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

Mihai Giurcanu[†] & Dananjani Liyanage[*] & Alex Trindade[**]

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

(∗) *Dept. of Mathematics & Statistics, University of Minnesota–Duluth*

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

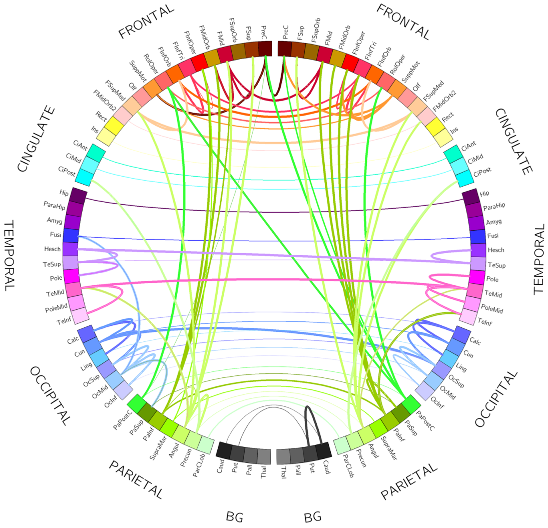
July 2026

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

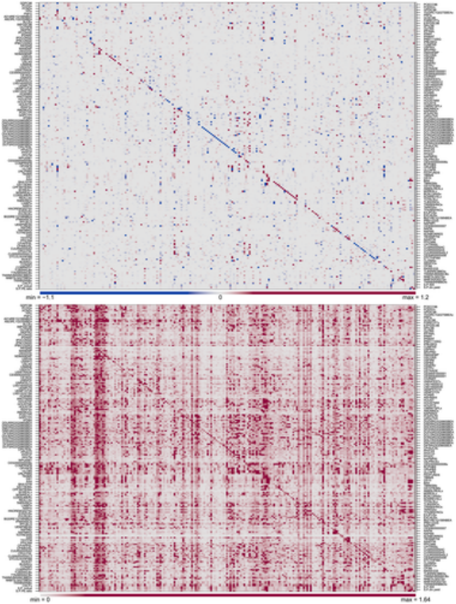
$$\text{VAR dimension} = d, \qquad \text{VAR order} = p$$

- **Total Dim:** $> pd^2$, inherently **sparse**...
- **Existing Methods:** shrinkage-based LASSO and variants
- **R Packages:** `bigtime` (formerly `sparsevar` & `BigVAR`)
- **New Methods:** p-value thresholding; BIC-based backward 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



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

- **stationary/stable/causal**
- Characteristic matrix: $\Phi(z) = I - \sum_{k=1}^{p} \Phi_k z^k$
- Total number of coefficients: vector of dim $m \equiv pd^2$

$$\boldsymbol{\theta} = \text{vec}\{\text{vec}(\Phi_1^T), \ldots, \text{vec}(\Phi_p^T)\}$$

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

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

where

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

- Write model in regression form for least-squares estimation (LSE):

$$\boldsymbol{X}_t = W_t\boldsymbol{\theta} + \boldsymbol{\epsilon}_t, \quad t = p+1, \ldots, n \qquad \Longrightarrow \qquad X = W\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

- LSE asymptotics function of:

$$\Omega = \mathbb{E}(W_t^T W_t) \qquad \Psi = \mathbb{V}(W_t^T \boldsymbol{\epsilon}_t) \qquad \Pi = \mathbb{E}(W_t^T W_t)^2$$

$$V_t = \text{vec}(W_t^T W_t), \quad \Gamma_V(h) = \mathbb{V}(V_{t+h} V_t^T), \quad \Gamma_V = \sum_{h=-\infty}^{\infty} \Gamma_V(h)$$

$$U_t = \text{vec}(W_t^T \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T W_t), \quad \Gamma_U(h) = \mathbb{V}(U_{t+h} U_t^T), \quad \Gamma_U = \sum_{h=-\infty}^{\infty} \Gamma_U(h)$$

Define regularity conditions:

A1. Stable VAR, $\epsilon_t \sim$ iid, Σ_ϵ non-singular, $m = o(n)$.

A2. Conditions on growth order of min and max eigenvalues of Ω, Ψ, Π, Γ_V, Γ_U, $\mathbb{V}(\epsilon_t \epsilon^T)$.

A3. $\sum_{h=-\infty}^{\infty} \|\Gamma_V(h)\| < \infty$ and $\sum_{h=-\infty}^{\infty} \|\Gamma_U(h)\| < \infty$.

Theorem (Asymptotics: all coefficients active)

Under conditions **A1–A3***:*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Omega^{-1}\Psi\Omega^{-1}\right)$$

Note: both $m \to \infty$ and $n \to \infty$!

$(J, m_J, \boldsymbol{\theta}_J)$: oracle index set, cardinality, vector of actives

- $\widehat{p}_i$: p-value for testing $H_0 : \theta_i = 0$
- Ordered p-values:

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

TLSE: use thresholding to select active coefficients

$$\widehat{p}_i < p_n \qquad \Longrightarrow \quad (\widehat{J},\ \widehat{\boldsymbol{\theta}}_{\widehat{J}})$$

Theorem (TLSE Asymptotics)

Under conditions **A1–A3**, *if* $p_n = o(1/m_K)$ *with* $\log(p_n) = o(n)$, *we have:*

- $P(\widehat{J} = J) \to 1$ **(oracle property)**
- $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\widehat{J}} - \boldsymbol{\theta}_J) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_J^{-1}\Psi_J\Omega_J^{-1})$

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

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

Consider family of t-holds (satisfies conditions of Theom):

$$p_n = \frac{1}{n^\rho \log(\log n)}, \qquad 0 \le \rho \le 1$$

Special (extreme) cases:

- **TLSE0:** Set $\rho = 0 \implies p_n = \dfrac{1}{\log(\log n)}$
- **TLSE1:** Set $\rho = 1 \implies p_n = \dfrac{1}{n \log(\log n)}$

- Ordered p-values and corresponding concomitant VAR coefficients:

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

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

$$\{\widehat{\theta}_{(1)}, \ldots, \widehat{\theta}_{(j)}\}$$

- Nested sequence of models:

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

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

$$GIC(J, b) = RSS(\widehat{\boldsymbol{\theta}}_J) + b\ m_J$$

- **Output:** Optimal model & active index set

$$\widehat{\mathcal{M}}_* \qquad \Longrightarrow \qquad \widehat{J}_*$$

BLSE Algorithm: delivers $(\widehat{\mathcal{M}}_*, \widehat{J}_*)$

Theorem

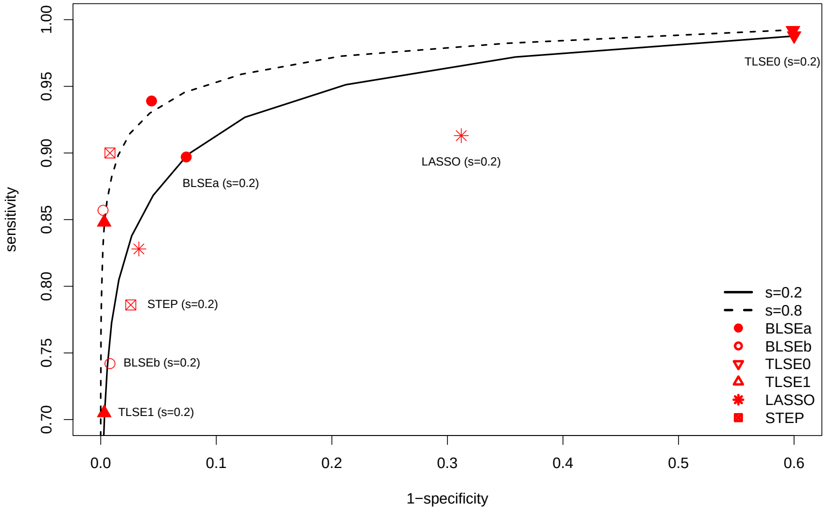
Under conditions **A1–A3**, $m_J = o(n^{1/6})$, *and* $b = o(n^{5/6}/m_J)$:

- $P(J \subseteq \widehat{J}_*) \to 1$ **(sure screening property)**
- $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\widehat{J}_*} - \boldsymbol{\theta}_J) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Omega_J^{-1}\Psi_J\Omega_J^{-1})$

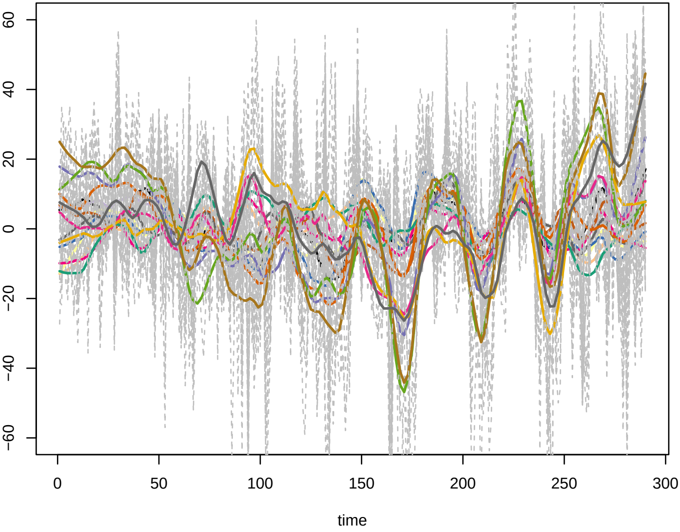
Tighter conditions also ensure $P(\widehat{J}_* = J) \to 1$ **(oracle property)**

- **BLSEa:** set $b = 2$ as in AIC $\Rightarrow$ sure screening property
- **BLSEb:** set $b = \log n$ as in BIC $\Rightarrow$ oracle property

Simulations: TLSE ROC Curve $0 \leq \rho \leq 1$ (10-dim sparse VAR(1), $n = 200$)



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

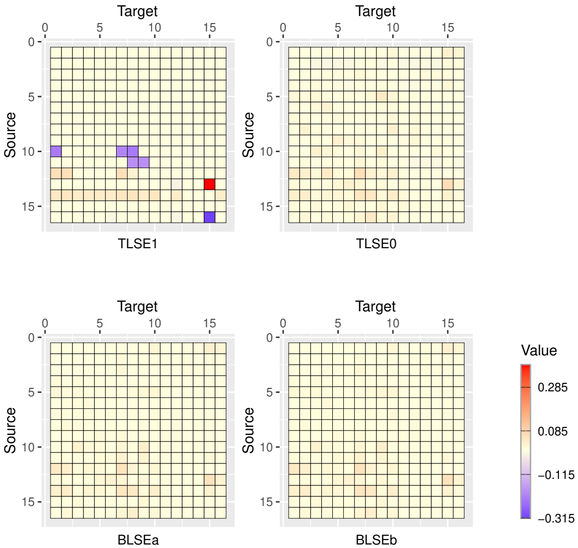


- **Granger (1969):** X "causes" Y if prediction of Y "improves" when past values of X are used, given all other relevant information (Z)
- **Geweke (1984):** Improvement measured by reduction in error variance of Y when X is included ($\Sigma_{XY|Z}$), relative to reduced model excluding X ($\Sigma_{YY|Z}$)

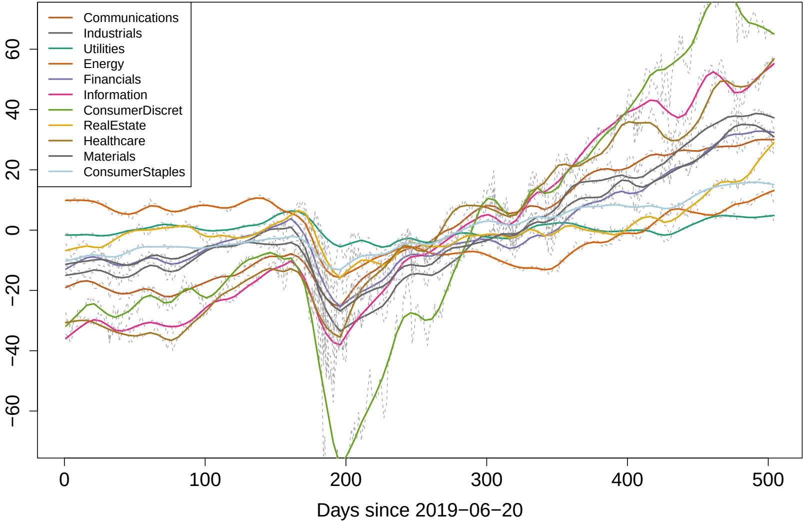
$$GC(X \to Y|Z) = \log\left(\frac{|\Sigma_{YY|Z}|}{|\Sigma_{XY|Z}|}\right)$$

- **For a VAR (Barnett & Seth, 2014):** measure GC strength of component i on j as

$$\nu_{i,j} = \mathrm{GC}(X_{t,i} \to X_{t,j} \mid X_{t,-\{i,j\}})$$


Target
0 5 10 15
Source
0 5 10 15
TLSE1
Target
0 5 10 15
Source
0 5 10 15
TLSE0
Target
0 5 10 15
Source
0 5 10 15
BLSEa
Target
0 5 10 15
Source
0 5 10 15
BLSEb
Value
0.285
0.085
−0.115
−0.315

Real Data: S&P 500 Index on $d = 11$ Economic Sectors (averages)



- **Persistent problem in time series:** combine both lead/lag dependence (directed) with instantaneous dependence (undirected) into one overall measure of causality
- **For VAR:** coefficients Φ_k account for long-run dynamic relationship; contemporaneous dependence is contained in Σ_ϵ
- With $\Phi(z)$ the characteristic matrix, define long-run concentration matrix

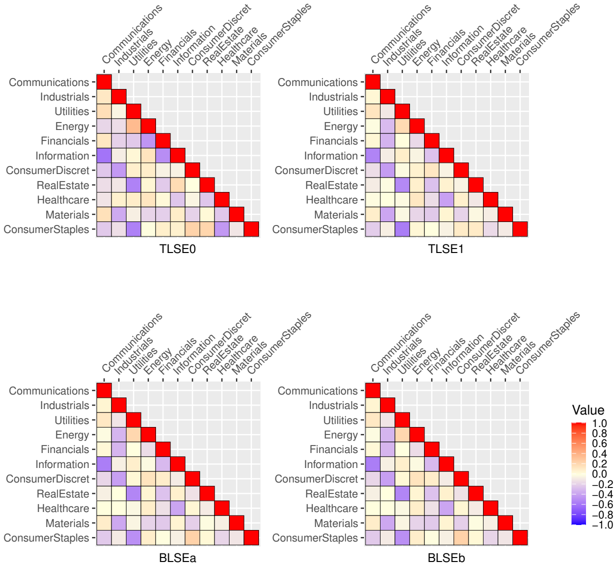
$$K = \Phi(1)^T \Sigma_\epsilon^{-1} \Phi(1) \equiv (k_{i,j})$$

- **Barigozzi & Brownlees (2019):** long-run PSC coefficient between $X_{t,i}$ and $X_{t,j}$

$$\rho_{i,j} = \frac{k_{i,j}}{\sqrt{k_{i,i} k_{j,j}}}$$

(Equal to the zero-frequency **partial spectral coherence**...)

S&P 500: PSC Coefficient



Acknowledgements

- Dr Travis Thompson (fMRI data); Andras Horvath (S&P 500 data)

References

- Barigozzi & Brownlees (2019), "Nets: Network estimation for time series", *Journal of Applied Econometrics*, 34(3):347–364.
- 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.
- Duggento et al (2018), "Multivariate granger causality unveils directed parietal to prefrontal cortex connectivity during task-free mri", *Scientific Reports*, 8(1):5571.
- 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.
- Giurcanu (2017), "Oracle M-estimation for time series models", *Journal of Time Series Analysis*, 38(3):479–504.

Thank You!