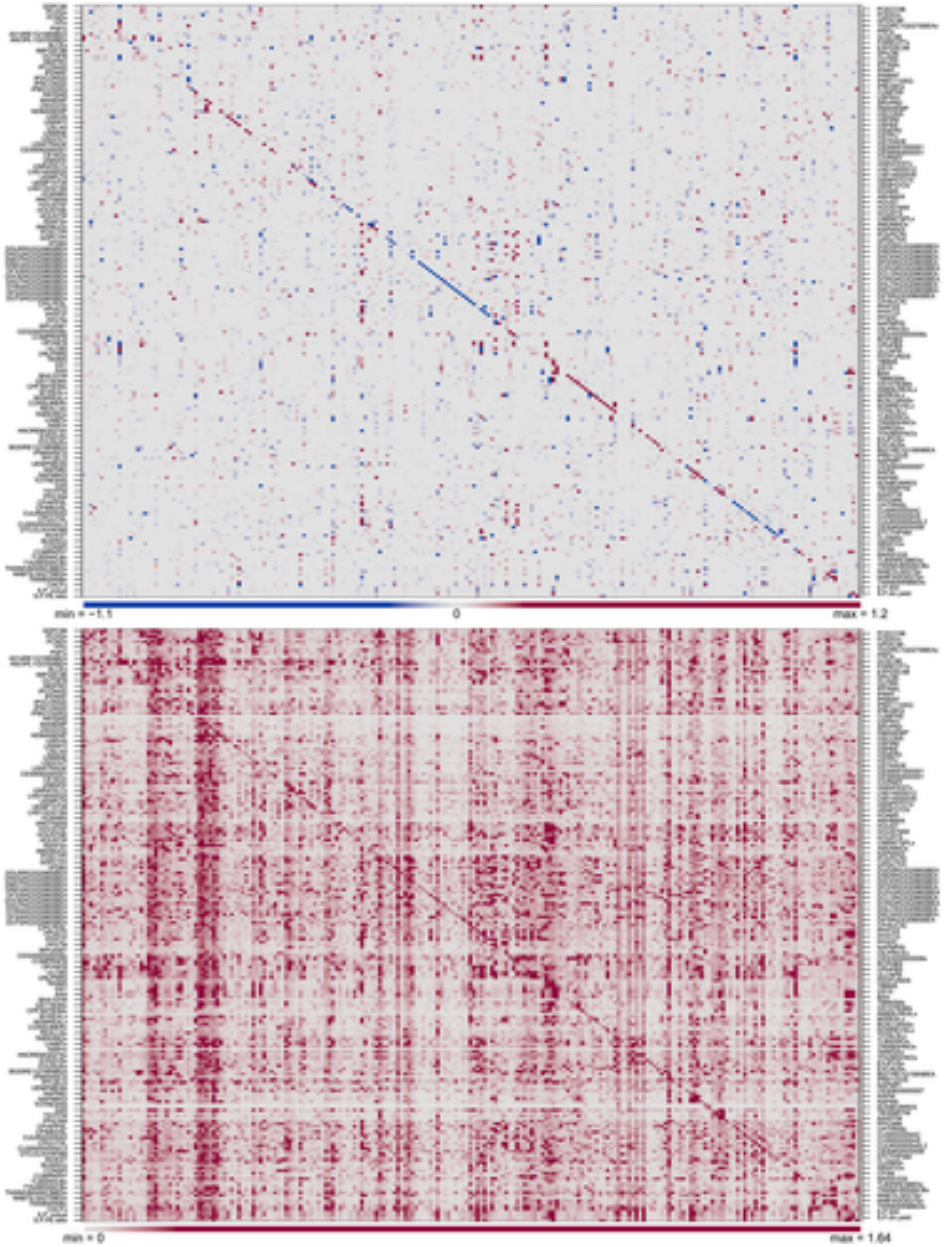
$$\text{VAR dimension} = d$$

- **Existing Methods:** Lasso; Elasticnet; SCAD; MCP; shrinkage priors (Bayesian)
- **R Packages:** `sparsevar`; `BigVAR`
- **New Methods:** p-value thresholding; BIC-based forward selection
- **Simulations:** Compare sparsity pattern recovery and accuracy
- **Illustrations:** fMRI data, S&P 500 Data

Information Flow in the Brain (Duggento et al, 2018): $d = 116$ regions



Macroeconomic Forecasting (Kastner & Huber, 2019): $d = 215$ quantities



$$\boldsymbol{x}_t = \Phi_1 \boldsymbol{x}_{t-1} + \cdots + \Phi_p \boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t, \qquad \{\boldsymbol{\epsilon}_t\} \sim \text{iid}(\mathbf{0}, \Sigma_\epsilon)$$

- **stationary/stable/causal**
- G is Toeplitz matrix of first p process autocovariances
- total number of parameters/coefficients: $m := pd^2$

$$\boldsymbol{\psi} = \text{vec}([\Phi_1, \ldots, \Phi_p])$$

- **Index set** of all elements of $\boldsymbol{\psi}$:

$$I = \{1, \ldots, m\} = J \cup K$$

where

- $J =$ active elements $(\neq 0)$
- $K =$ non-active elements $(= 0)$

Both least-squares (LSEs) & max likelihood (MLEs) have same asymptotics:

$$\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} (\mathbf{0}, \Omega), \qquad \Omega = \Sigma_\epsilon \otimes G^{-1}$$

- $\hat{p}_i$: p-value for testing null that $\psi_i = 0$
- Ordered p-values:

$$\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(m)}$$

A1. The VAR is strictly stationary and stable, and the covariance matrix $\Sigma_\epsilon \otimes G^{-1}$ is well defined and non-singular.

A2. The iid series $\{\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{dt})'\}$ has a continuous distribution with a positive definite covariance Σ_ϵ, and satisfies $\mathbb{E}|\epsilon_{it}\epsilon_{jt}\epsilon_{kt}\epsilon_{lt}| < \infty$ for any $i, j, k, l \in \{1, \ldots, d\}$, and all t.

A3. The total number of coefficient parameters m is such that $m < n$ for all sufficiently large n, and the threshold activation parameter is of order: $h_n = O(\log n)$.

A4. The threshold satisfies $p_n \to 0$ and $\log(p_n) = o(n)$ as $n \to \infty$.

Declare as non-active coefficients for which p-value exceeds threshold:

$$\hat{p}_i > p_n$$

Theorem

Let $\psi_{J_0} \neq \mathbf{0}$ denote the vector of active parameters of the stationary and stable VAR model, and $\hat{\psi}_{\hat{J}}$ its TLSE estimator. Then, under conditions **A1–A4**, *we have as $n \to \infty$:*

- $P(\hat{J} = J_0) \to 1$
- $\sqrt{n}(\hat{\psi}_{\hat{J}} - \psi_{J_0}) \xrightarrow{d} N(\mathbf{0}, \Omega_{J_0})$

Declare as non-active coefficients for which p-value exceeds threshold:

$$\hat{p}_i > p_n$$

Consider following t-holds (satisfy A3 & A4):

- **TLSE1:** $p_n = \dfrac{1}{\sqrt{h_n}\log(h_n)}$
- **TLSE2:** $p_n = \dfrac{1}{\sqrt{h_n \log(h_n)}}$
- **TLSE3:** $p_n = \dfrac{h_n}{n\log(h_n)}$

Activation parameter choice:

$$h_n = m \log n$$

- Ordered p-values and corresponding concomitant VAR coefficients:

$$\{\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(m)}\} \qquad \Longleftrightarrow \qquad \{\psi_{(1)}, \ldots, \psi_{(m)}\}$$

- For $j = 1, \ldots, m$: model $\widehat{\mathcal{M}}_j$ contains coefficients

$$\{\psi_{(1)}, \ldots, \psi_{(j)}\}$$

- Nested sequence of models:

$$\widehat{\mathcal{M}}_1 \subset \cdots \subset \widehat{\mathcal{M}}_m$$

- **BLSE:** Model with **smallest BIC** in nested sequence

$$\widehat{\mathcal{M}}_*$$

BLSE Algorithm:

- delivers Model $\widehat{\mathcal{M}}_*$
- identifies the (active) index set $\widehat{J}_*$

Theorem

For the stationary and stable VAR(p) with Assumptions **A1**–**A4**:

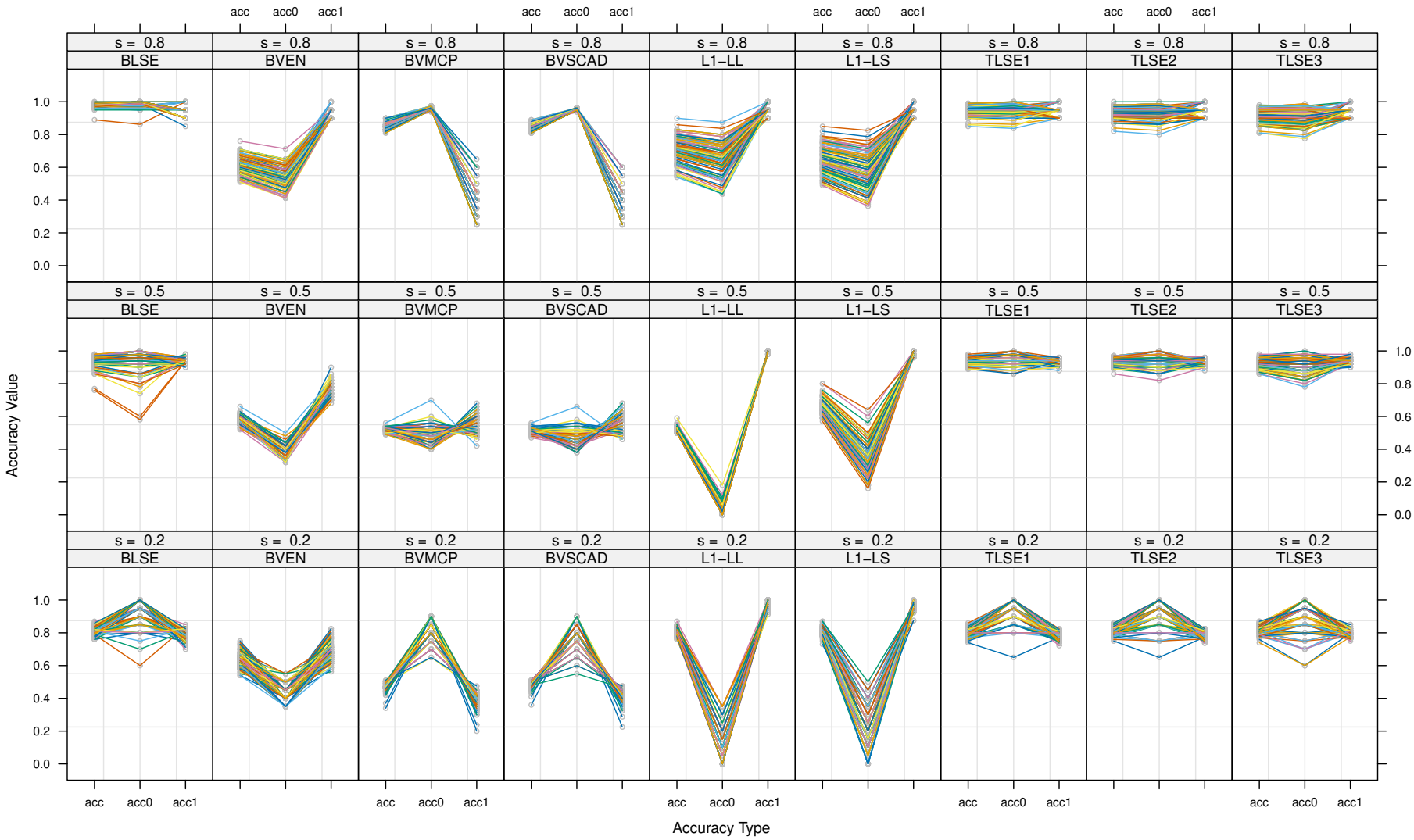
$$P(J_0 \subset \widehat{J}_*) \to 1, \qquad \text{as } n \to \infty$$

- **Models (2):**
  - **Medium VAR:** $d = 10$ dim VAR(1), $n = 200$.
  - **Large VAR:** $d = 30$ dim VAR(1), $n = 2000$.
- **Sparsity Values (3):** $s = \{0.2, 0.5, 0.8\}$.
- **Methods (9):**
  - **L1-LS:** OLS-based lasso (`sparsevar`)
  - **L1-LL:** MLE-based lasso (`sparsevar`)
  - **BVEN:** OLS-based elastic net (`BigVAR`)
  - **BVSCAD:** OLS-smoothly clipped absolute deviation (`BigVAR`)
  - **BVMCP:** OLS-minimax concave penalty (`BigVAR`)
  - **BLSE**
  - **TLSE1**
  - **TLSE2**
  - **TLSE3**

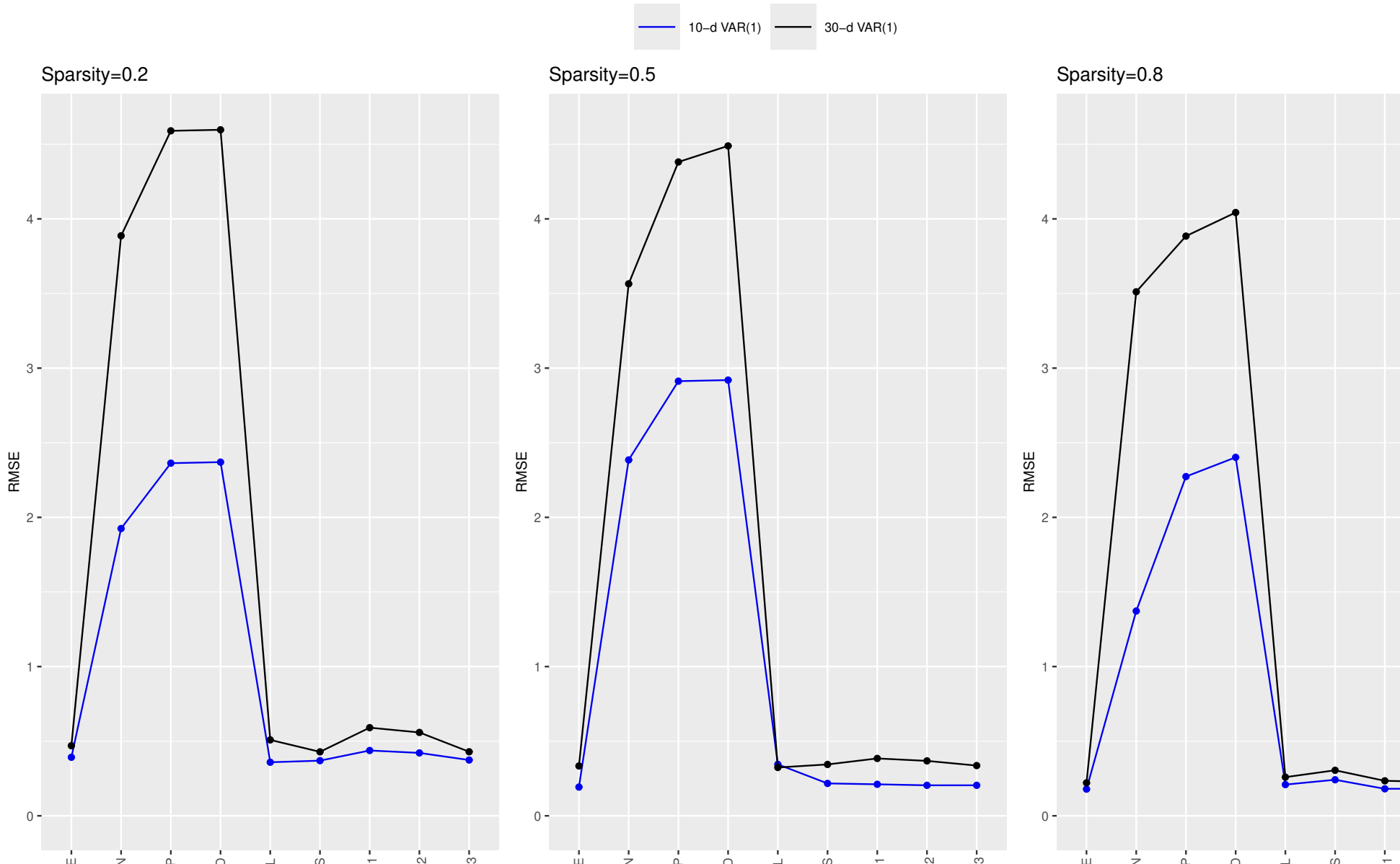
- **Sparsity pattern recovery.** Use usual metrics from binary classification:
  - acc_1: true positive rate (sensitivity)
  - acc_0: true negative rate (specificity)
  - acc: overall accuracy (proportion correctly classified)
- **Coefficient estimation.**
  - RMSE: only for the truly active coefficients
- **1-Step forecast error.**
  - RMSFE

(All based on 100 replications)

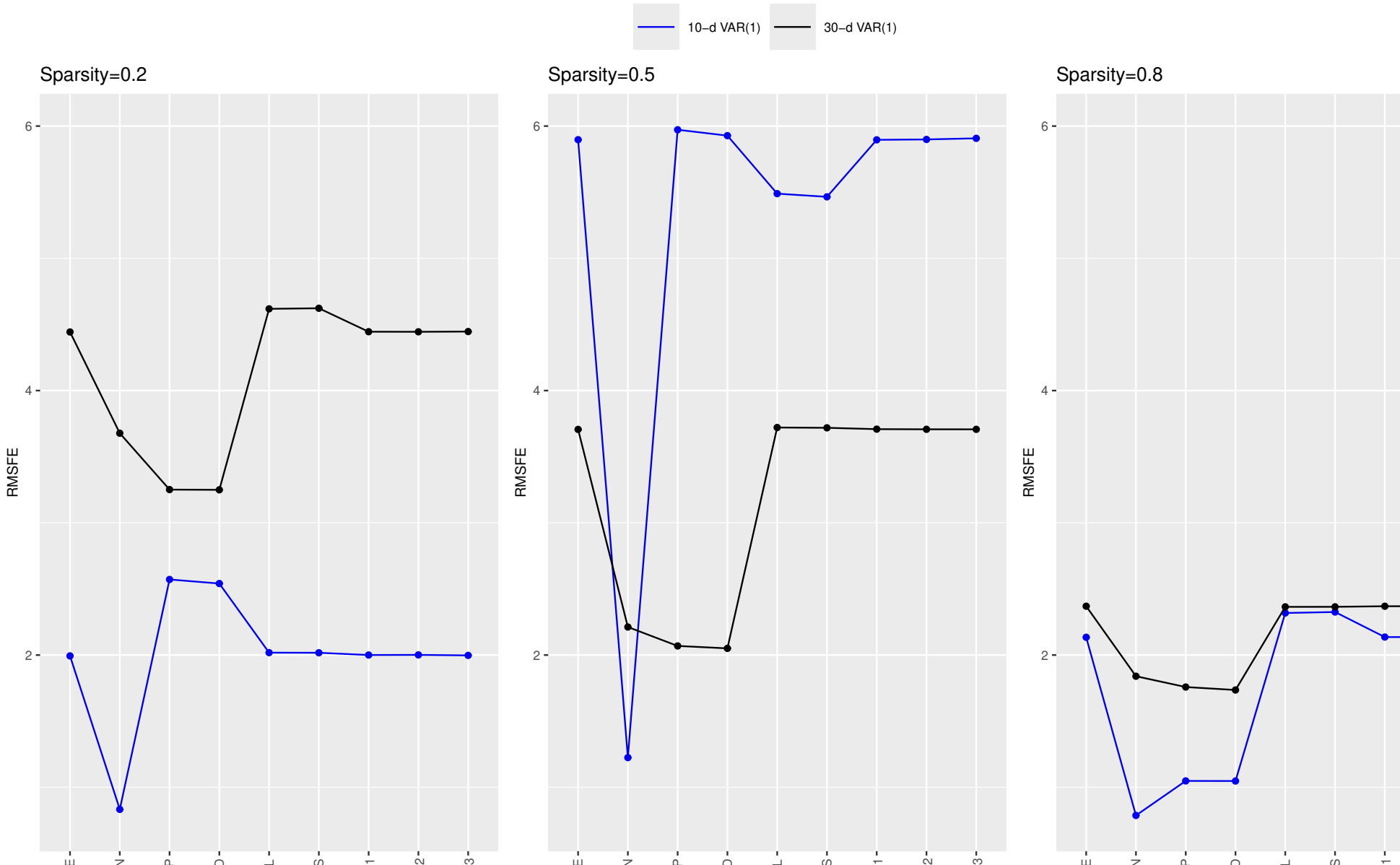
Simulations: Sparsity Pattern Recovery (Medium VAR)



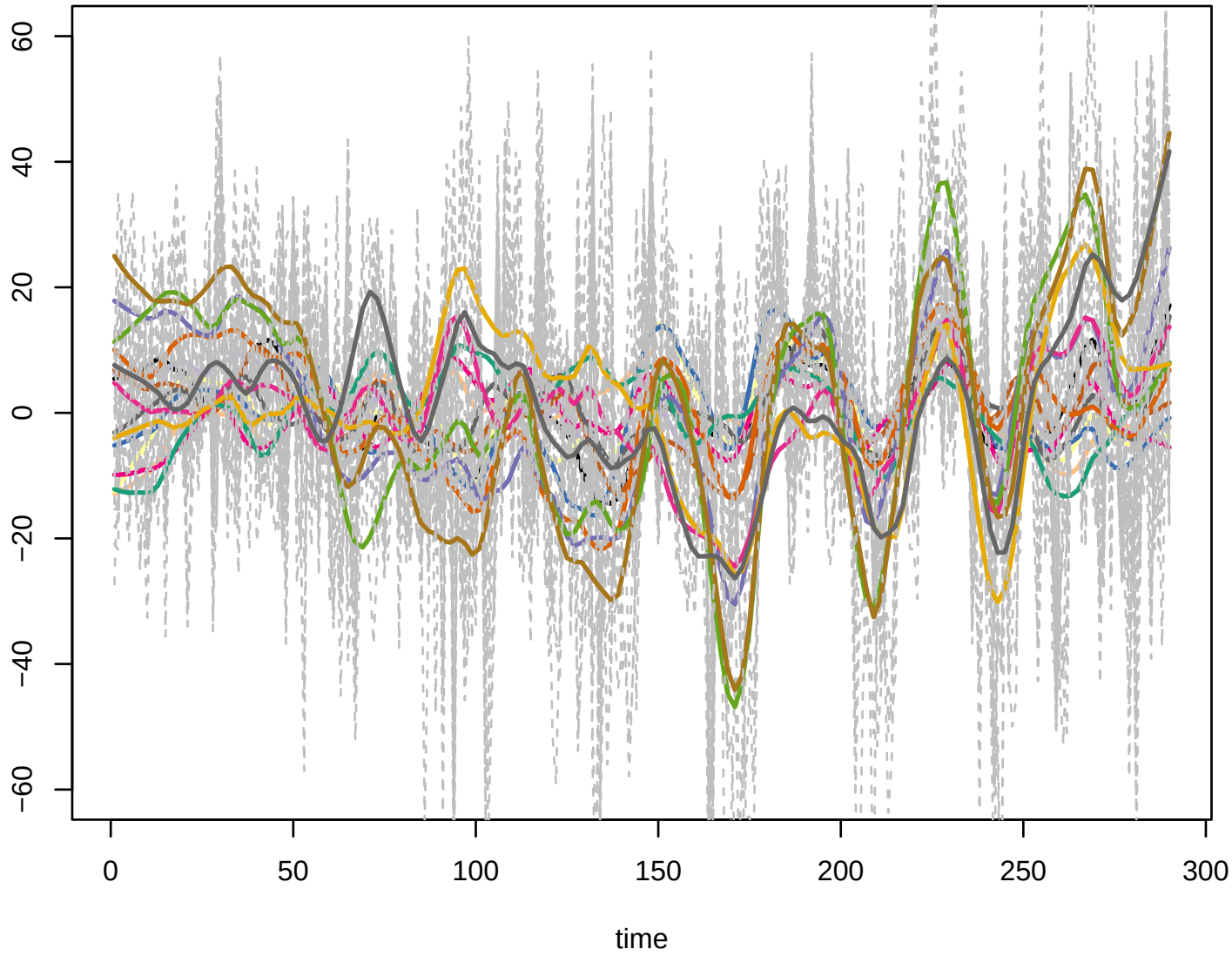
Simulations: Coefficient Estimation (RMSE)



Simulations: 1-Step Forecast Error (RMSFE)



Real Data: Brain fMRI Signals on $d = 16$ Adjacent Voxels

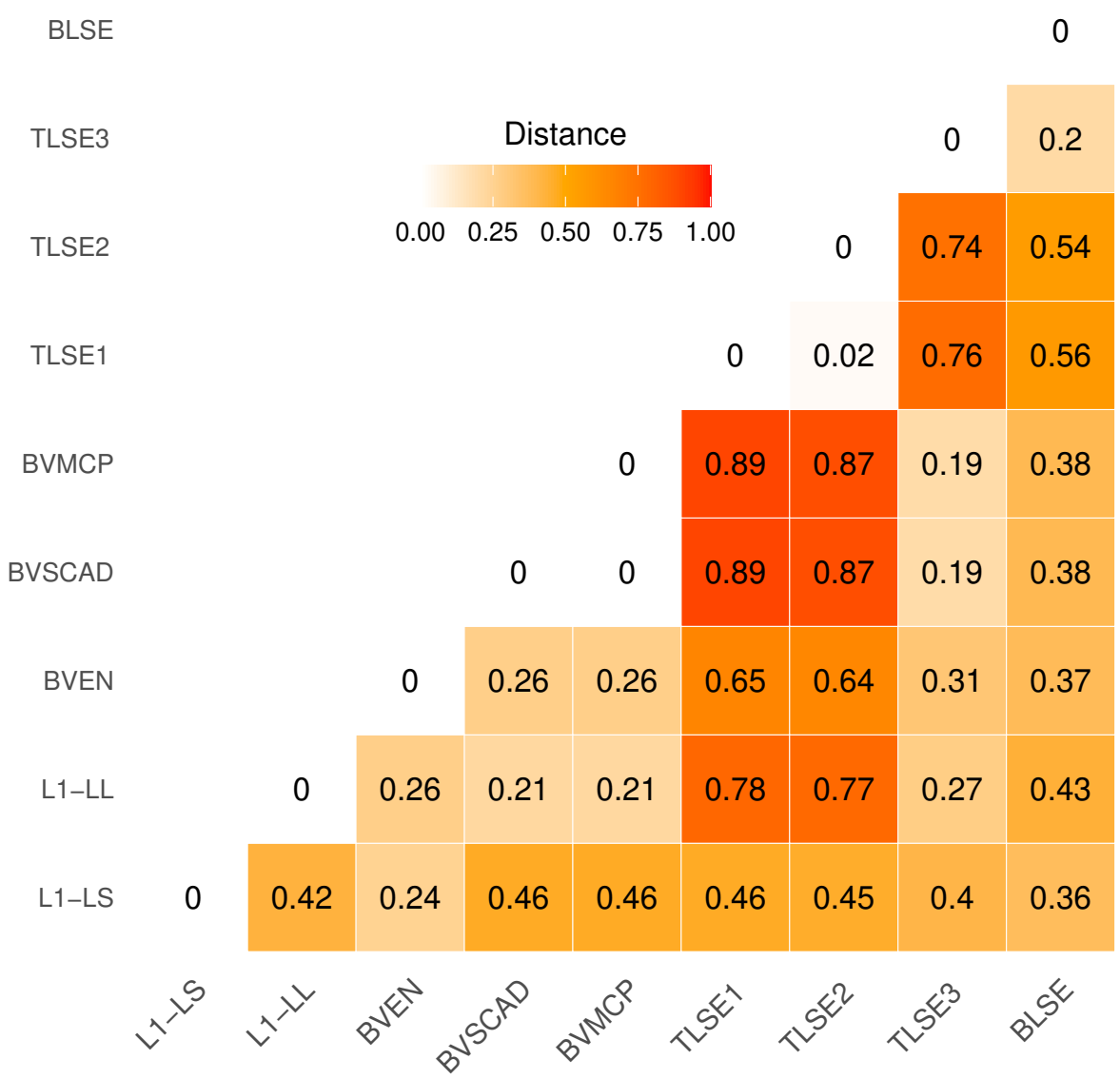


Brain fMRI Data: Sparsity Percentages by Method

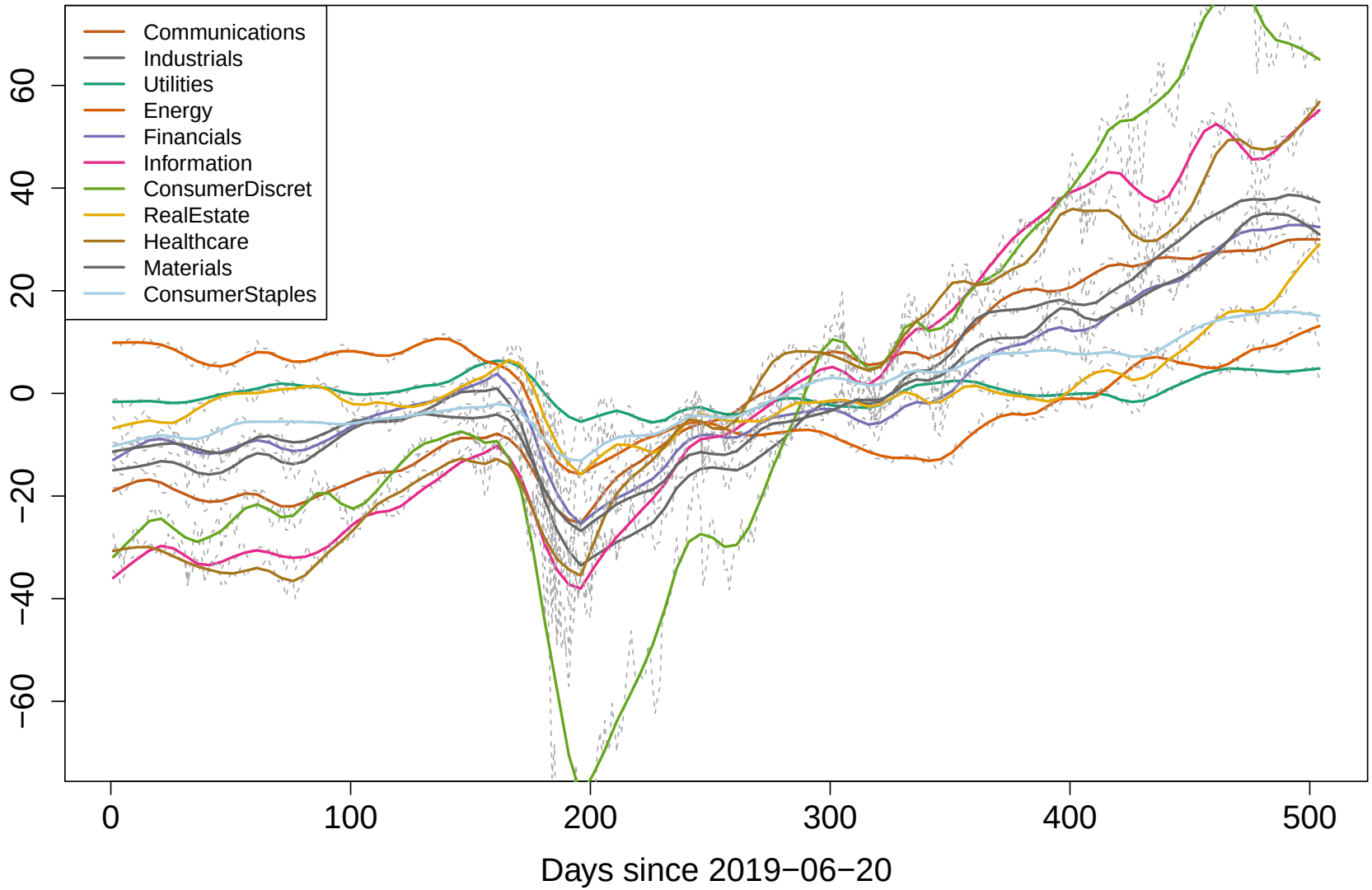
Table: Percentage of non-active coefficients (sparsity) for the Φ_1 matrix of VAR(1) models fitted by each method.

TLSE1	TLSE2	TLSE3	BLSE	BVEN	BVMCP	BVSCAD	L1-LL	L1-LS
90	86	54	36	27	7	7	16	45

Brain fMRI Data: Sparsity Pattern Distances (Scaled Manhattan Metric)



Real Data Illustration: S&P 500 Index Over $d = 11$ Economic Sectors

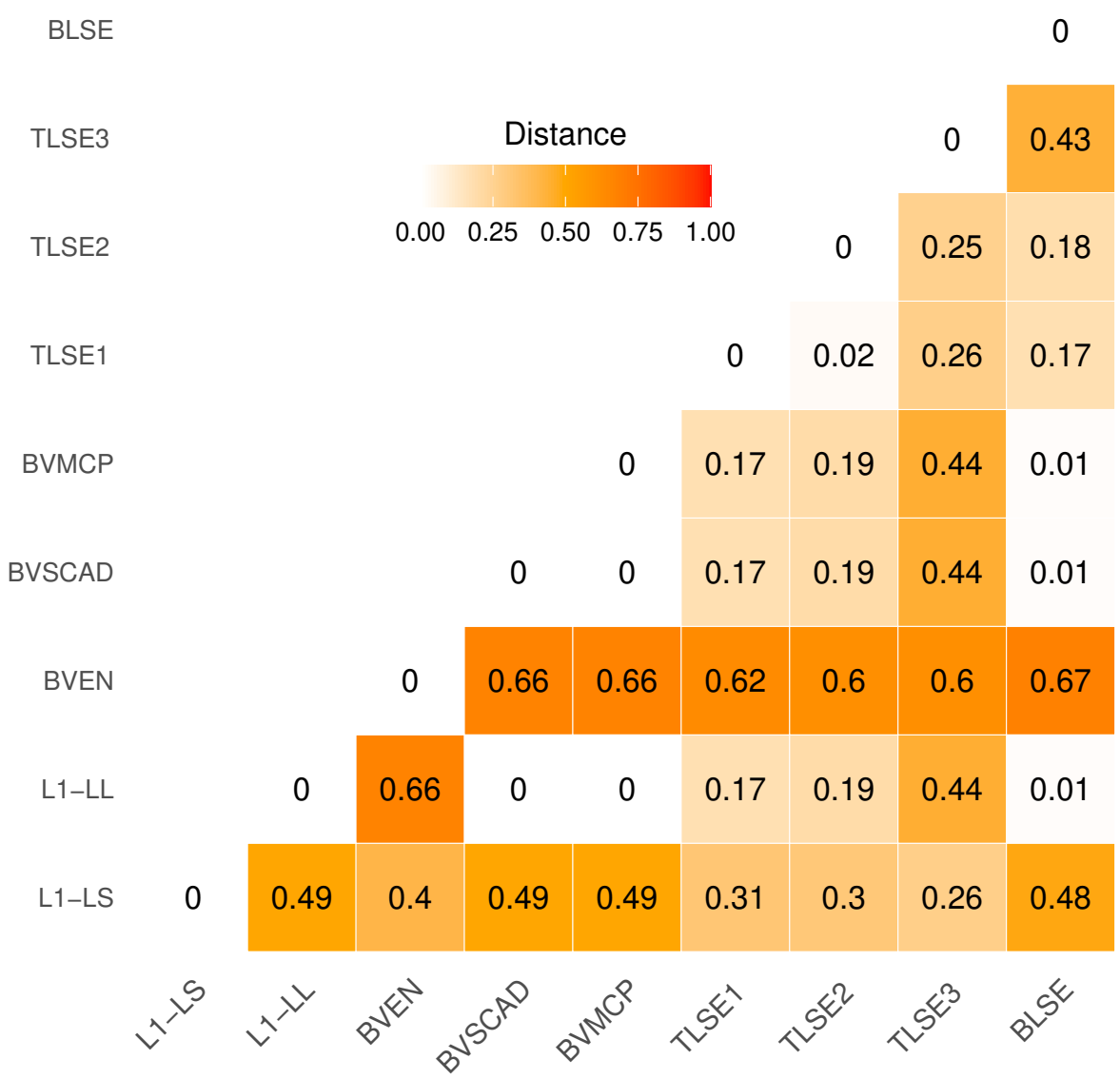


S&P 500 Data: Sparsity Percentages by Method

Table: Percentage of non-active coefficients (sparsity) for the Φ_1 matrix of VAR(1) models fitted by each method.

TLSE1	TLSE2	TLSE3	BLSE	BVEN	BVMCP	BVSCAD	L1-LL	L1-LS
77	74	69	99	34	100	100	100	51

S&P 500 Data: Sparsity Pattern Distances (Scaled Manhattan Metric)



Granger (1969):

Y Granger-Causes X if prediction of X improves when one uses past values of Y, given that all other relevant information, Z, is taken into account.

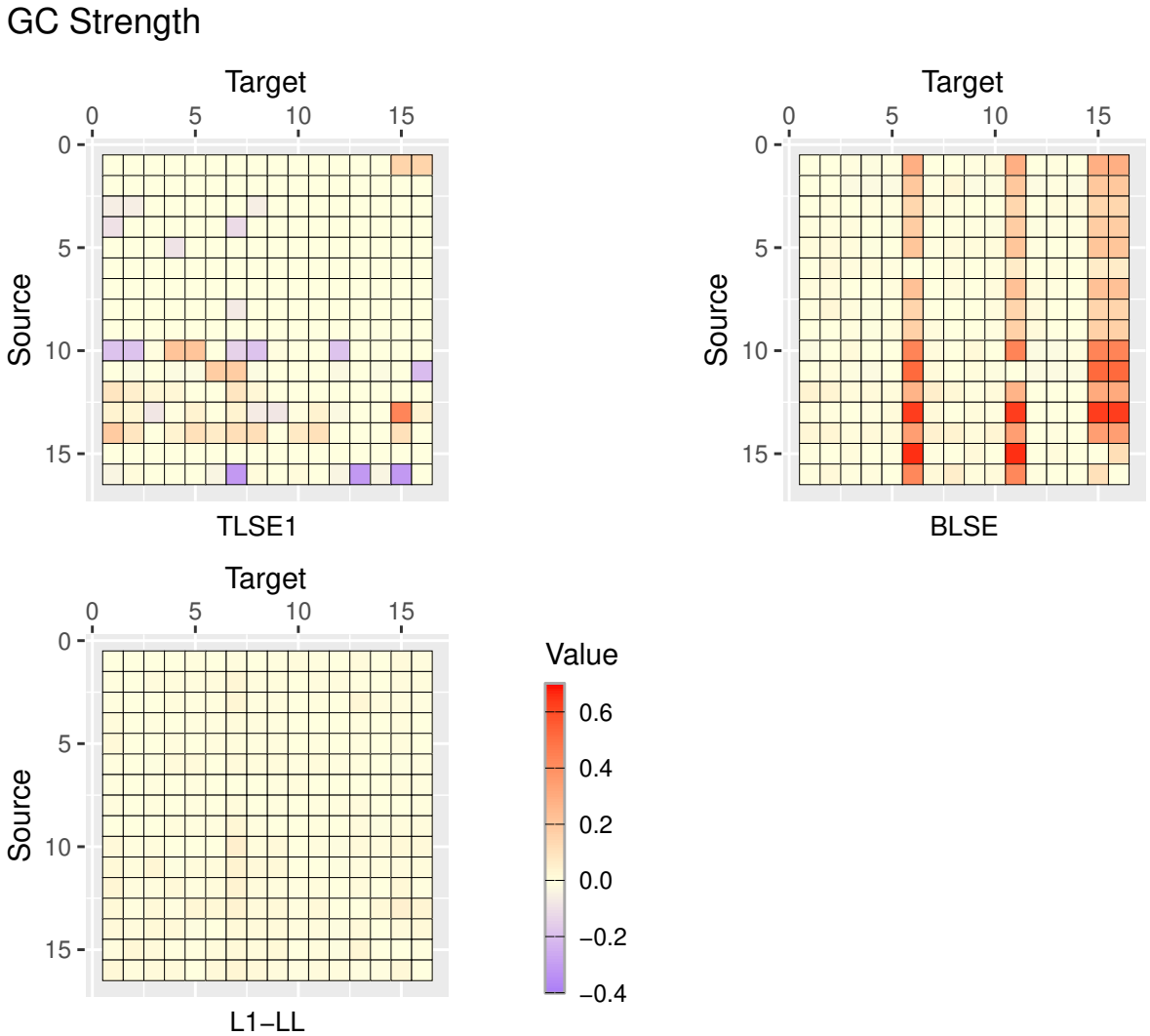
For VAR Model (Geweke, 1984):

- $\Sigma_{XY|Z}$: prediction error covariance for $X|Z$ when Y is included
- $\Sigma_{XX|Z}$: prediction error covariance for $X|Z$ when Y is excluded

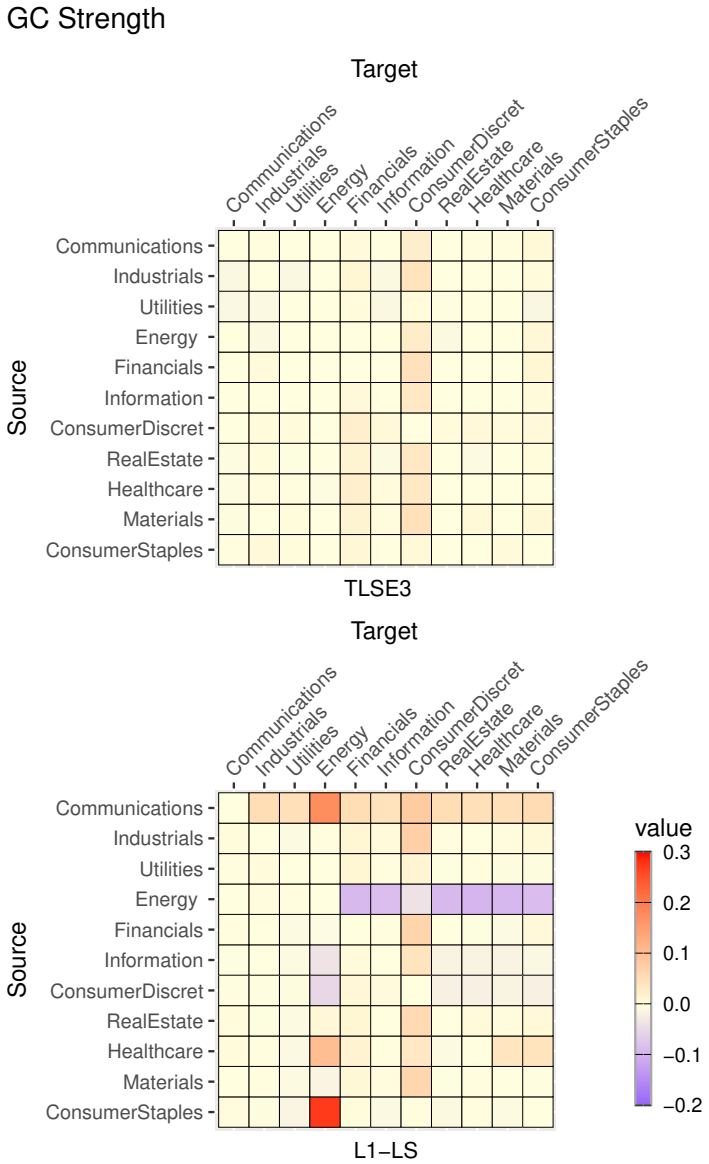
$$GC(Y \to X|Z) = \log\left(\frac{|\Sigma_{XX|Z}|}{|\Sigma_{XY|Z}|}\right).$$

- $GC(Y \to X|Z) > 0$: suggests Y is G-Causal for X
- $GC(Y \to X|Z) \leq 0$: suggests Y is not G-Causal for X

Brain fMRI Data: GC-Flow



S&P 500 Data: GC-Flow



References

- Barnett & Seth (2014), "The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference", *Journal of Neuroscience Methods*, 223, 50–68.
- Basu, Li, & Michailidis (2019). "Low rank and structured modeling of high-dimensional vector autoregressions", *IEEE Transactions on Signal Processing*, 67, 1207—1222.
- Davis, Zang & Zheng (2016). "Sparse vector autoregressive modeling", *Journal of Computational and Graphical Statistics*, 25, 1077–1096.
- Geweke (1984). "Measures of conditional linear dependence and feedback between time series", *Journal of the American Statistical Association*, 79, 907–915.
- Kastner & Huber (2020). "Sparse bayesian vector autoregressions in huge dimensions", *Journal of Forecasting*, 39, 1142-–1165.
- Nicholson, Matteson, & Bien (2017). "VARX-L: Structured regularization for large vector autoregressions with exogenous variables", *International Journal of Forecasting*, 33, 627—651.

Thank You!