Factor-adjusted network estimation and forecasting for high-dimensional time series

Haeran Cho*

Joint work with Matteo Barigozzi (U. Bologna) & Dom Owens (U. Bristol)

CRiSM Seminar, Warwick
March 2003

*School of Mathematics, University of Bristol. Supported by the Leverhulme Trust Research Project Grant (RPG-2019-390).

Vector autoregressive (VAR) models

Model a zero-mean process $\mathbf{X}_t = (X_{1t}, \ldots, X_{pt})^\top$ as

$$\mathbf{X}_t = \mathbf{A}_1 \mathbf{X}_{t-1} + \ldots + \mathbf{A}_d \mathbf{X}_{t-d} + \mathbf{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim_{\text{iid}} (\mathbf{0}, \mathbf{I}).$$

Applications in finance (Barigozzi & Hallin, 2017; Barigozzi & Brownlees, 2019; Basu et al., 2019), neuroscience (Kirch et al., 2015; Wang et al., 2012), systems biology (Shojaie & Michailidis, 2010).

VAR modelling enables **inferring dynamic interdependence between the variables** as well as **forecasting** of the future.

Three networks under VAR model

$$\mathbf{X}_t = \mathbf{A}_1 \mathbf{X}_{t-1} + \ldots + \mathbf{A}_d \mathbf{X}_{t-d} + \mathbf{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim_{\text{iid}} (\mathbf{0}, \mathbf{I}).$$

Let $\mathcal{V} = \{1, \ldots, p\}$ denote the set of vertices.

Directed network $\mathcal{N}^{\mathsf{G}} = (\mathcal{V}, \mathcal{E}^{\mathsf{G}})$ representing Granger causal linkages:

$$\mathcal{E}^{\mathsf{G}} = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : \ A_{\ell, ii'} \neq 0 \text{ for some } 1 \leq \ell \leq d\right\}.$$

An edge $(i, i') \in \mathcal{E}^{\mathsf{G}}$ indicates that $X_{i', t-\ell}$ Granger causes X_{it} at some lag $1 \leq \ell \leq d$.

Undirected network $\mathcal{N}^{\mathsf{C}} = (\mathcal{V}, \mathcal{E}^{\mathsf{C}})$ representing contemporaneous dependence in $\mathbf{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t$ by means of partial correlations (PC):
With the precision matrix $\mathbf{\Delta} = [\delta_{ii'}] = \mathbf{\Gamma}^{-1}$,

$$\mathcal{E}^{\mathsf{C}} = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : \ i \neq i' \text{ and } -\frac{\delta_{ii'}}{\sqrt{\delta_{ii} \cdot \delta_{i'i'}}} \neq 0\right\}.$$

Undirected network $\mathcal{N}^{\mathsf{L}} = (\mathcal{V}, \mathcal{E}^{\mathsf{L}})$ summarising lead-lag and contemporaneous relations in $\mathbf{X}_t$ by means of the long-run partial correlations (LRPC): With $\boldsymbol{\Omega} = [\omega_{ii'},\, 1 \le i, i' \le p] = \boldsymbol{\Sigma}_x^{-1}(0) = 2\pi(\mathcal{A}(1))^\top \boldsymbol{\Delta} \mathcal{A}(1)$,

$$\mathcal{E}^{\mathsf{L}} = \left\{ (i, i') \in \mathcal{V} \times \mathcal{V} :\ i \ne i' \text{ and } -\frac{\omega_{ii'}}{\sqrt{\omega_{ii} \cdot \omega_{i'i'}}} \ne 0 \right\},$$

where $\mathcal{A}(1) = \mathbf{I} - \sum_{\ell=1}^{d} \mathbf{A}_\ell$.

Aim: Estimate the three networks permitting $p \to \infty$.

Stability and sparsity in high dimensions

VAR modelling quickly becomes high-dimensional as p increases, hence ℓ_1-regularisation methods have been developed assuming sparsity of $\mathbf{A}_\ell$ (Basu & Michailidis, 2015; Han et al., 2015).

For their consistency, it is required that (Basu & Michailidis, 2015)

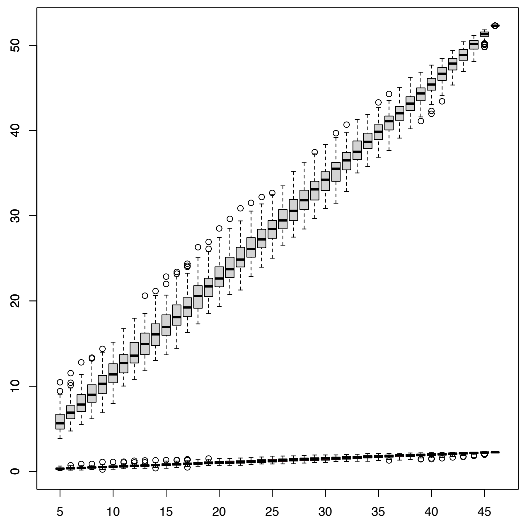
$$\sup_{\omega \in [-\pi, \pi]} \Lambda_{\max}(\mathbf{\Sigma}_x(\omega)) < \infty$$

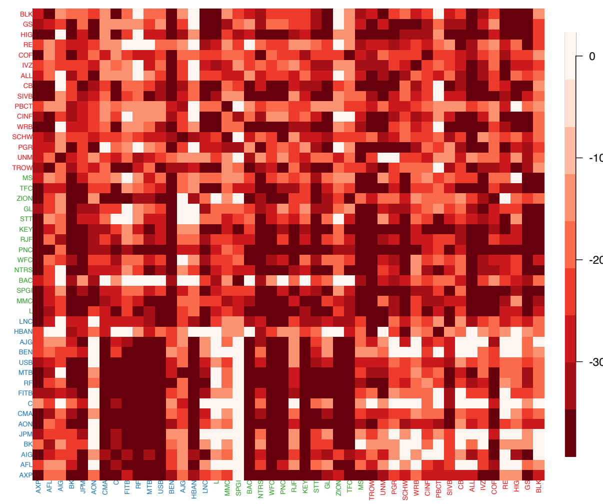
† uniform boundedness of the largest eigenvalue of the spectral density matrix

which implies that $\mathbf{A}_\ell$ is (weakly) sparse (Lin & Michailidis, 2020).

Difficult to identify sparse predictive representations for real-life datasets observed e.g. in economics and finance (Giannone et al., 2021).

Panel of volatility measures with $p = 46$ and $n = 252$ (03/2008 to 02/2009).





Left: two largest eigenvalues of the long-run covariance matrix $\widehat{\mathbf{\Sigma}}_x(0)$ with subsets of cross-sections randomly sampled 100 times for each given dimension $p \in \{5, \ldots, 46\}$. **Right:** logged and truncated p-values from fitting a VAR(5) model via ridge regression (truncation level chosen by Bonferroni correction; minimum p-value over the five lags).

Stability and sparsity in high dimensions

VAR modelling quickly becomes high-dimensional as p increases, hence ℓ_1-regularisation methods have been developed assuming sparsity of $\mathbf{A}_\ell$ (Basu & Michailidis, 2015; Han et al., 2015).

For their consistency, it is required that (Basu & Michailidis, 2015)

$$\sup_{\omega \in [-\pi, \pi]} \Lambda_{\max}(\boldsymbol{\Sigma}_x(\omega)) < \infty$$

† uniform boundedness of the largest eigenvalue of the spectral density matrix

which implies that $\mathbf{A}_\ell$ is (weakly) sparse (Lin & Michailidis, 2020).

Factor-adjusted regression (Fan et al., 2020, 2021, 2023; Krampe & Margaritella, 2021): **dominant** (auto)correlations are addressed by a finite number of **common factors**, justifying sparsity imposed on remaining **idiosyncratic** component.

Our contributions

Propose FNETS methodology for **network estimation** and **forecasting** under factor-adjusted VAR model.

Fully address the challenges arising from not directly observing the latent VAR process, both methodologically and theoretically.

– Most general approach to dynamic factor modelling.
– ℓ_1-regularised Yule-Walker estimators, distinguished from the existing factor-adjusted regression modelling approach.

Show estimation and forecasting consistency in a general setting permitting **heavy-tailedness** and **'weak' factors**.

R package `fnets` available on CRAN.

Factor-adjusted VAR model

$\mathbf{X}_t \in \mathbb{R}^p$ is decomposed into two **latent** components: $\mathbf{X}_t = \boldsymbol{\xi}_t + \boldsymbol{\chi}_t$,

$$\boldsymbol{\xi}_t = \sum_{\ell=1}^d \mathbf{A}_\ell \boldsymbol{\xi}_{t-\ell} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t, \; \boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \mathbf{I}_p),$$
$$\boldsymbol{\chi}_t = \underbrace{\mathcal{B}(L)}_{\text{loadings}} \underbrace{\mathbf{u}_t}_{\text{factors}} = \sum_{\ell=0}^\infty \mathbf{B}_\ell \mathbf{u}_{t-\ell}, \; \mathbf{u}_t \sim (\mathbf{0}, \mathbf{I}_q).$$

That is, $\mathbf{X}_t = \sum_{\ell=0}^\infty \left(\sum_{k=1}^\ell \mathbf{A}_k \mathbf{B}_{\ell-k} \right) \mathbf{u}_{t-\ell} + \sum_{\ell=1}^d \mathbf{A}_\ell \mathbf{X}_{t-\ell} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t$.

Adopting generalised dynamic factor model (GDFM, Forni et al., 2000), $\boldsymbol{\chi}_t$ is driven by q-dimensional **common factors** $\mathbf{u}_{t-\ell}$, $\ell \ge 0$.
Cf. static factor model $\boldsymbol{\chi}_t = \sum_{\ell=0}^{M} \mathbf{B}_\ell \mathbf{u}_{t-\ell}$.

Our aim is to (i) **estimate networks** underpinning the latent VAR process $\boldsymbol{\xi}_t$ under appropriate sparsity assumptions, and (ii) **forecast** $\mathbf{X}_{n+a}$ for some $a \ge 1$ given $\mathbf{X}_t$, $t \le n$.

Assumptions for model identifiability

Let $\mathbf{\Sigma}_\chi(\omega)$ denote the spectral density matrix of $\boldsymbol{\chi}_t$ and $\mu_{\chi,j}(\omega)$ its j-th largest eigenvalue. Similarly define $\mathbf{\Sigma}_\xi(\omega)$ and $\mu_{\xi,j}(\omega)$.

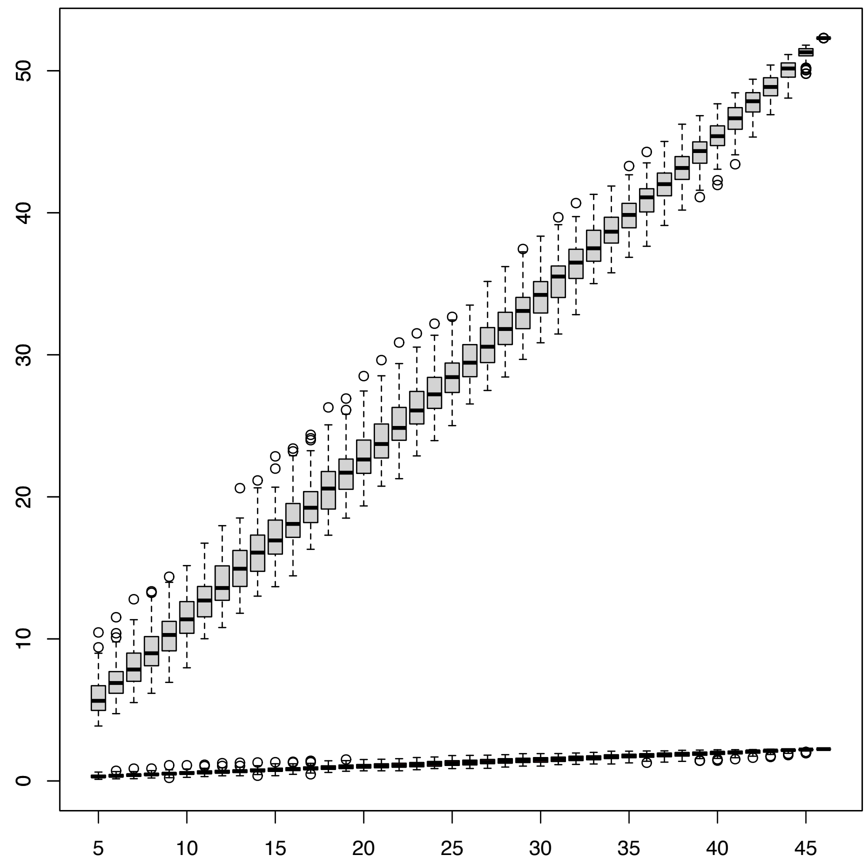
On factor-driven $\boldsymbol{\chi}_t$:

There exist $3/4 < \rho_q \le \ldots \le \rho_1 \le 1$, and functions $\omega \mapsto \alpha_{\chi,j}(\omega)$ and $\omega \mapsto \beta_{\chi,j}(\omega)$ for $\omega \in [-\pi, \pi]$, such that for all $p \ge p_0$,

$$\beta_{\chi,1}(\omega) \ge \frac{\mu_{\chi,1}(\omega)}{p^{\rho_1}} \ge \alpha_{\chi,1}(\omega) > \ldots > \beta_{\chi,q}(\omega) \ge \frac{\mu_{\chi,q}(\omega)}{p^{\rho_q}} \ge \alpha_{\chi,q}(\omega) > 0.$$

† Weak factors when $\rho_j < 1$, and cross-sectional ordering matters.

† Divergence of $\mu_{\chi,j}(\omega)$ is necessary and sufficient for GDFM representation.



0
10
20
30
40
50
5
10
15
20
25
30
35
40
45

For stability of $\boldsymbol{\xi}_t$ and controlling dependence (Zhang & Wu, 2021):

On the Wold representation, $\boldsymbol{\xi}_t = \sum_{\ell=0}^{\infty} \mathbf{D}_\ell \boldsymbol{\Gamma}^{1/2} \boldsymbol{\varepsilon}_{t-\ell}$, we have $|D_{\ell,ik}| \le C_{ik}(1+\ell)^{-\varsigma}$ for all $\ell \ge 0$, with

$$\max\left\{ \max_{1\le k\le p} \sum_{i=1}^{p} C_{ik},\ \max_{1\le i\le p} \sum_{k=1}^{p} C_{ik},\ \max_{1\le i\le p} \sqrt{\sum_{k=1}^{p} C_{ik}^2} \right\} \le \Xi.$$

for some constants $\Xi > 0$ and $\varsigma > 2$. † Holds e.g. when $d = 1$ and $|\mathbf{A}_1|_\infty < 1$.

Then, $\exists\ B_\xi > 0$ such that the largest eigenvalue of $\boldsymbol{\Sigma}_\xi(\omega)$ is uniformly bounded, i.e.

$$\sup_\omega \mu_{\xi,1}(\omega) \le B_\xi.$$

$\therefore$ Latent $\boldsymbol{\xi}_t$ and $\boldsymbol{\chi}_t$ are **identifiable** as by Weyl's inequality, $\mu_{x,q}(\omega) \to \infty$ while $\mu_{x,q+1}(\omega) \le B_\xi$ for all p.

Network estimation via FNETS

Step 1: Perform factor-adjustment and estimate $\mathbf{\Gamma}_\xi(\ell) = \mathsf{E}(\boldsymbol{\xi}_{t-\ell}\boldsymbol{\xi}_t^\top)$ via dynamic PCA in frequency domain.

Step 2: Estimation of $\mathbf{A}_\ell$, $1 \le \ell \le d$, via ℓ_1-regularised Yule-Walker estimation, from which we estimate $\mathcal{N}^{\mathsf{G}} = (\mathcal{V}, \mathcal{E}^{\mathsf{G}})$:

$$\mathcal{E}^{\mathsf{G}} = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : \, A_{\ell,ii'} \ne 0 \text{ for some } 1 \le \ell \le d\right\}.$$

Step 3: Estimation of (long-run) partial correlations by estimating $\mathbf{\Delta} = \mathbf{\Gamma}^{-1}$ and $\mathbf{\Omega} = \mathbf{\Sigma}_\xi^{-1}(0) = 2\pi(\mathcal{A}(1))^\top \mathbf{\Delta} \mathcal{A}(1)$, from which we estimate $\mathcal{N}^{\mathsf{C}} = (\mathcal{V}, \mathcal{E}^{\mathsf{C}})$ and $\mathcal{N}^{\mathsf{L}} = (\mathcal{V}, \mathcal{E}^{\mathsf{L}})$:

$$\mathcal{E}^{\mathsf{C}} = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : \, i \ne i' \text{ and } -\frac{\delta_{ii'}}{\sqrt{\delta_{ii} \cdot \delta_{i'i'}}} \ne 0\right\},$$

$$\mathcal{E}^{\mathsf{L}} = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : \, i \ne i' \text{ and } -\frac{\omega_{ii'}}{\sqrt{\omega_{ii} \cdot \omega_{i'i'}}} \ne 0\right\}$$

Step 1: Factor adjustment

Exploit the large eigengap between $\mu_{x,q}(\omega)$ and $\mu_{x,q+1}(\omega)$.

With sample ACV $\widehat{\mathbf{\Gamma}}_x(\ell)$, estimate the spectral density of $\mathbf{X}_t$ by

$$\widehat{\mathbf{\Sigma}}_x(\omega) = \frac{1}{2\pi} \sum_{\ell=-m}^{m} K\left(\frac{\ell}{m}\right)\widehat{\mathbf{\Gamma}}_x(\ell) \exp(-\iota\ell\omega),$$

with the Bartlett kernel $K(\cdot)$ and bandwidth $m \asymp n^\beta$, $\beta \in (0,1)$.
Performing PCA on $\widehat{\mathbf{\Sigma}}_x(\omega_k)$ at Fourier frequencies $\omega_k = \frac{2\pi k}{2m+1}$, $|k| \le m$,

$$\widehat{\mathbf{\Sigma}}_x(\omega_k) = \sum_{j=1}^{p} \widehat{\mu}_{x,j}(\omega_k)\widehat{\mathbf{e}}_{x,j}(\omega_k)(\widehat{\mathbf{e}}_{x,j}(\omega_k))^*,$$

we estimate $\mathbf{\Sigma}_\chi(\omega_k)$ and $\mathbf{\Gamma}_\chi(\ell)$ as

$$\widehat{\mathbf{\Sigma}}_\chi(\omega_k) = \sum_{j=1}^{q} \widehat{\mu}_{x,j}(\omega_k)\widehat{\mathbf{e}}_{x,j}(\omega_k)(\widehat{\mathbf{e}}_{x,j}(\omega_k))^*,$$
$$\widehat{\mathbf{\Gamma}}_\chi(\ell) = \frac{2\pi}{2m+1} \sum_{k=-m}^{m} \widehat{\mathbf{\Sigma}}_\chi(\omega_k) \exp(\iota\ell\omega_k).$$

Straightforwardly, $\widehat{\mathbf{\Gamma}}_\xi(\ell) = \widehat{\mathbf{\Gamma}}_x(\ell) - \widehat{\mathbf{\Gamma}}_\chi(\ell)$.

Step 2: Estimation of VAR parameters

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_n^\top \\ \vdots \\ \boldsymbol{\xi}_{d+1}^\top \end{bmatrix}}_{\mathcal{Y} \in \mathbb{R}^{(n-d) \times p}} = \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{n-1}^\top & \boldsymbol{\xi}_{n-2}^\top & \cdots & \boldsymbol{\xi}_{n-d}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_d^\top & \boldsymbol{\xi}_{d-1}^\top & \cdots & \boldsymbol{\xi}_1^\top \end{bmatrix}}_{\mathcal{X} \in \mathbb{R}^{(n-d) \times (dp)}} \underbrace{\begin{bmatrix} \mathbf{A}_1^\top \\ \vdots \\ \mathbf{A}_d^\top \end{bmatrix}}_{\boldsymbol{\beta}^0 \in \mathbb{R}^{(dp) \times p}} + \underbrace{\begin{bmatrix} \boldsymbol{\varepsilon}_n^\top \\ \vdots \\ \boldsymbol{\varepsilon}_{d+1}^\top \end{bmatrix}}_{\mathcal{E} \in \mathbb{R}^{(n-d) \times p}} \boldsymbol{\Gamma}^{1/2}.$$

Existing approach: Estimate the latent $\boldsymbol{\xi}_t$, $1 \le t \le n$, and apply ℓ_1-regularisation methods such as Lasso to estimate $\boldsymbol{\beta}^0$ (**FARM**, Fan et al., 2021).

$\Longrightarrow$ Loss of statistical efficiency due to estimating the $n \times p$ matrix containing $\boldsymbol{\xi}_t$, $1 \le t \le n$.

ℓ_1-regularised Yule-Walker estimators

Via Yule-Walker equation, we have $\mathbb{G}\boldsymbol{\beta}^0 = \mathbb{g}$ where

$$\mathbb{G} = \begin{bmatrix} \boldsymbol{\Gamma}_\xi(0) & \boldsymbol{\Gamma}_\xi(-1) & \ldots & \boldsymbol{\Gamma}_\xi(-d+1) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_\xi(d-1) & \boldsymbol{\Gamma}_\xi(d-2) & \ldots & \boldsymbol{\Gamma}_\xi(0) \end{bmatrix} \text{ and } \mathbb{g} = \begin{bmatrix} \boldsymbol{\Gamma}_\xi(1) \\ \vdots \\ \boldsymbol{\Gamma}_\xi(d) \end{bmatrix}.$$

With $\widehat{\boldsymbol{\Gamma}}_\xi(\ell)$ from Step 1 estimating $\boldsymbol{\Gamma}_\xi(\ell)$, we obtain surrogate matrices $\widehat{\mathbb{G}}$ and $\widehat{\mathbb{g}}$.

Lasso (ℓ_1-penalised minimisation):

$$\widehat{\boldsymbol{\beta}}^{\mathsf{las}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{pd \times p}} \mathrm{tr}\left(\boldsymbol{\beta}^\top \widehat{\mathbb{G}} \boldsymbol{\beta} - 2\boldsymbol{\beta}^\top \widehat{\mathbb{g}}\right) + \lambda^{\mathsf{las}} |\boldsymbol{\beta}|_1.$$

Dantzig selector (constrained ℓ_1-minimisation):

$$\widehat{\boldsymbol{\beta}}^{\mathsf{DS}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{pd \times p}} |\boldsymbol{\beta}|_1 \quad \text{subject to} \quad \left|\widehat{\mathbb{G}}\boldsymbol{\beta} - \widehat{\mathbb{g}}\right|_\infty \le \lambda^{\mathsf{DS}}.$$

Step 3: Estimation of (long-run) partial correlations

From Yule-Walker equation, estimate $\mathbf{\Gamma} = \mathsf{Cov}(\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t)$ by

$$\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{\Gamma}}_\xi(0) - \textstyle\sum_{\ell=1}^d \widehat{\mathbf{A}}_\ell \widehat{\mathbf{\Gamma}}_\xi(\ell).$$

We estimate $\mathbf{\Delta} = \mathbf{\Gamma}^{-1}$ via constrained ℓ_1-minimisation (Cai et al., 2011),

$$\widehat{\mathbf{\Delta}} = \arg\min_{\mathbf{M} \in \mathbb{R}^{p \times p}} |\mathbf{M}|_1 \quad \text{subject to} \quad \left| \widehat{\mathbf{\Gamma}}\mathbf{M} - \mathbf{I} \right|_\infty \le \eta.$$

Replacing $\mathcal{A}(1) = \mathbf{I} - \sum_\ell \mathbf{A}_\ell$ and $\mathbf{\Delta}$ with their estimators, we estimate $\mathbf{\Omega} = \mathbf{\Sigma}_\xi^{-1}(0) = 2\pi(\mathcal{A}(1))^\top \mathbf{\Delta} \mathcal{A}(1)$ by

$$\widehat{\mathbf{\Omega}} = 2\pi(\widehat{\mathcal{A}}(1))^\top \widehat{\mathbf{\Delta}} \widehat{\mathcal{A}}(1).$$

Theoretical consistency

$$\mathbf{X}_t = \underbrace{\sum_{m=0}^{\infty} \mathbf{B}_m \mathbf{u}_{t-m}}_{\boldsymbol{\chi}_t} + \underbrace{\sum_{\ell=1}^{d} \mathbf{A}_\ell \boldsymbol{\xi}_{t-\ell} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\varepsilon}_t}_{\boldsymbol{\xi}_t} .$$

We assume $\max_{1 \le j \le q} \mathsf{E}(|u_{jt}|^\nu), \max_{1 \le i \le p} \mathsf{E}(|\varepsilon_{it}|^\nu) \le \mu_\nu$ for some $\underline{\nu > 4}$.

Each $\{u_{jt}\}_{t \in \mathbb{Z}}$ (resp. $\{\varepsilon_{it}\}_{t \in \mathbb{Z}}$) is a sequence of martingale differences and u_{jt}, $1 \le j \le q$ (resp. ε_{it}, $1 \le i \le p$) are i.i.d. for given t.

To ease presentation, let all the factors be <u>strong</u> (i.e. $\mu_{\chi,q}(\omega) \asymp p$).

Consistency of Step 1: Factor adjustment

As $n, p \to \infty$, we have $\mathsf{P}(\mathcal{E}_{n,p}) \to 1$ where

$$\mathcal{E}_{n,p} = \Big\{ \max_{\ell \le d} |\widehat{\mathbf{\Gamma}}_\xi(\ell) - \mathbf{\Gamma}_\xi(\ell)|_\infty \lesssim \underbrace{\frac{1}{m} \vee \vartheta_{n,p}}_{\text{bias-var trade-off in } \widehat{\mathbf{\Sigma}}_x(\omega)} \vee \underbrace{\frac{1}{\sqrt{p}}}_{\text{latency of } \boldsymbol{\xi}_t} \Big\},$$

and $\vartheta_{n,p} = \left(\frac{m p^{2/\nu} \log^{7/2}(p)}{n^{1-2/\nu}} \vee \sqrt{\frac{m \log(mp)}{n}} \right)$.

In what follows, all results are conditional on $\mathcal{E}_{n,p}$.

Consistency of Step 2

Degrees of sparsity associated with $\boldsymbol{\beta}^0$ are defined as

$$s_{\mathsf{in}} = \max_{1 \le j \le p} s_{0,j} \text{ where } s_{0,j} = \sum_{\ell=1}^{d} |\mathbf{A}_{\ell,j\cdot}|_0 = |\boldsymbol{\beta}^0_{\cdot j}|_0.$$

Suppose that $\lambda^{\mathsf{las}} \gtrsim (\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2})$ and $s_{\mathsf{in}}(\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2}) \lesssim m_\xi \le \inf_\omega \mu_{\xi,p}(\omega)$. Then, for all $j = 1, \ldots, p$,

$$\left|\widehat{\boldsymbol{\beta}}^{\mathsf{las}}_{\cdot j} - \boldsymbol{\beta}^0_{\cdot j}\right|_\infty \le \left|\widehat{\boldsymbol{\beta}}^{\mathsf{las}}_{\cdot j} - \boldsymbol{\beta}^0_{\cdot j}\right|_2 \le \frac{32\sqrt{s_{\mathsf{in}}}\lambda^{\mathsf{las}}}{\pi m_\xi} = \mathfrak{t}.$$

Under 'beta-min' condition: $\min_{(i,j):\, \beta^0_{ij} \neq 0} |\beta^0_{ij}| \ge 2\mathfrak{t}$, $\mathcal{N}^{\mathsf{G}}$ is consistently estimated by hard-thresholding $\widehat{\boldsymbol{\beta}}^{\mathsf{las}}$ as

$$\widetilde{\boldsymbol{\beta}}^{\mathsf{las}}(\mathfrak{t}) = \left[\widehat{\beta}^{\mathsf{las}}_{ij} \cdot \mathbb{I}_{\{|\widehat{\beta}^{\mathsf{las}}_{ij}| > \mathfrak{t}\}},\ 1 \le i \le pd,\ 1 \le j \le p\right].$$

Consistency of Step 3

Suppose that $\eta \gtrsim \|\mathbf{\Delta}\|_1 s_{\text{in}}(\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2})$. Then,

$$|\widehat{\mathbf{\Delta}} - \mathbf{\Delta}|_\infty \lesssim \|\mathbf{\Delta}\|_1 \eta = \|\mathbf{\Delta}\|_1^2 s_{\text{in}}(\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2}).$$

Additionally, if $\widehat{\mathbf{\Omega}}$ is obtained with $\widetilde{\beta}^{\text{las}}(\mathfrak{t})$,

$$|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}|_\infty \lesssim \|\mathcal{A}(1)\|_1(\|\mathbf{\Delta}\| s_{\text{out}} \mathfrak{t} + \|\mathcal{A}(1)\|_1 \; \|\mathbf{\Delta}\|_1 \eta),$$

where $s_{\text{out}} = \max_{1 \le j \le p} \sum_{\ell=1}^{d} |\mathbf{A}_{\ell,\cdot j}|_0$.

Edge sets of $\mathcal{N}^{\text{C}}$ and $\mathcal{N}^{\text{L}}$ are estimated by hard-thresholding $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{\Omega}}$, respectively, and consistency is achieved under analogous ‘beta-min’ conditions.

Consider the case where $\underline{\boldsymbol{\chi}_t = \sum_{m=0}^{M} \mathbf{B}_m \mathbf{u}_{t-m}}$, $\underline{n \asymp p}$, $\underline{\nu > 8}$. Then,

$$\max_j |\widehat{\boldsymbol{\beta}}^{\mathsf{las}}_{\cdot j} - \boldsymbol{\beta}^0_{\cdot j}|_2 = O_P(\sqrt{s_{\mathsf{in}} \log(p)/n})$$

$\widehat{\beta}^{\mathsf{las}}$ and its thresholded counterpart performs as well as the benchmark derived under independence and Gaussianity in the Lasso literature:

Cf. Lasso estimator applied to estimated $\boldsymbol{\xi}_t$ attains $O_P(\sqrt{s_{\mathsf{in}}} n^{-1/2+5/\nu})$ under strong mixingness (**FARM**, Fan et al., 2021).

$$|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}|_\infty = O_P(\|\boldsymbol{\Delta}\|_1^2 s_{\mathsf{in}} \sqrt{\log(p)/n}) \text{ and}$$
$$|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}|_\infty = O_P((s_{\mathsf{out}} \sqrt{s_{\mathsf{in}}} \vee \|\boldsymbol{\Delta}\|_1^2 s_{\mathsf{in}}) \sqrt{\log(p)/n}).$$

$\widehat{\boldsymbol{\Delta}}$ performs close to (up to s_{in}) the CLIME estimator (Cai et al., 2011) for sparse precision matrix from independent random vectors.

$\therefore$ FNETS estimators perform as well as benchmarks obtained where $\boldsymbol{\xi}_t$ is directly observed under independence.

Forecasting via FNETS

Forecast $\mathbf{X}_{n+a}$ given $\mathbf{X}_t,\, t \le n$ for some $a \ge 1$, by

$$\widehat{\mathbf{X}}_{n+a|n} = \widehat{\boldsymbol{\chi}}_{n+a|n} + \widehat{\boldsymbol{\xi}}_{n+a|n}, \quad \text{estimating}$$

$$\mathbf{X}_{n+a|n} = \underbrace{\mathsf{Proj}(\boldsymbol{\chi}_{n+a}|\boldsymbol{\chi}_v,\, v \le t)}_{\boldsymbol{\chi}_{n+a|n}} + \underbrace{\mathsf{Proj}(\boldsymbol{\xi}_{n+a}|\boldsymbol{\xi}_v,\, v \le t)}_{\boldsymbol{\xi}_{n+a|n}}$$

Forecasting of factor-driven component

Under a more restricted, **static** factor model $\boldsymbol{\chi}_t = \sum_{\ell=0}^{M} \mathbf{B}_\ell \mathbf{u}_{t-\ell}$ with the number of static factors $r = (M+1)q$,

$$\boldsymbol{\chi}_{n+a|n} = \boldsymbol{\Gamma}_\chi(-a)\boldsymbol{\Gamma}_\chi^-(0)\boldsymbol{\chi}_n.$$

This motivates $\widehat{\boldsymbol{\chi}}^{\text{res}}_{n+a|n} = \widehat{\boldsymbol{\Gamma}}_\chi(-a)\widehat{\boldsymbol{\Gamma}}_\chi^-(0)\mathbf{X}_n$, which achieves consistent estimation of $\boldsymbol{\chi}_{n+a|n}$:

$$\left|\widehat{\boldsymbol{\chi}}^{\text{res}}_{n+a|n} - \boldsymbol{\chi}_{n+a|n}\right|_\infty = O_p\left(\vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}}\right).$$

We also obtain the in-sample estimator of $\boldsymbol{\chi}_t$, $t \le n$, as

$\widehat{\boldsymbol{\chi}}^{\text{res}}_t = \widehat{\mathbf{E}}_\chi \widehat{\mathbf{E}}_\chi^\top \mathbf{X}_t$, where $\widehat{\mathbf{E}}_\chi$ contains r leading eigenvectors of $\widehat{\boldsymbol{\Gamma}}_\chi(0)$.

Forecasting of the latent VAR process

Under the VAR model,

$$\widehat{\boldsymbol{\xi}}_{n+a|n} = \sum_{\ell=1}^{\max(1,a)-1} \widehat{\mathbf{A}}_\ell \widehat{\boldsymbol{\xi}}_{n+a-\ell|n} + \sum_{\ell=\max(1,a)}^{d} \widehat{\mathbf{A}}_\ell \widehat{\boldsymbol{\xi}}_{n+a-\ell}.$$

which inherits the theoretical properties of $\widehat{\mathbf{A}}_\ell$ and in-sample estimator $\widehat{\boldsymbol{\xi}}_t = \mathbf{X}_t - \widehat{\boldsymbol{\chi}}_t$, $t \le n$, as

$$\begin{aligned} &\left| \widehat{\boldsymbol{\xi}}_{n+1|n} - \boldsymbol{\xi}_{n+1|n} \right|_\infty \\ &= O_P\left(\left(s_{\mathsf{in}} \log^{1/2}(p) p^{1/\nu} + \|\boldsymbol{\beta}^0\|_1 \right) \left(\vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right). \end{aligned}$$

Simulations

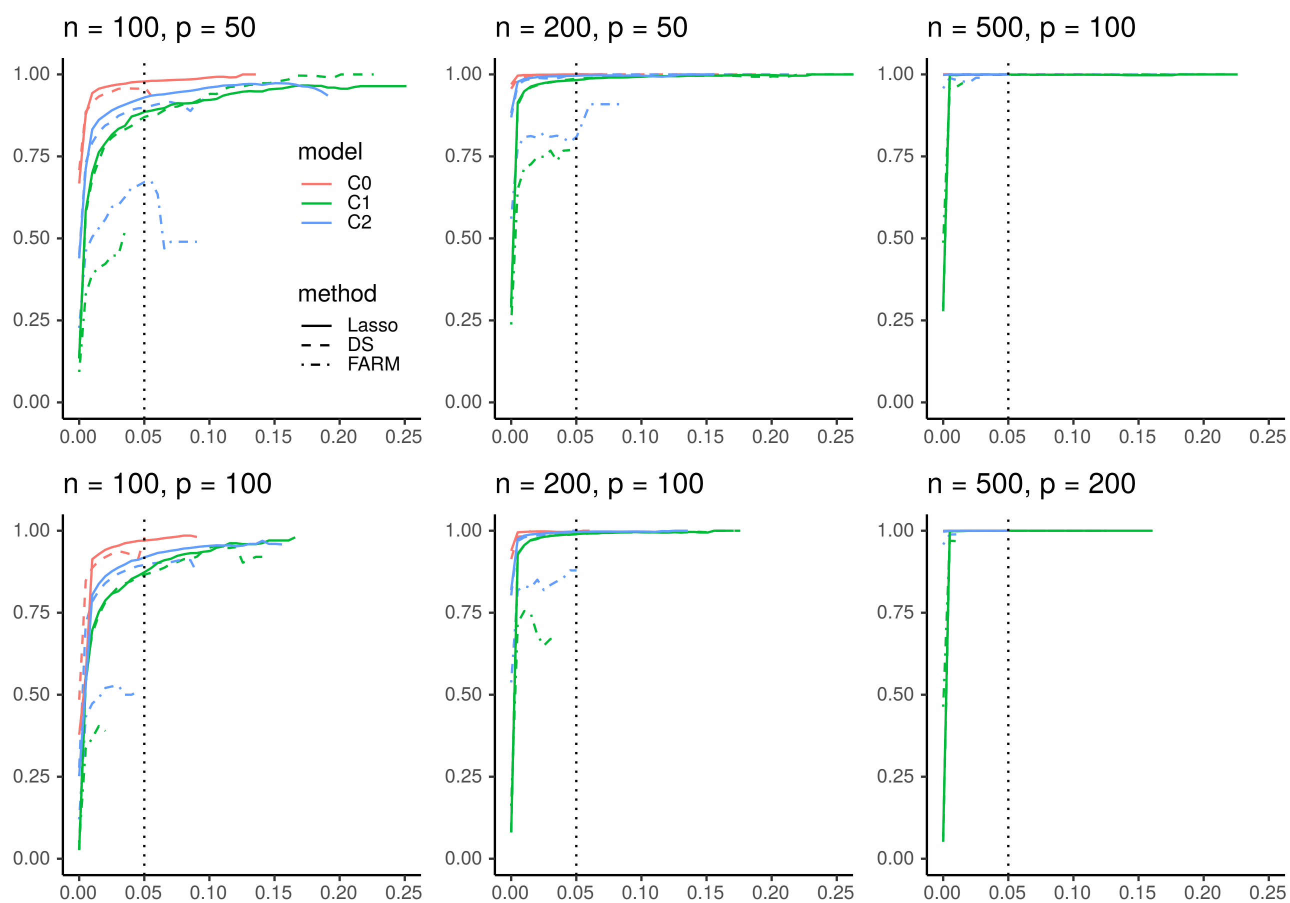
VAR process: $\boldsymbol{\xi}_t = \mathbf{A}\boldsymbol{\xi}_{t-1} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$ with supp$(\mathbf{A})$ generated as an Erdős-Rényi random graph, and
(E1) $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ and $\boldsymbol{\Gamma} = \mathbf{I}$.
(E2) $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ and $\boldsymbol{\Gamma} \neq \mathbf{I}$.
(E3) $\sqrt{5/3} \cdot \varepsilon_{it} \sim t_5$ and $\boldsymbol{\Gamma} = \mathbf{I}$.

Factor-driven component:
(C0) $\boldsymbol{\chi}_t = \mathbf{0}$ ('oracle').
(C1) $\boldsymbol{\chi}_t$ does not admit a static representation with $q = 2$.
(C2) $\boldsymbol{\chi}_t$ admits a static representation with $q = 2$ and $r = 4$.

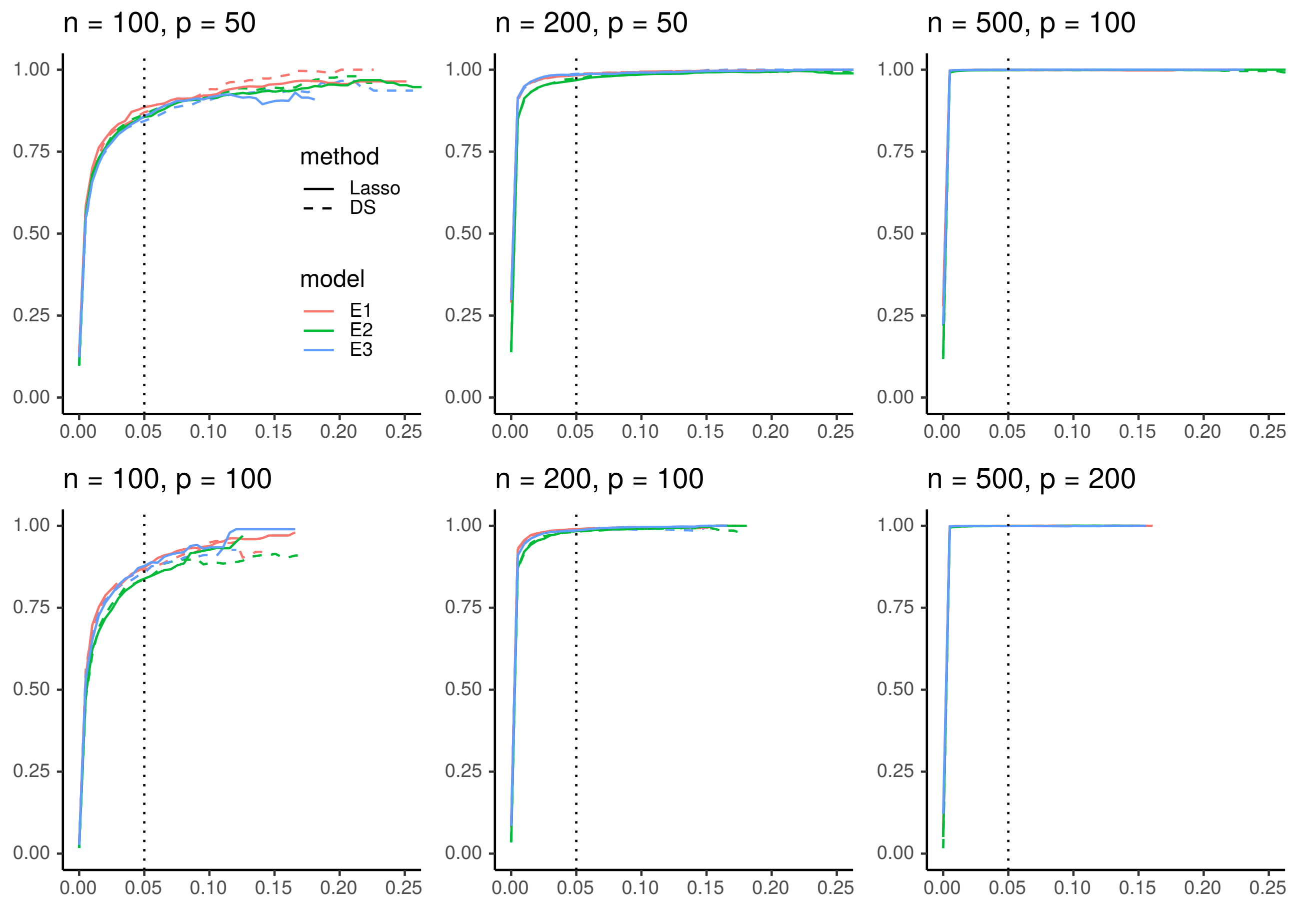
Compared FNETS against FARM (Fan et al., 2021): time-domain PCA + fitting a VAR model to estimated $\boldsymbol{\xi}_t$ via Lasso.

Results: Network estimation

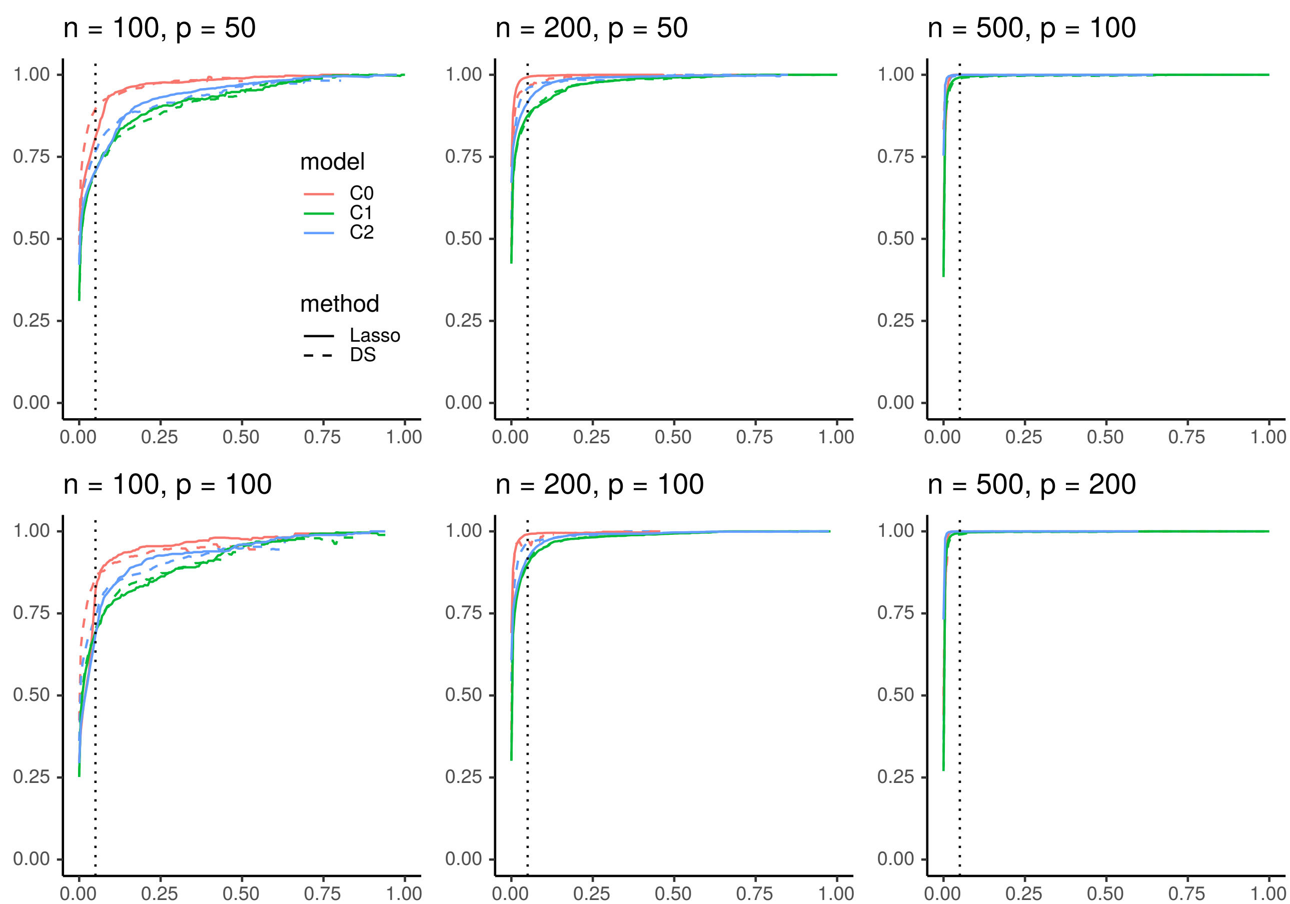
ROC curves of TPR against FPR for estimation of $\mathbf{A}$: (E1) + (C0)–(C2).



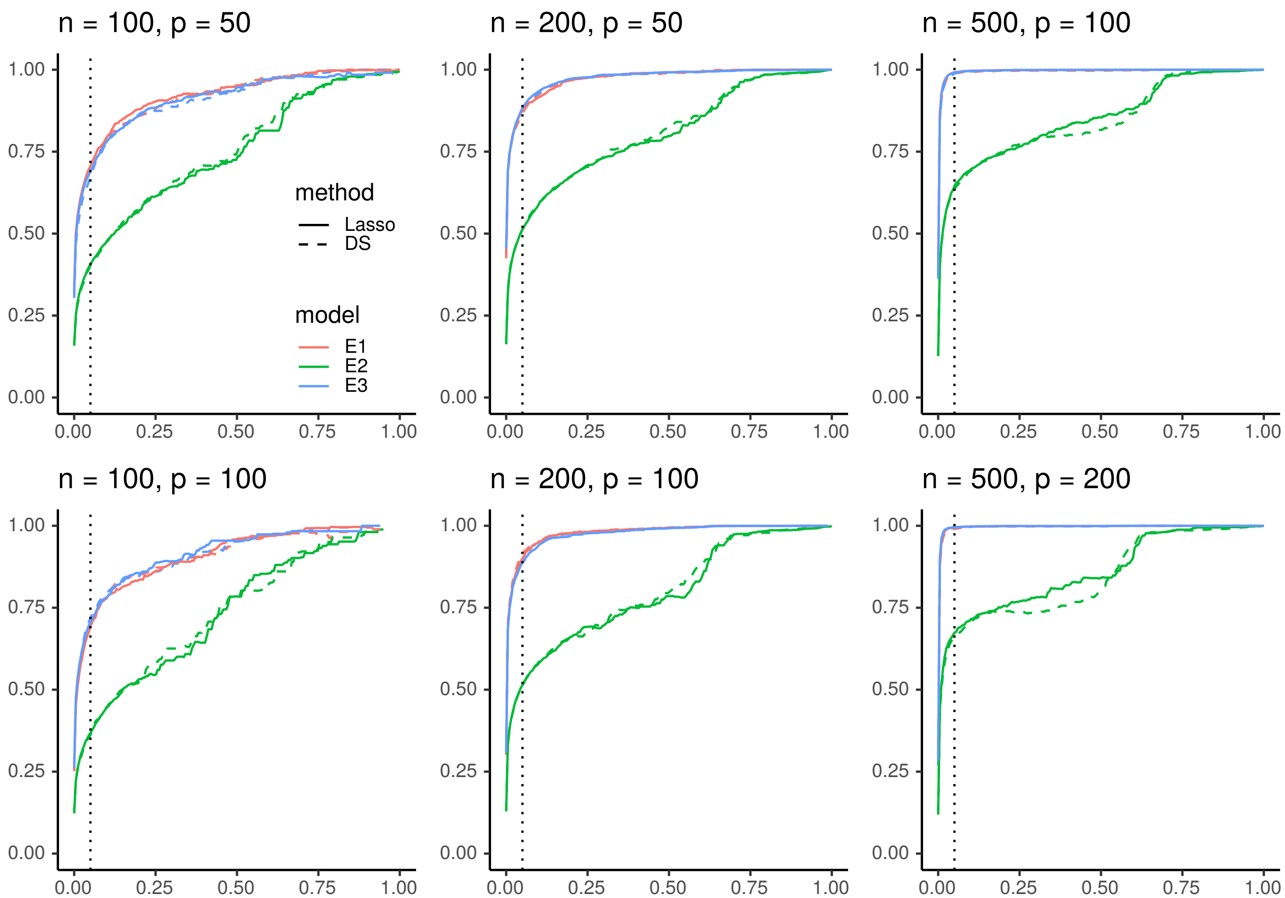
ROC curves of TPR against FPR for estimation of $\mathbf{A}$: (E1)–(E3) + (C1).



ROC curves of TPR against FPR for estimation of $\boldsymbol{\Omega}$: (E1) + (C0)–(C2).

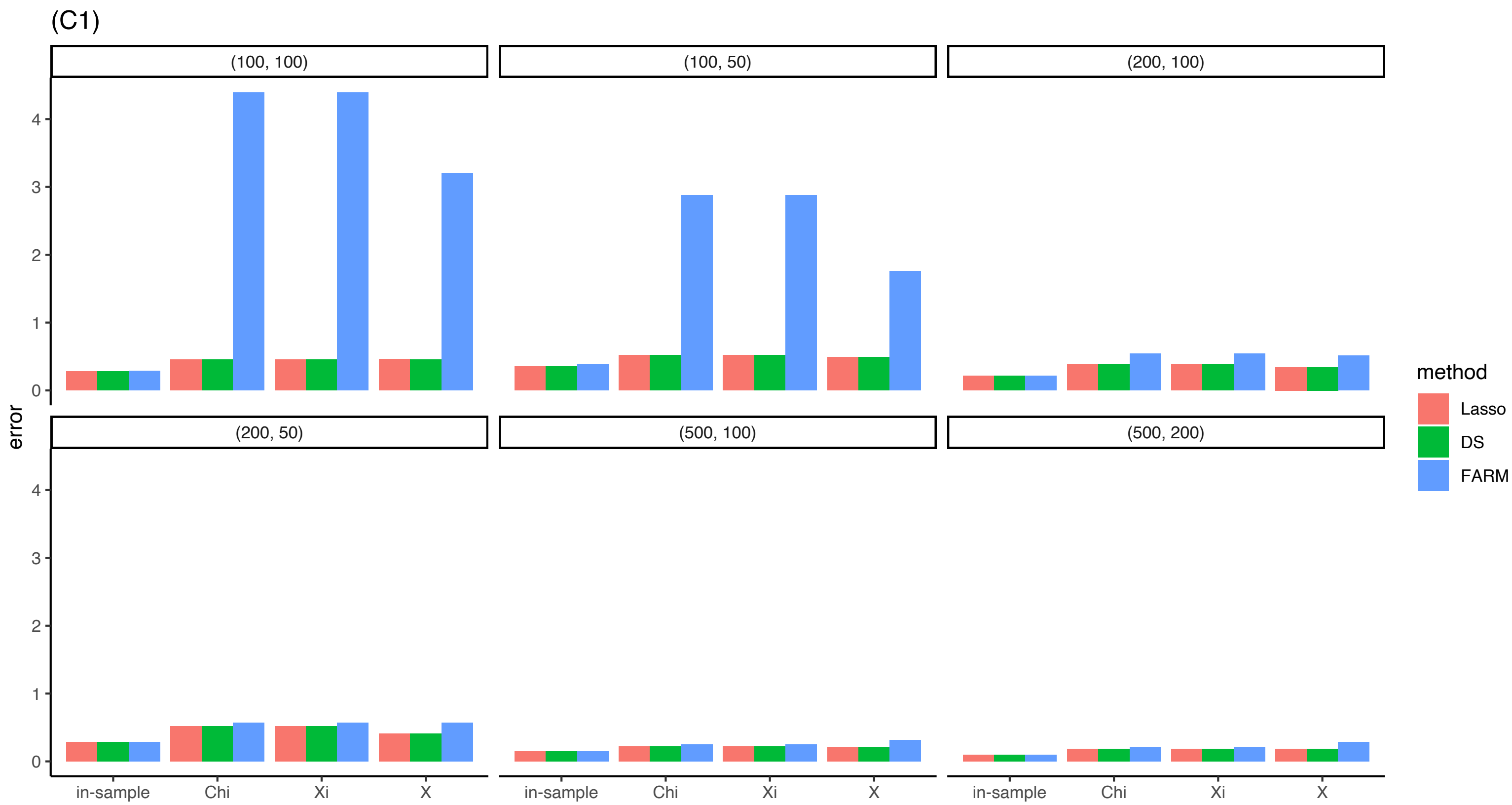


ROC curves of TPR against FPR for estimation of $\boldsymbol{\Omega}$: (E1)–(E3) + (C1).

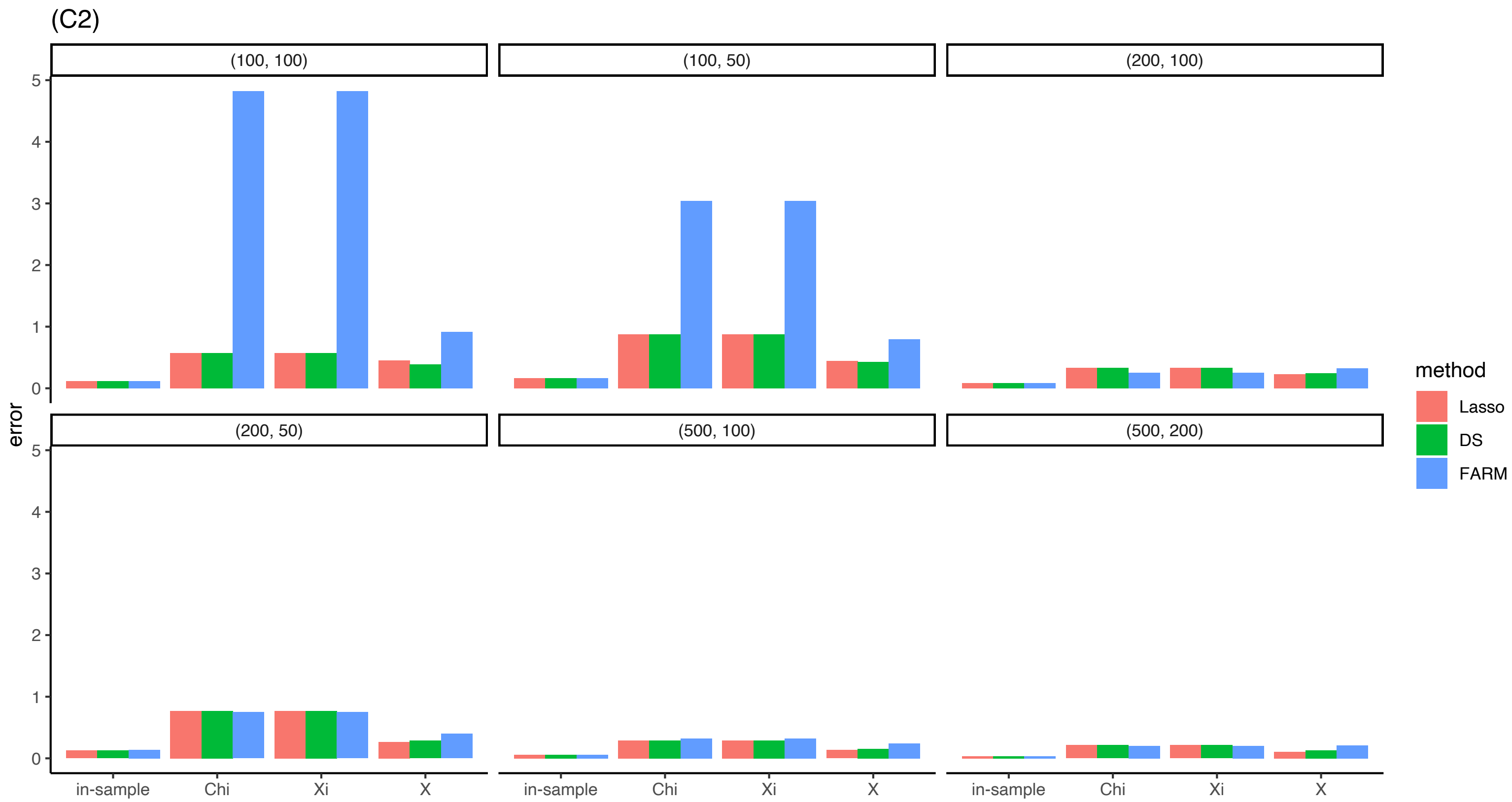


Results: Forecasting

Error measured by $\frac{|\widehat{\gamma}_{n+1|n} - \gamma_{n+1|n}|_2^2}{|\gamma_{n+1|n}|_2^2}$ when (E1) + (C1):



Error measured by $\frac{|\widehat{\gamma}_{n+1|n}-\gamma_{n+1|n}|_2^2}{|\gamma_{n+1|n}|_2^2}$ when (E1) + (C2):



Application to a panel of volatility measures

A panel of volatility measures from $p = 46$ stock prices of US companies all classified as 'financials' according to the Global Industry Classification Standard.

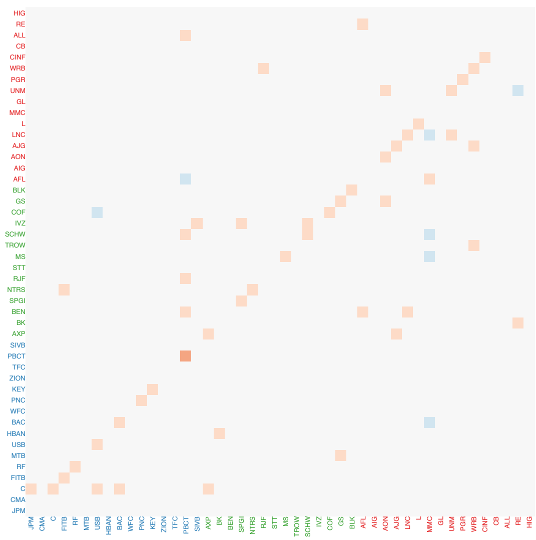
Measure the volatility using the high-low range as

$$\sigma_{it}^2 = 0.361(p_{it}^{\text{high}} - p_{it}^{\text{low}})^2,$$

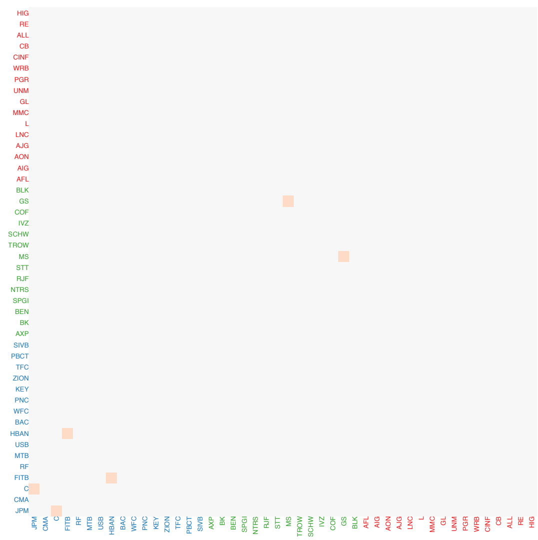
where p_{it}^{high} and p_{it}^{low} the maximum and the minimum log-price of stock i on day t, and set $X_{it} = \log(\sigma_{it}^2)$.

Network analysis: 03/2006–02/2007

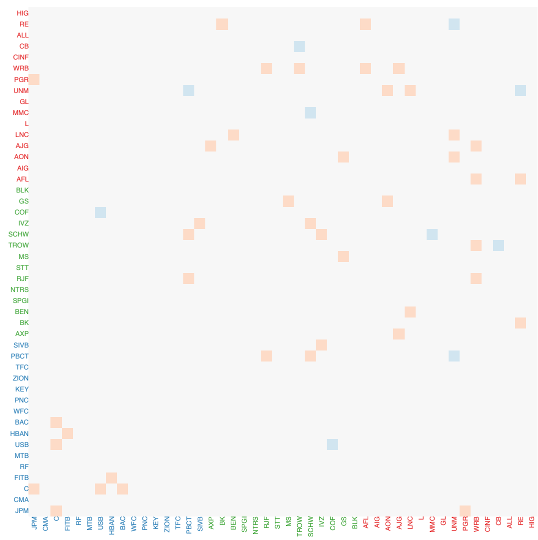
$\mathcal{N}^{\mathrm{G}}$



$\mathcal{N}^{\mathrm{C}}$

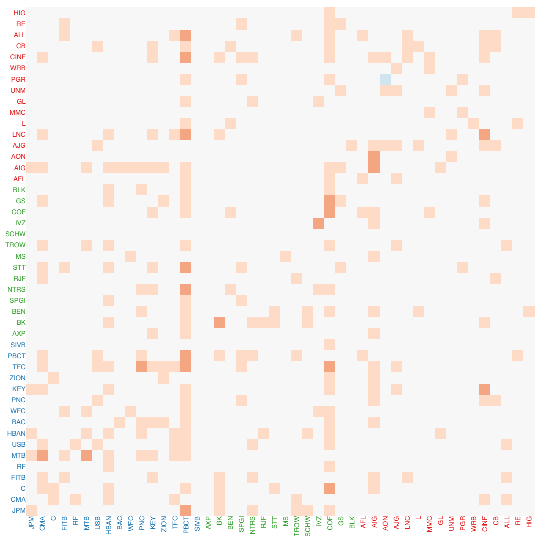


$\mathcal{N}^{\mathrm{L}}$

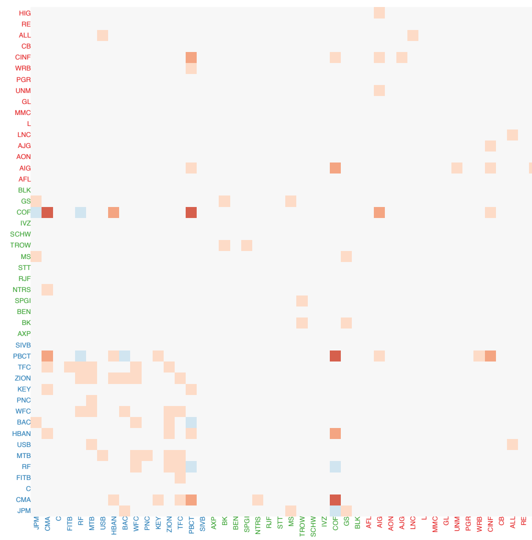


Network analysis: 03/2007–02/2008

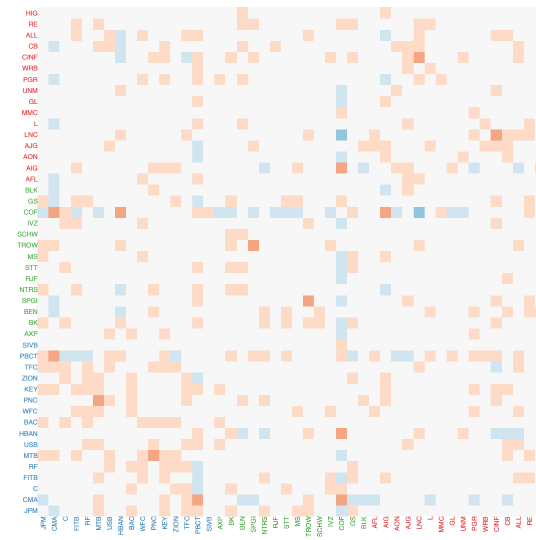
$\mathcal{N}^{\mathrm{G}}$



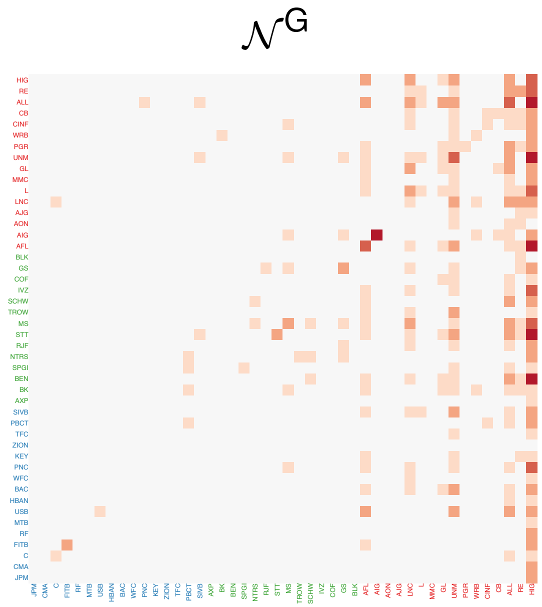
$\mathcal{N}^{\mathrm{C}}$

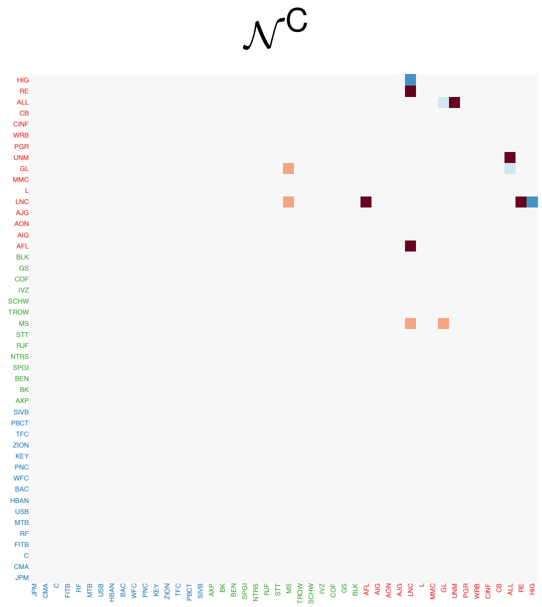


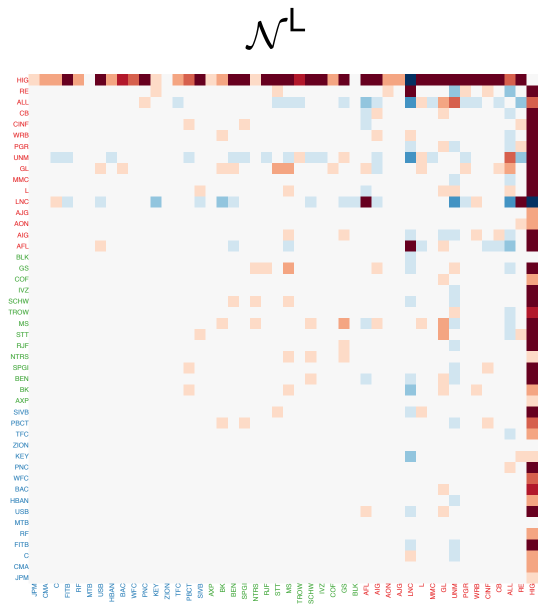
$\mathcal{N}^{\mathrm{L}}$



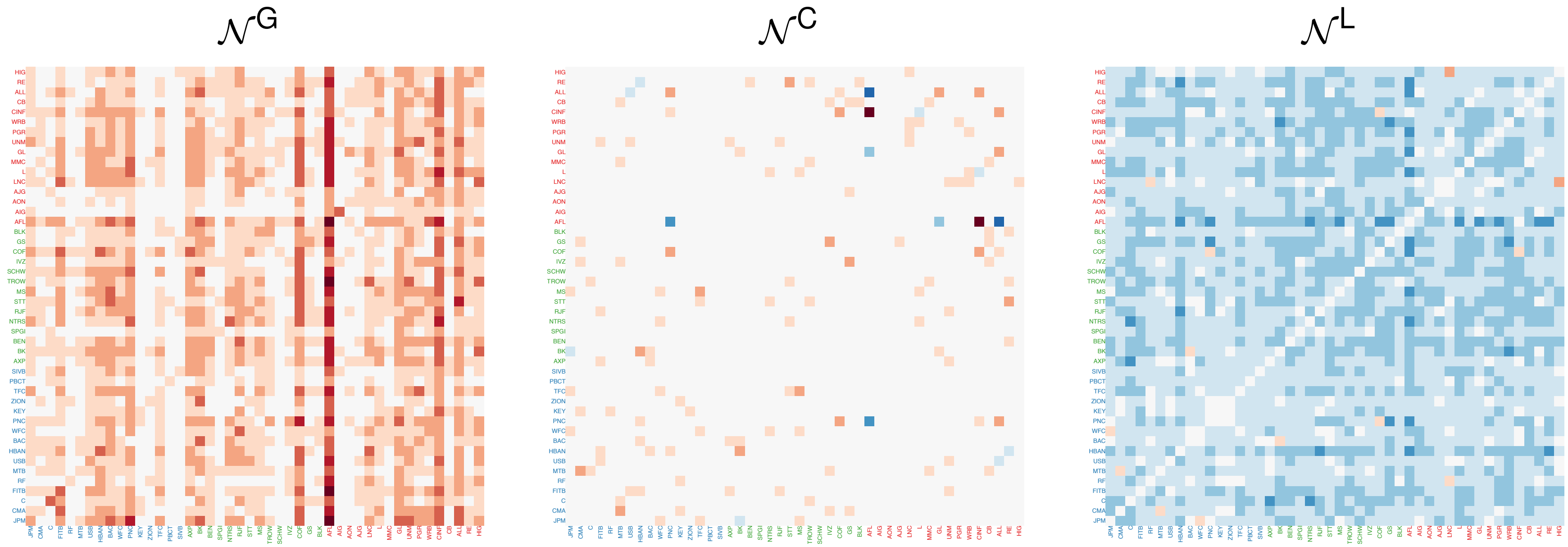
Network analysis: 03/2008–02/2009







Network analysis: 03/2009–02/2010



Forecasting exercise

Rolling window exercise for trading days in 2012 with $n = 252$:

		FNETS					
		Restricted		Unrestricted			
		$\widehat{\boldsymbol{\beta}}^{\text{las}}$	$\widehat{\boldsymbol{\beta}}^{\text{DS}}$	$\widehat{\boldsymbol{\beta}}^{\text{las}}$	$\widehat{\boldsymbol{\beta}}^{\text{DS}}$	AR	FARM
FE$^{\text{avg}}$	Mean	**0.7258**	0.7651	0.7466	0.9665	0.7572	0.7616
	Median	**0.6029**	0.6163	0.6412	0.6756	0.6511	0.6243
	SE	0.4929	0.5081	**0.3748**	1.088	0.4162	0.4946
FE$^{\text{max}}$	Mean	**0.8433**	0.8752	0.8729	0.9359	0.879	0.8745
	Median	**0.7925**	0.8217	0.8088	0.8708	0.8437	0.8259
	SE	0.2331	0.2406	0.2246	0.3246	**0.2169**	0.2337

$$\text{FE}^{\text{avg}}_{T+1} = \frac{\sum_i (X_{i,T+1} - \widehat{X}_{i,T+1|T})^2}{\sum_i X^2_{i,T+1}} \text{ and } \text{FE}^{\text{max}}_{T+1} = \frac{\max_i |X_{i,T+1} - \widehat{X}_{i,T+1|T}|}{\max_i |X_{i,T+1}|}.$$

Conclusions

FNETS consists of estimation and forecasting methods that fully take into account that the VAR process of interest is latent.

Consistency established under general conditions permitting heavy tails and weak factors.

Benefits from regularised Yule-Walker estimation demonstrated both theoretically and empirically.

M. Barigozzi, H. Cho and D. Owens (2022) FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series. *arXiv:2201.06110*

R package `fnets` available on CRAN with accompanying paper
D. Owens, H. Cho and M. Barigozzi (2023) fnets: An R Package for Network Estimation and Forecasting via Factor-Adjusted VAR Modelling. *arXiv:2301.11675*

References

Barigozzi, M. and Hallin, M. (2017). A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66:581–605.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43:1535–1567.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Fan, J., Masini, R., and Medeiros, M. C. (2021). Bridging factor and sparse models. *arXiv preprint arXiv:2102.11341*.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: identification and estimation. *The Review of Economics and Statistics*, 82:540–554.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199:74–92.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. ECB Working Paper 2542, European Central Bank.

Han, F., Lu, H., and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16(97):3115–3150.

Lin, J. and Michailidis, G. (2020). Regularized estimation of high-dimensional factor-augmented vector autoregressive (FAVAR) models. *Journal of Machine Learning Research*, 21:1–51.

Zhang, D. and Wu, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics*, 49(1):233–254.

Tuning parameter selection

For χ_t:

Kernel bandwidth is set at $m = \lfloor 4(n/\log(n))^{1/3} \rfloor$ based on when ν is sufficiently large and $n \asymp p$.

Various factor number estimators exist; we adopt an information criterion-based estimator of Hallin & Liška (2007).

For $\boldsymbol{\xi}_t$:

Cross validation (CV) for jointly selecting λ^{las} or λ^{DS} and VAR order d. Partitioning the data into $\mathcal{I}^{\text{train}} = \{1, \ldots, \lceil \alpha n \rceil\}$ and $\mathcal{I}^{\text{test}} = \{\lceil \alpha n \rceil + 1, \ldots, n\}$, obtain $\widehat{\boldsymbol{\beta}}^{\text{train}}(\lambda, b)$ from $\{\mathbf{X}_t,\, t \in \mathcal{I}^{\text{train}}\}$, evaluate

$$\text{CV}(\lambda, b) = \text{tr}(\widehat{\boldsymbol{\Gamma}}^{\text{test}}_{\xi}(0) - (\widehat{\boldsymbol{\beta}}^{\text{train}}(\lambda, b))^\top \widehat{\mathbb{g}}^{\text{test}}(b) - (\widehat{\mathbb{g}}^{\text{test}}(b))^\top \widehat{\boldsymbol{\beta}}^{\text{train}}(\lambda, b) + (\widehat{\boldsymbol{\beta}}^{\text{train}}(\lambda, b))^\top \widehat{\mathbb{G}}^{\text{test}}(b) \widehat{\boldsymbol{\beta}}^{\text{train}}(\lambda, b)),$$

approximating the prediction error when $\boldsymbol{\xi}_t$ is not directly observed.

For selecting η, we adopt the Burg matrix divergence-based CV measure:

$$\text{CV}(\eta) = \text{tr}\left(\widehat{\boldsymbol{\Delta}}^{\text{train}}(\eta)\widehat{\boldsymbol{\Gamma}}^{\text{test}}\right) - \log\left|\widehat{\boldsymbol{\Delta}}^{\text{train}}(\eta)\widehat{\boldsymbol{\Gamma}}^{\text{test}}\right| - p.$$