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 $X_t = (X_{1t}, \ldots, X_{pt})^\top$ as

$$X_t = A_1 X_{t-1} + \ldots + A_d X_{t-d} + \Gamma^{1/2} \varepsilon_t, \quad \varepsilon_t \sim_{iid} (0, I).$$

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

\[ X_t = A_1 X_{t-1} + \ldots + A_d X_{t-d} + \Gamma^{1/2} \varepsilon_t, \quad \varepsilon_t \sim \text{iid } (0, I). \]

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

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

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

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

Undirected network \( \mathcal{N}^C = (V, \mathcal{E}^C) \) representing contemporaneous dependence in \( \Gamma^{1/2} \varepsilon_t \) by means of partial correlations (PC):

With the precision matrix \( \Delta = [\delta_{ii'}] = \Gamma^{-1} \),

\[ \mathcal{E}^C = \left\{(i, i') \in V \times V : i \neq i' \text{ and } -\frac{\delta_{ii'}}{\sqrt{\delta_{ii} \delta_{i'i'}}} \neq 0 \right\}. \]
Undirected network $\mathcal{N}^L = (\mathcal{V}, \mathcal{E}^L)$ summarising lead-lag and contemporaneous relations in $X_t$ by means of the long-run partial correlations (LRPC): With $\Omega = [\omega_{ii'}, 1 \leq i, i' \leq p] = \Sigma_x^{-1}(0) = 2\pi (A(1))^\top \Delta A(1)$,

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

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

**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 \( \ell_1 \)-regularisation methods have been developed assuming sparsity of \( A_\ell \) (Basu & Michailidis, 2015; Han et al., 2015).

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

\[
\sup_{\omega \in [-\pi, \pi]} \Lambda_{\max}(\Sigma_x(\omega)) < \infty
\]

† uniform boundedness of the largest eigenvalue of the spectral density matrix which implies that \( A_\ell \) 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 $\hat{\Sigma}_x(0)$ 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 $\ell_1$-regularisation methods have been developed assuming sparsity of $A_\ell$ (Basu & Michailidis, 2015; Han et al., 2015).

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

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

† uniform boundedness of the largest eigenvalue of the spectral density matrix which implies that $A_\ell$ 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.
– $\ell_1$-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

$X_t \in \mathbb{R}^p$ is decomposed into two latent components: $X_t = \xi_t + \chi_t$,

\[
\begin{align*}
\xi_t &= \sum_{\ell=1}^{d} A_{\ell} \xi_{t-\ell} + \Gamma^{1/2} \varepsilon_t, \quad \varepsilon_t \sim (0, I_p), \\
\chi_t &= B(L) \begin{pmatrix} \varepsilon_t \end{pmatrix} = \sum_{\ell=0}^{\infty} B_{\ell} u_{t-\ell}, \quad u_t \sim (0, I_q).
\end{align*}
\]

That is, $X_t = \sum_{\ell=0}^{\infty} \left( \sum_{k=1}^{\ell} A_{k} B_{\ell-k} \right) u_{t-\ell} + \sum_{\ell=1}^{d} A_{\ell} X_{t-\ell} + \Gamma^{1/2} \varepsilon_t$.

Adopting generalised dynamic factor model (GDFM, Forni et al., 2000), $\chi_t$ is driven by $q$-dimensional common factors $u_{t-\ell}, \ell \geq 0$.

Cf. static factor model $\chi_t = \sum_{\ell=0}^{M} B_{\ell} u_{t-\ell}$.

Our aim is to (i) estimate networks underpinning the latent VAR process $\xi_t$ under appropriate sparsity assumptions, and (ii) forecast $X_{n+a}$ for some $a \geq 1$ given $X_t, t \leq n$. 
Assumptions for model identifiability

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

On factor-driven $\chi_t$:

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

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

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

† Divergence of $\mu_{\chi,j}(\omega)$ is necessary and sufficient for GDFM representation.
For stability of $\xi_t$ and controlling dependence (Zhang & Wu, 2021):

On the Wold representation, $\xi_t = \sum_{\ell=0}^{\infty} D_\ell \Gamma^{1/2} \varepsilon_{t-\ell}$, we have $|D_{\ell,ik}| \leq C_{ik}(1 + \ell)^{-\varsigma}$ for all $\ell \geq 0$, with

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

for some constants $\Xi > 0$ and $\varsigma > 2$.  † Holds e.g. when $d = 1$ and $|A_1|_{\infty} < 1$.

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

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

$\therefore$ Latent $\xi_t$ and $\chi_t$ are identifiable as by Weyl’s inequality, $\mu_{x,q}(\omega) \to \infty$ while $\mu_{x,q+1}(\omega) \leq B_\xi$ for all $p$.  

11
Network estimation via FNETS

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

**Step 2:** Estimation of \( A_{\ell}, 1 \leq \ell \leq d \), via \( \ell_1 \)-regularised Yule-Walker estimation, from which we estimate \( \mathcal{N}^G = (\mathcal{V}, \mathcal{E}^G) \):

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

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

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

\[
\mathcal{E}^L = \left\{(i, i') \in \mathcal{V} \times \mathcal{V} : i \neq i' \text{ and } -\frac{\omega_{ii'}}{\sqrt{\omega_{ii}\omega_{i'i'}}} \neq 0 \right\}.
\]
Step 1: Factor adjustment

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

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

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

with the Bartlett kernel $K(\cdot)$ and bandwidth $m \asymp n^\beta$, $\beta \in (0, 1)$.

Performing PCA on $\widehat{\Sigma}_x(\omega_k)$ at Fourier frequencies $\omega_k = \frac{2\pi k}{2m+1}$, $|k| \leq m$,

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

we estimate $\Sigma_\chi(\omega_k)$ and $\Gamma_\chi(\ell)$ as

$$\widehat{\Sigma}_\chi(\omega_k) = \sum_{j=1}^{q} \hat{\mu}_{x,j}(\omega_k) \hat{e}_{x,j}(\omega_k)(\hat{e}_{x,j}(\omega_k))^*,$$

$$\widehat{\Gamma}_\chi(\ell) = \frac{2\pi}{2m+1} \sum_{k=-m}^{m} \widehat{\Sigma}_\chi(\omega_k) \exp(i\ell\omega_k).$$

Straightforwardly, $\widehat{\Gamma}_\xi(\ell) = \widehat{\Gamma}_x(\ell) - \widehat{\Gamma}_\chi(\ell)$. 
Step 2: Estimation of VAR parameters

\[
\begin{bmatrix}
\xi_{n} \top \\
\vdots \\
\xi_{d+1} \top
\end{bmatrix}
\in \mathbb{R}^{(n-d) \times p} =
\begin{bmatrix}
\xi_{n-1} \top \\
\vdots \\
\xi_{d} \top
\end{bmatrix}
\begin{bmatrix}
\xi_{n-2} \top \\
\vdots \\
\xi_{d-1} \top
\end{bmatrix} + \begin{bmatrix}
\xi_{n-d} \top \\
\vdots \\
\xi_{1} \top
\end{bmatrix}
\begin{bmatrix}
A_{1} \top \\
\vdots \\
A_{d} \top
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{n} \top \\
\vdots \\
\varepsilon_{d+1} \top
\end{bmatrix} \Gamma^{1/2}.
\]

**Existing approach:** Estimate the latent \( \xi_{t} \), \( 1 \leq t \leq n \), and apply \( \ell_{1} \)-regularisation methods such as Lasso to estimate \( \beta^{0} \) (FARM, Fan et al., 2021).

\( \implies \) Loss of statistical efficiency due to estimating the \( n \times p \) matrix containing \( \xi_{t} \), \( 1 \leq t \leq n \).
**$\ell_1$-regularised Yule-Walker estimators**

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

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

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

**Lasso ($\ell_1$-penalised minimisation):**

$$\hat{\beta}_{\text{las}} = \arg \min_{\beta \in \mathbb{R}^{pd \times p}} \text{tr} \left( \beta^T \hat{G} \beta - 2 \beta^T \hat{g} \right) + \lambda_{\text{las}} |\beta|_1.$$  

**Dantzig selector (constrained $\ell_1$-minimisation):**

$$\hat{\beta}_{\text{DS}} = \arg \min_{\beta \in \mathbb{R}^{pd \times p}} |\beta|_1 \quad \text{subject to} \quad \left| \hat{G} \beta - \hat{g} \right|_\infty \leq \lambda_{\text{DS}}.$$
Step 3: Estimation of (long-run) partial correlations

From Yule-Walker equation, estimate $\Gamma = \text{Cov}(\Gamma^{1/2}\varepsilon_t)$ by

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

We estimate $\Delta = \Gamma^{-1}$ via constrained $\ell_1$-minimisation (Cai et al., 2011),

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

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

$$\hat{\Omega} = 2\pi (\hat{A}(1))^\top \hat{\Delta} \hat{A}(1).$$
Theoretical consistency

\[
X_t = \sum_{m=0}^{\infty} B_m u_{t-m} + \sum_{\ell=1}^{d} A_\ell \xi_{t-\ell} + \Gamma^{1/2} \varepsilon_t.
\]

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

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

To ease presentation, let all the factors be strong (i.e. \( \mu_{X,q}(\omega) \asymp p \)).
Consistency of Step 1: Factor adjustment

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

$$\mathcal{E}_{n,p} = \left\{ \max_{\ell \leq d} |\hat{\Gamma}_\xi(\ell) - \Gamma_\xi(\ell)|_\infty \lesssim \frac{1}{m} \lor \vartheta_{n,p} \lor \frac{1}{\sqrt{p}} \right\},$$

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

In what follows, all results are conditional on $\mathcal{E}_{n,p}$. 
Consistency of Step 2

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

$$s_{in} = \max_{1 \leq j \leq p} s_{0,j} \quad \text{where} \quad s_{0,j} = \sum_{\ell=1}^{d} |A_{\ell,j}|_0 = |\beta_{.j}|_0.$$  

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

$$\left| \hat{\beta}_{.j}^{\text{las}} - \beta_{.j}^0 \right|_\infty \leq \left| \hat{\beta}_{.j}^{\text{las}} - \beta_{.j}^0 \right|_2 \leq \frac{32 \sqrt{s_{in} \lambda^{\text{las}}}}{\pi m_\xi} = t.$$

Under ‘beta-min’ condition: $\min_{(i,j): \beta_{ij}^0 \neq 0} |\beta_{ij}^0| \geq 2t$, $\mathcal{N}^G$ is consistently estimated by hard-thresholding $\hat{\beta}^{\text{las}}$ as

$$\hat{\beta}^{\text{las}}(t) = \left[ \hat{\beta}_{ij}^{\text{las}} \cdot I\{|\hat{\beta}_{ij}^{\text{las}}| > t\}, 1 \leq i \leq pd, 1 \leq j \leq p \right].$$
Consistency of Step 3

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

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

Additionally, if $\hat{\Omega}$ is obtained with $\tilde{\beta}_{\text{las}}(t)$,

$$|\hat{\Omega} - \Omega|_\infty \lesssim \|A(1)\|_1 (\|\Delta\|_{s_{\text{out}}t} + \|A(1)\|_1 \|\Delta\|_1\eta),$$

where $s_{\text{out}} = \max_{1 \leq j \leq p} \sum_{\ell=1}^d |A_{\ell,j}|_0$.

Edge sets of $\mathcal{N}^C$ and $\mathcal{N}^L$ are estimated by hard-thresholding $\hat{\Delta}$ and $\hat{\Omega}$, respectively, and consistency is achieved under analogous ‘beta-min’ conditions.
Consider the case where $\chi_t = \sum_{m=0}^{M} B_m u_{t-m}, \ n \geq p, \ \nu > 8$. Then,

$$\max_j |\hat{\beta}_{j}^{\text{las}} - \beta_j^0|_2 = O_P(\sqrt{s_{\text{in}} \log(p)/n})$$

$\hat{\beta}^{\text{las}}$ 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 $\xi_t$ attains $O_P(\sqrt{s_{\text{in}} n^{-1/2+5/\nu}})$ under strong mixingness (FARM, Fan et al., 2021).

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

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

∴ FNETS estimators perform as well as benchmarks obtained where $\xi_t$ is directly observed under independence.
Forecasting via FNETS

Forecast $X_{n+a}$ given $X_t$, $t \leq n$ for some $a \geq 1$, by

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

$$X_{n+a|n} = \underbrace{\text{Proj}(\chi_{n+a|\chi_v, v \leq t})}_{\chi_{n+a|n}} + \underbrace{\text{Proj}(\xi_{n+a|\xi_v, v \leq t})}_{\xi_{n+a|n}}$$
Forecasting of factor-driven component

Under a more restricted, static factor model \( \chi_t = \sum_{\ell=0}^{M} B_{\ell} u_{t-\ell} \) with the number of static factors \( r = (M + 1)q \),

\[
\chi_{n+a|n} = \Gamma \chi(-a) \Gamma^{-}(0) \chi_n.
\]

This motivates \( \hat{\chi}_{n+a|n}^{\text{res}} = \hat{\Gamma} \chi(-a) \hat{\Gamma}^{-}(0) X_n \), which achieves consistent estimation of \( \chi_{n+a|n} \):

\[
\left| \hat{\chi}_{n+a|n}^{\text{res}} - \chi_{n+a|n} \right|_{\infty} = O_p \left( \vartheta_{n,p} \lor \frac{1}{m} \lor \frac{1}{\sqrt{p}} \right).
\]

We also obtain the in-sample estimator of \( \chi_t, t \leq n \), as

\[
\hat{\chi}_t^{\text{res}} = \hat{E}_\chi \hat{E}_\chi^\top X_t, \text{ where } \hat{E}_\chi \text{ contains } r \text{ leading eigenvectors of } \hat{\Gamma}_\chi(0).
\]
Forecasting of the latent VAR process

Under the VAR model,
\[ \hat{\xi}_{n+a|n} = \sum_{\ell=1}^{\max(1,a)-1} \hat{A}_\ell \hat{\xi}_{n+a-\ell|n} + \sum_{\ell=\max(1,a)}^{d} \hat{A}_\ell \hat{\xi}_{n+a-\ell}, \]

which inherits the theoretical properties of \( \hat{A}_\ell \) and in-sample estimator \( \hat{\xi}_t = X_t - \hat{\chi}_t, \ t \leq n, \) as

\[ \left| \hat{\xi}_{n+1|n} - \xi_{n+1|n} \right|_{\infty} = O_P \left( (s_{\in} \log^{1/2}(p)p^{1/\nu} + \| \beta^0 \|_1) \left( \vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right). \]
Simulations

**VAR process:** $\xi_t = A\xi_{t-1} + \Gamma^{1/2}\varepsilon_t$ with $\text{supp}(A)$ generated as an Erdős-Rényi random graph, and

(E1) $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ and $\Gamma = I$.
(E2) $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ and $\Gamma \neq I$.
(E3) $\sqrt{5/3} \cdot \varepsilon_{it} \sim t_5$ and $\Gamma = I$.

**Factor-driven component:**
(C0) $\chi_t = 0$ (‘oracle’).
(C1) $\chi_t$ does not admit a static representation with $q = 2$.
(C2) $\chi_t$ 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 $\xi_t$ via Lasso.
Results: Network estimation

ROC curves of TPR against FPR for estimation of $A: (E1) + (C0)-(C2)$. 

$\begin{align*}
\text{n} = 100, \ p = 50 \\
\text{n} = 200, \ p = 50 \\
\text{n} = 500, \ p = 100 \\
\text{n} = 500, \ p = 200
\end{align*}$
ROC curves of TPR against FPR for estimation of $A$: (E1)–(E3) + (C1).
ROC curves of TPR against FPR for estimation of $\Omega$: (E1) + (C0)−(C2).
ROC curves of TPR against FPR for estimation of $\Omega$: (E1)-(E3) + (C1).
Results: Forecasting

Error measured by

\[
\frac{|\hat{\gamma}_{n+1} - \gamma_{n+1}|^2}{|\gamma_{n+1}|^2}
\]

when (E1) + (C1):

\[(200, 50) \quad (500, 100) \quad (500, 200) \quad (100, 100) \quad (100, 50) \quad (200, 100)\]
Error measured by \( \frac{\|\hat{\gamma}_{n+1} - \gamma_{n+1}\|^2_n}{\|\gamma_{n+1}\|^2_n} \) 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 $p_{it}^{\text{high}}$ and $p_{it}^{\text{low}}$ 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
Network analysis: 03/2007–02/2008
Network analysis: 03/2008–02/2009
Network analysis: 03/2009–02/2010
Forecasting exercise

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

<table>
<thead>
<tr>
<th>FNETS</th>
<th>Restricted</th>
<th>Unrestricted</th>
<th>AR</th>
<th>FARM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{\text{las}}$</td>
<td>$\hat{\beta}^{\text{DS}}$</td>
<td>$\hat{\beta}^{\text{las}}$</td>
<td>$\hat{\beta}^{\text{DS}}$</td>
</tr>
<tr>
<td>$\text{FE}^{\text{avg}}$</td>
<td>Mean</td>
<td>0.7258</td>
<td>0.7651</td>
<td>0.7466</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.6029</td>
<td>0.6163</td>
<td>0.6412</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.4929</td>
<td>0.5081</td>
<td>0.3748</td>
</tr>
<tr>
<td>$\text{FE}^{\text{max}}$</td>
<td>Mean</td>
<td>0.8433</td>
<td>0.8752</td>
<td>0.8729</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.7925</td>
<td>0.8217</td>
<td>0.8088</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.2331</td>
<td>0.2406</td>
<td>0.2246</td>
</tr>
</tbody>
</table>

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


R package fnets available on CRAN with accompanying paper


Tuning parameter selection

For $\chi_t$:

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

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

Cross validation (CV) for jointly selecting $\lambda^{\text{las}}$ or $\lambda^{\text{DS}}$ 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 $\hat{\beta}^{\text{train}}(\lambda, b)$ from $\{X_t, t \in \mathcal{I}^{\text{train}}\}$, evaluate

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

approximating the prediction error when $\xi_t$ is not directly observed.

For selecting $\eta$, we adopt the Burg matrix divergence-based CV measure:

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