Imperial College London

UNIVERSITY OF OXFORD

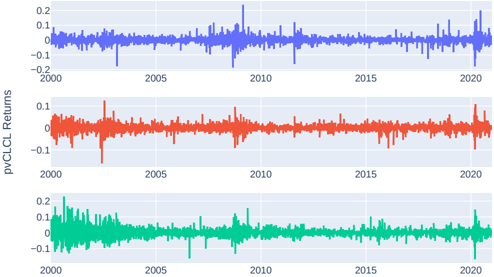
15 October 2024

NIRVAR: Network Informed Restricted Vector Autoregression

Brendan Martin, Francesco Sanna Passino, Mihai Cucuringu, Alessandra Luati

Setting

Data: Panel of high dimensional time series, $\{(X_{1,t}, ..., X_{N,t})'\}_{t \in \mathbb{Z}}$.



Goal:

- Find a coarse-grained description of the system by grouping the panel components.
- Utilise this coarse-grained information in a parametric time series modelling context to aid with estimation and prediction.

Motivation

Vector Autoregression (VAR) is a widely used model for panels of multivariate time series, $\{(X_{1,t}, ..., X_{N,t})'\}_{t\in\mathbb{Z}}$; e.g. environmental science, econometrics, neuroscience.

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Sigma),$$

The task: We want to estimate the VAR parameters, Φ, in order to understand the co-movement between panel components and for prediction tasks.

Challenge: Φ has N^2 parameters. For large panels, it is often the case that $N^2 > T$, where T is the number of observations.

Solutions: Utilise techniques from high dimensional statistics: dimensionality reduction, factor models, sparse regression via penalisation, network based approaches.

Related Literature

- **Factor Models:** The large panel of time series are modelled as stemming from a relatively small number of common latent factors [Stock and Watson, 2002].

- **Factors + Sparse Regression:** Fan, Masini, and Medeiros [2023] combine the dimensionality reduction of factor modelling with the parsimony of sparse linear regression and give a novel test for covariance structure. Their proposed model is called the Factor Augmented Regression Model (**FARM**).

- **Network VAR:** Knight, Leeming, Nason, and Nunes [2020] introduce **GNAR** which, given an observed network, fits a flexible network autoregressive model. Barigozzi, Cho, and Owens [2023] propose an ℓ_1-regularised Yule-Walker method for estimating a factor adjusted, idiosyncratic VAR model (**FNETS**).

- **Community Detection:** Guðmundsson and Brownlees [2021] use estimated VAR coefficients to embed and cluster the panel components.

NIRVAR Model

NIRVAR Model

$$
\begin{aligned}
\boldsymbol{X}_t &= A \odot \tilde{\Phi} \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Sigma), \\
A &\sim \text{SBM}\left(B, \pi\right),
\end{aligned}
$$

where $\odot$ is the Hadamard (entry-wise) product and $\tilde{\Phi}$ is an $N \times N$ matrix of fixed weights.

Random Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the edge set, $\mathcal{E}$, are random variables defined over some probability space.

Stochastic Block Model (SBM): a random graph in which each vertex belongs to one of K communities, called blocks.

Subset VAR: The NIRVAR model is a **random coefficient** subset VAR model in which the zero-restrictions are determined by a SBM.

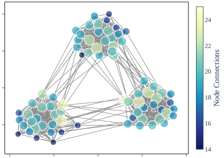
Stochastic Block Model

A Stochastic Block Model (SBM) is a random graph in which each vertex belongs to one of K communities, called blocks. Vertices are assigned to blocks via

$$z : \{1, ..., N\} \to \{1, ..., K\}$$

The probability of an edge forming between i and j is $B_{z_i z_j}$, where B is the $K \times K$ block probability matrix.

$$B = \begin{pmatrix} 0.75 & 0.03 & 0.03 \\ 0.03 & 0.75 & 0.03 \\ 0.03 & 0.03 & 0.75 \end{pmatrix}$$



We write $A \sim \text{SBM}(B, \pi)$, where $\pi = (\pi_1, \dots, \pi_K)$ represent the prior probabilities of each node to belong to the k-th community, with $\pi_k \geq 0$ for all $k \in [K]$ and $\sum_{k=1}^{K} \pi_k = 1$.

Stochastic Block Model: Fixed Block Memberships

When the block memberships are taken to be nonrandom, the matrix of edge probabilities, $P \in [0,1]^{N \times N}$, can be written as

$$P = ZBZ',$$

where $Z \in \{0,1\}^{N \times K}$ with $Z_{ij} = 1$ if and only if $z(i) = j$.

Then $A \sim \text{Bernoulli}(P)$, so $\mathbb{E}(A) = P$.

A is a "noisy version" of P.

NOTE: We assume B is full rank: $\text{rank}(B) = K$.

Latent Position Random Graphs

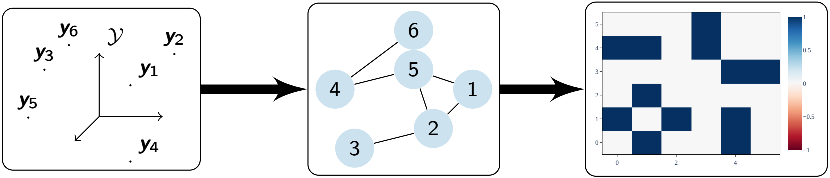
A SBM is an example of a **random dot product graph** which itself is a **latent position random graph**.

Latent Position Random Graph: Each vertex in $\mathcal{V}$ has an associated latent position, $\boldsymbol{y}_i \in \mathcal{Y} \subset \mathbb{R}^d, i \in \mathcal{V}$, where $\mathcal{Y}$ is some latent space. The probability, p_{ij}, of an edge forming between two vertices, i and j, is independent of all other edges and is given by $p_{ij} = \kappa(\boldsymbol{y}_i, \boldsymbol{y}_j)$, where $\kappa : \mathcal{Y} \times \mathcal{Y} \to [0, 1]$ is a so-called kernel function.

Key point: $d < N \implies$ dimensionality reduction!

If κ is the dot product, then the latent position random graph is called a **random dot product graph**.

Latent Position Random Graphs



- The latent positions, $\mathbf{y}_i$ and $\mathbf{y}_j$, determine the probability of an edge forming between vertex i and j.
- The realised edges can be represented by an adjacency matrix, A.

Representing a SBM as a random dot product graph

SBMs can be represented as a random dot product graph with every vertex in the same block having the same latent position:

$$z(i) = z(j) \iff \boldsymbol{y}_i = \boldsymbol{y}_j,$$

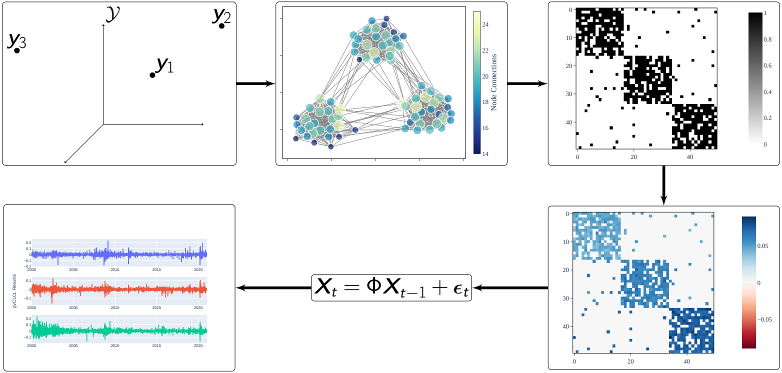
where, again, $z : \{1, ..., N\} \to \{1, ..., K\}$ is the block assignment function.

Block matrix: $B_{z_i z_j} = \boldsymbol{y}_i' \boldsymbol{y}_j$

Adjacency matrix: $A_{ij} \sim \text{Bernoulli}(\boldsymbol{y}_i' \boldsymbol{y}_j)$

Conclusion: A SBM can be written as a random dot product graph having K **distinct** d-dimensional latent positions.

Pictorial Representation of NIRVAR Model

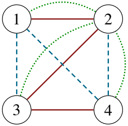


Multiplex Networks

NIRVAR can model data that is represented by a multiplex network, corresponding to a network containing **multiple types of edges**, expressed via a graph G containing multiple layers of connectivity, one layer for each type of edge.

$$\boldsymbol{X}_t^{(q)} = \sum_{r=1}^{Q} (A_q^{(r)} \odot \tilde{\Phi}_q^{(r)}) \boldsymbol{X}_{t-1}^{(r)} + \boldsymbol{\epsilon}_t^{(q)}, \qquad \boldsymbol{\epsilon}_t^{(q)} \sim \mathcal{N}(0, \Sigma)$$

$$A_{ij}^{(r)} \sim \text{Bernoulli}\left(\boldsymbol{y}_i^{(r)\prime} \boldsymbol{y}_j^{(r)} \right)$$



Layer 1

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Layer 2

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Layer 3

$$\mathbf{A}_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Example: Each stock in the S&P500 has multiple attributes or **features** such as open-to-close returns and previous-close-to-close returns.

Estimation

Two Step Approach

Recovering the edge set, $\mathcal{E}$, is challenging. We aim instead to recover the community, z_i, corresponding to panel component i.

The two step estimation approach is

1. Determine z_i by finding an embedding, $\hat{\boldsymbol{y}}_i$, for each panel component, i. Define the binary matrix, $\hat{A}_{ij}$, as

$$\hat{A}_{ij} = \begin{cases} 1 & \text{if} \quad \hat{\boldsymbol{y}}_i, \hat{\boldsymbol{y}}_j \quad \text{are in the } \textbf{same GMM cluster} \\ 0 & \text{otherwise.} \end{cases}$$

2. Set $\hat{\Phi}_{ij} = 0$ if $\hat{A}_{ij} = 0$ and estimate the remaining unrestricted parameters via ordinary least squares (**OLS**).

How to recover the latent communities?

- Compute the **sample covariance matrix**, S_T, between panel components.

- **Estimate the dimension** of the latent space by counting the number, $\hat{d} \in \mathbb{N}$, of eigenvalues of S_T that are greater than the Marčencko-Pastur distribution cutoff [Marčenko and Pastur, 1967].

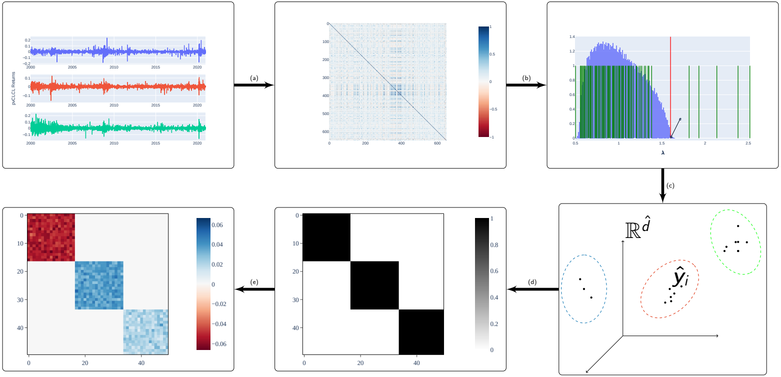
- Let $\Lambda, U \in \mathbb{R}^{N \times \hat{d}}$ be the matrices containing the $\hat{d}$ largest eigenvalues and corresponding eigenvectors of S_T. Compute the **embedding**,

$$\hat{Y} = U\Lambda^{1/2} \in \mathbb{R}^{N \times \hat{d}}.$$

- **Cluster** each $\hat{\boldsymbol{y}}_i \in \mathbb{R}^{\hat{d}}$ into $\hat{K} = \hat{d}$ groups using a Gaussian Mixture Model.

NOTE: Assuming Y has full rank, $\text{rank}(\mathbb{E}(A)) = \text{rank}(YY') = \text{rank}(Y) = d$. For a SBM, Y has K distinct rows, thus $\text{rank}(Y) = K$. Therefore, we set $K = d$.

NIRVAR Estimation in Pictures



Recap: NIRVAR Model and Estimator

NIRVAR model in vectorised form:

Let $X = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)$, $Z = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{T-1})$ and $U = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)$. Then

$$\boldsymbol{\psi} = \left(Z' \otimes I_N\right) \boldsymbol{\beta} + \mathrm{u},$$

where $\boldsymbol{\psi} := \mathrm{vec}(X)$, $\boldsymbol{\beta} := \mathrm{vec}(\Phi)$, and $\mathrm{u} := \mathrm{vec}(U)$.

The constraints on $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta} = R(A)\boldsymbol{\gamma}(A)$ where $\boldsymbol{\gamma}(A)$ is an M-dimensional vector containing the non-zero elements of $\boldsymbol{\beta}$. So $M = |A|_0$.

NIRVAR least squares estimator:

$$\hat{\gamma}(\hat{A}) = \{R(\hat{A})'(ZZ' \otimes \Sigma^{-1})R(\hat{A})\}^{-1} R(\hat{A})'(Z \otimes \Sigma^{-1})\,\boldsymbol{\psi}.$$

$$\hat{\boldsymbol{\beta}}(\hat{A}) = R(\hat{A})\hat{\gamma}(\hat{A}).$$

Modifying the Estimator

The estimation procedure follows the following workflow: dimension reduction via principal component analysis, embedding, clustering, and graph construction.

Each step of the workflow can be modified. For example

- Embed the **precision matrix** instead of the covariance matrix.
- Choose the embedding dimension using a scree plot.
- Construct the graph directly as $\hat{Y}\hat{Y}'$.

Latent Position Recovery

Theoretical justification for the spectral embedding estimator?

Currently, we **cannot** say whether $\hat{\boldsymbol{y}}_i$ is a consistent estimator of $\boldsymbol{y}_i$.

Ideally, we would like to prove that $\hat{\boldsymbol{y}}_i$ consistent and asymptotically normal with respect its population counterpart, $\boldsymbol{y}_i$.

In the case of **symmetric** Φ, we can, however, show that Γ and Φ share the same eigenbasis:

Proposition

Let $\boldsymbol{X}_t \sim \mathrm{NIRVAR}(\Phi)$ where Φ is assumed to be symmetric. Consider the eigendecomposition $\Phi = U_\Phi \Lambda_\Phi U_\Phi' + U_{\Phi,\perp}\Lambda_{\Phi,\perp}U_{\Phi,\perp}'$, where $U_\Phi \in \mathbb{O}(N \times d)$ and Λ_Φ is a $d \times d$ diagonal matrix comprising the d largest eigenvalues in absolute value of Φ. Then the rank d truncated eigendecomposition of the covariance matrix $\Gamma = \mathbb{E}(\boldsymbol{X}_t \boldsymbol{X}_t')$ is $\Gamma = U_\Phi \Lambda_\Gamma U_\Phi'$ in which Λ_Γ is a $d \times d$ diagonal matrix with diagonal elements $(\lambda_\Gamma)_i = 1/\{1 - (\lambda_\Phi)_i^2\}$ where $(\lambda_\Phi)_i$ is the corresponding diagonal entry of Λ_Φ.

Sketch Proof

Proof.

$$\Gamma - \Phi\Gamma\Phi' = \Sigma.$$

This is an example of a Lyapunov matrix equation and its formal solution is given by

$$\Gamma = \sum_{k=0}^{\infty} (\Phi)^k \, \Sigma \left(\Phi'\right)^k ,$$

which converges when $\rho(\Phi) < 1$. With $\Sigma = \sigma^2 I_N$ we have

$$\begin{aligned}
\Gamma &= \sigma^2 \sum_{k=0}^{\infty} (\Phi)^k \left(\Phi'\right)^k \\
&= \sigma^2 \sum_{k=0}^{\infty} (U_\Phi \Lambda_\Phi U_\Phi' + U_{\Phi,\perp}\Lambda_{\Phi,\perp}U_{\Phi,\perp}')^k [(U_\Phi \Lambda_\Phi U_\Phi' + U_{\Phi,\perp}\Lambda_{\Phi,\perp}U_{\Phi,\perp}')']^k \\
&= \sigma^2 \left(U_\Phi \Lambda_\Gamma U_\Phi' + U_{\Phi,\perp}\Lambda_{\Gamma,\perp}U_{\Phi,\perp}' \right)
\end{aligned}$$

$\square$

Properties of the NIRVAR Estimator

Bias

The NIRVAR estimator is **biased** whenever $\hat{A}_{ij} = 0$ and $A_{ij} = 1$ (model misspecification) but unbiased otherwise.

The bias is given by

$$C := \{R(\hat{A})'(ZZ' \otimes \Sigma^{-1})R(\hat{A})]^{-1} R(\hat{A})'(ZZ' \otimes \Sigma^{-1})R(A).$$

The estimator is therefore suited to highly **assortative** SBMs.

Consistency, and Asymptotic Normality

Proposition (Consistency, Asymptotic Normality)

The NIRVAR estimator, $\hat{\gamma}(\hat{A})$ is a consistent estimator of $C\boldsymbol{\gamma}(A)$ where C determines the bias, and

$$\sqrt{T}\left\{\hat{\gamma}(\hat{A}) - C\boldsymbol{\gamma}(A)\right\} \xrightarrow{d} \mathcal{N}\left(0, \left\{R(\hat{A})'\left(\Gamma \otimes \boldsymbol{\Sigma}^{-1}\right)R(\hat{A})\right\}^{-1}\right),$$

where $\Gamma := \mathbb{E}(Z_t Z_t') = plim ZZ'/T$.

Simulation Studies

Between block probability

Define $p^{(\text{in})}$ to be the *intra*-block probability and $p^{(\text{out})}$ to be the *inter*-block probability. So

$$B = \begin{pmatrix} p^{(\text{in})} & p^{(\text{out})} & \cdots \\ p^{(\text{out})} & p^{(\text{in})} & \\ \vdots & & \ddots \end{pmatrix}$$

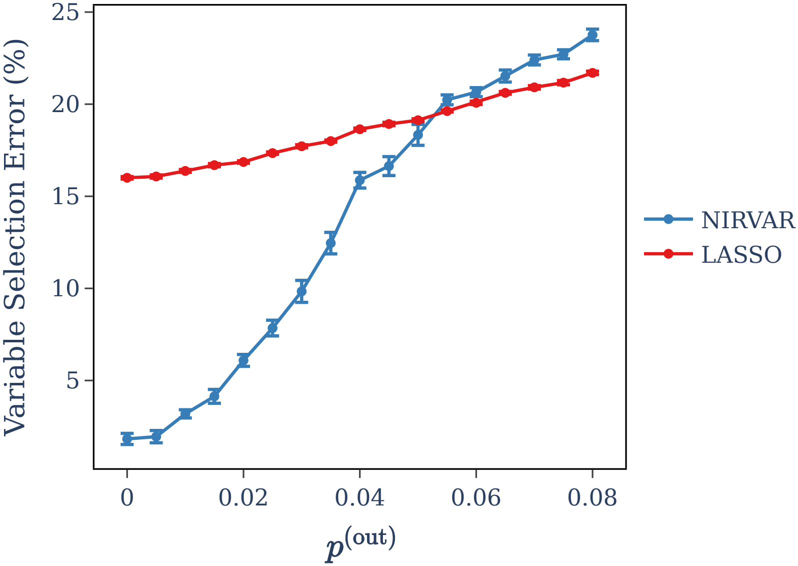
We simulate from a NIRVAR model with $N = 100$, $T = 1000$, $K = 10$, and $p^{(\text{in})} = 1$.

We vary $p^{(\text{out})}$ and computed the percentage of incorrect entries of $\hat{A}$ as a function of $p^{(\text{out})}$.

The percentage error was calculated as $100 \times \sum_{i,j=1}^{N} \mathbb{1}\{\hat{A}_{ij} \neq A_{ij}\}/N^2$

We compared to the percentage variable selection error of a LASSO estimator, with penalty chosen using the Akaike Information Criterion (AIC).

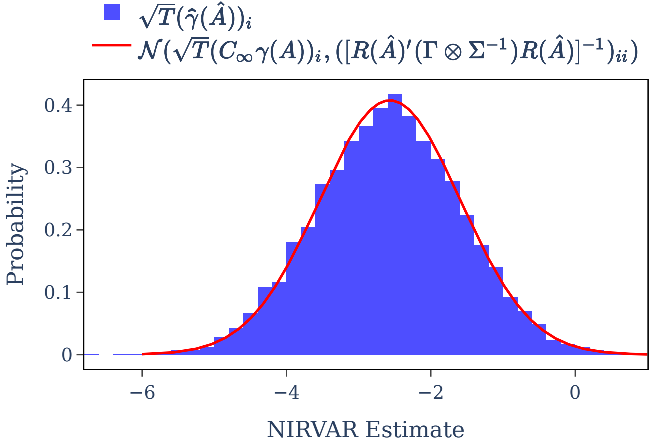
Between block probability



Large sample distribution of the NIRVAR estimator

We simulated 10,000 replica datasets from a NIRVAR model with $N = 50$, $T = 5000$, $K = 5$, $p^{(\text{in})} = 0.75$, and $p^{(\text{out})} = 0.2$.

We compare the asymptotic distribution with the empirical distribution of $\sqrt{T}\hat{\gamma}(\hat{A})_i$.



Latent Position Recovery

We fix two ground truth block latent positions $(Y_B)_1 = (0.05, 0.95)'$ and $(Y_B)_2 = (0.95, 0.05)'$.

We simulate from the corresponding NIRVAR model with $N = 150$, $T = 2000$, $K = 2$, $z_1, \ldots, z_{75} = 1$ and $z_{76}, \ldots, z_{150} = 2$. We repeat this 4000 times.

We compare the NIRVAR embedded points, $\hat{Y}$, to the ground truth latent positions.

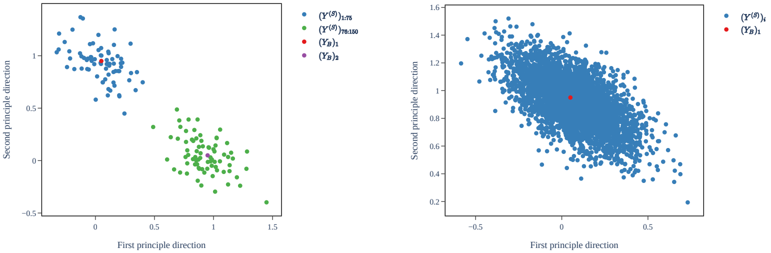


Figure 1: Left: NIRVAR embedded points alongside the ground truth block latent positions Right: replicas of NIRVAR embedding, $\hat{\boldsymbol{y}}_i$, alongside $\boldsymbol{y}_{z_i}$.

Applications

Macroeconomic Application: FRED-MD

FRED-MD is a publicly accessible database of monthly observations of macroeconomic variables [1] [McCracken and Ng, 2016].

The prediction task is one-step ahead forecasts of the first order difference of the logarithm of the monthly industrial production (IP) index.

We backtest NIRVAR, FARM, FNETS, and GNAR from January 2000 - December 2019 using a rolling window framework with a lookback window of 480 observations.

Table 1: Overall MSE of each model for the task of forecasting US IP.

Metric	NIRVAR	NIRVAR*	FARM	FNETS	GNAR
Overall MSE	**0.0087**	0.0097	0.0089	0.0096	0.0101

*NIRVAR estimator using the 8 FRED-MD defined groups.

[1] https://research.stlouisfed.org/econ/mccracken/fred-databases/

Macroeconomic Application: FRED-MD

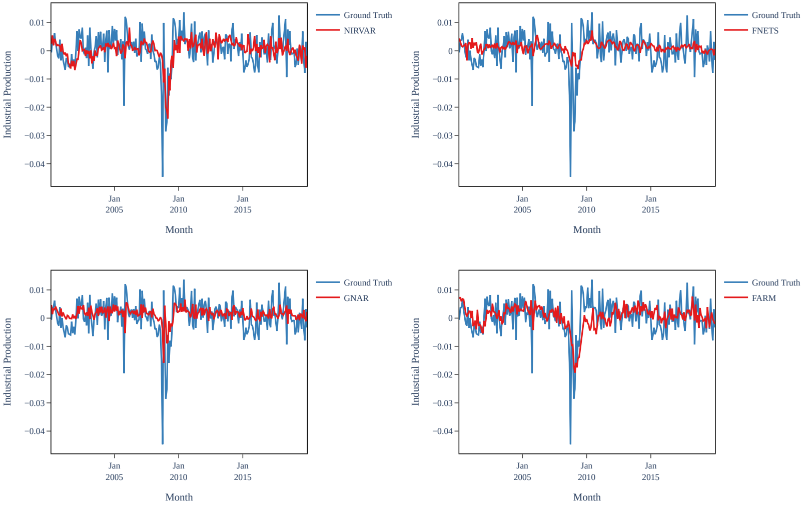
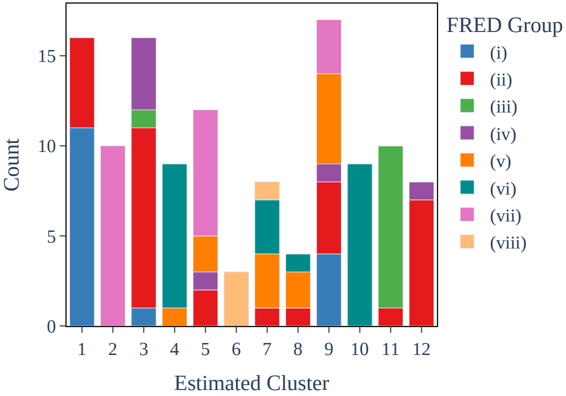


Figure 2: Predicted (log-differenced) IP against the realised (log-differenced) IP for each model.

Macroeconomic Application: FRED-MD



The FRED-MD variables are divided into eight groups: (i) output and income; (ii) labour market; (iii) housing; (iv) consumption, orders, and inventories; (v) money and credit; (vi) interest and exchange rates; (vii) prices.

Application to Financial Returns Prediction

The previous close-to-close (pvCLCL) and open-to-close (OPCL) market excess returns of 648 financial assets from 03/01/2000 - 31/12/2020 were derived from databases provided by the CRSP. [2]

The task is to predict the sign of the next day pvCLCL market excess returns (a positive (negative) sign corresponds to a long (short) position in the asset).

We backtest NIRVAR, FARM, FNETS, and GNAR using a rolling window from 01/01/2004 - 31/12/2020 with a look-back window of four years.

[2] CRSP, LLC, is an affiliate of the University of Chicago Booth School of Business.

Application to Financial Returns Prediction

The previous close-to-close (pvCLCL) and open-to-close (OPCL) market excess returns of 648 financial assets from 03/01/2000 - 31/12/2020 were derived from databases provided by the CRSP. [2]

The task is to predict the sign of the next day pvCLCL market excess returns (a positive (negative) sign corresponds to a long (short) position in the asset).

We backtest NIRVAR, FARM, FNETS, and GNAR using a rolling window from 01/01/2004 - 31/12/2020 with a look-back window of four years.

Table 2: Statistics on the financial returns predictive performance over the backtesting period.

Metric	NIRVAR	FARM	FNETS	GNAR
Sharpe Ratio	2.82	0.22	0.78	0.70
Sortino Ratio	4.80	0.36	1.39	1.13
Mean Turnover (%)	50.3	51.1	50.0	43.0
Maximum Drawdown (%)	61	531	107	257
Hit Ratio (%)	50.7	48.7	50.2	41.5
Long Ratio (%)	50.1	49.0	50.1	40.7
Mean Daily PnL (bpts)	3.00	0.44	0.89	1.10
Market Correlation	0.017	0.000	0.004	0.011

[2] CRSP, LLC, is an affiliate of the University of Chicago Booth School of Business.

Application to Financial Returns Prediction

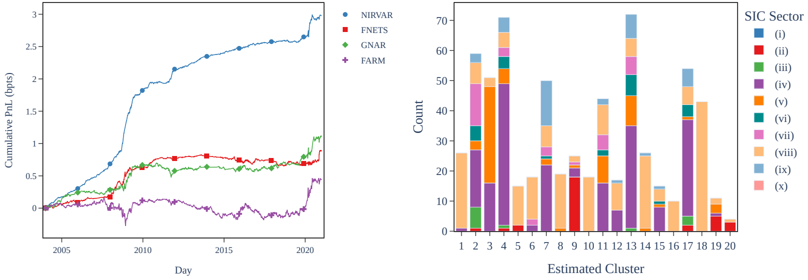


Figure 3: Left: the cumulative PnL in bpts over the backtesting period. Right: comparison of the NIRVAR estimated clusters with the SIC groups on 31/12/2020.

(i) Agriculture, Forestry, and Fishing, (ii) Mining, (iii) Construction, (iv) Manufacturing, (v) Transportation and Public Utilities, (vi)Wholesale Trade, (vii) Retail Trade, (viii) Finance, Insurance, and Real Estate, (ix) Services, and (x) Public Administration.

Santander Bicycle Rides

The first differences of the log daily number of bicycle rides from $N = 774$ Santander stations in London from 07/03/2018 until 10/03/2020 ($T = 735$) were obtained using records from TfL Open Data [3].

[3] https://cycling.data.tfl.gov.uk/

Santander Bicycle Rides: NIRVAR Clusters

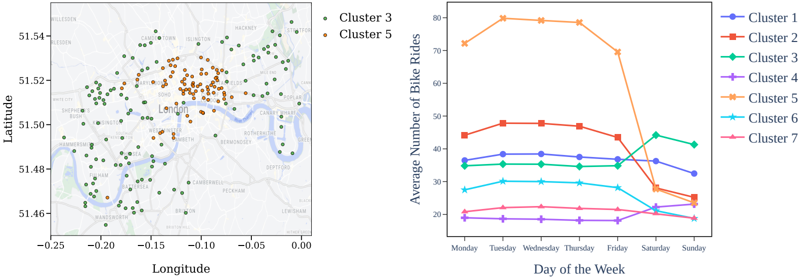


Figure 4: NIRVAR on the Santander Cycles dataset. (a) Clusters 3 and 5. (b) Average number of bicycle rides across each of the $K = 7$ clusters for weekday.

Santander Bicycle Rides: Forecasting

One-day-ahead predictions from NIRVAR, FARM, FNETS and GNAR are obtained using a rolling window backtesting framework from 09/02/2020 until 10/03/2020 (30 days).

The overall MSE is homogeneous across models with NIRVAR achieving the lowest value (**0.364**), followed by FARM (0.370), GNAR (0.374) and FNETS (0.388).

Summary and Future Directions

- We model a panel of multivariate time series as a VAR process whose parameter matrix is a realisation of a SBM.

- We introduce an **estimation framework for sparse VAR** models that
  1. determines the restrictions to be placed on the VAR parameters
  2. estimates the remaining unrestricted parameters via OLS estimation.

- The framework allows for a network based approach to determining the VAR parameters when the underlying network is **unobserved**.

Summary and Future Directions

- We model a panel of multivariate time series as a VAR process whose parameter matrix is a realisation of a SBM.
- We introduce an **estimation framework for sparse VAR** models that
  1. determines the restrictions to be placed on the VAR parameters
  2. estimates the remaining unrestricted parameters via OLS estimation.
- The framework allows for a network based approach to determining the VAR parameters when the underlying network is **unobserved**.

Current and future work:

- Prove that NIRVAR **recovers the latent positions** of the SBM up to an orthogonal transformation.
- Incorporate a **factor model** into the NIRVAR framework.
- Modify the method to allow for **vertices between blocks**.
- Extend the model to incorporate a **time varying adjacency matrix**.

THANK YOU!

References

M. Barigozzi, H. Cho, and D. Owens. Fnets: Factor-adjusted network estimation and forecasting for high-dimensional time series. *Journal of Business & Economic Statistics*, pages 1–13, 2023.

J. Fan, R. P. Masini, and M. C. Medeiros. Bridging factor and sparse models. *The Annals of Statistics*, 51(4):1692–1717, 2023.

G. S. Guðmundsson and C. Brownlees. Detecting groups in large vector autoregressions. *Journal of Econometrics*, 225(1):2–26, 2021.

M. Knight, K. Leeming, G. Nason, and M. Nunes. Generalized network autoregressive processes and the gnar package. *Journal of Statistical Software*, 96:1–36, 2020.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460): 1167–1179, 2002.

NIRVAR: **Python:**