FNETS: Factor-Adjusted Network Estimation and Forecasting for High-Dimensional Time Series

Matteo Barigozzi, Haeran Cho & Dom Owens

View supplementary material

Published online: 04 Oct 2023.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

FNETS: Factor-Adjusted Network Estimation and Forecasting for High-Dimensional Time Series

Matteo Barigozzi[a], Haeran Cho[b] ⬤, and Dom Owens[b]

[a]Department of Economics, Università di Bologna, Bologna, Italy; [b]School of Mathematics, University of Bristol, Bristol, UK

ABSTRACT

We propose FNETS, a methodology for network estimation and forecasting of high-dimensional time series exhibiting strong serial- and cross-sectional correlations. We operate under a factor-adjusted vector autoregressive (VAR) model which, after accounting for pervasive co-movements of the variables by *common* factors, models the remaining *idiosyncratic* dynamic dependence between the variables as a sparse VAR process. Network estimation of FNETS consists of three steps: (i) factor-adjustment via dynamic principal component analysis, (ii) estimation of the latent VAR process via ℓ_1-regularized Yule-Walker estimator, and (iii) estimation of partial correlation and long-run partial correlation matrices. In doing so, we learn three networks underpinning the VAR process, namely a directed network representing the Granger causal linkages between the variables, an undirected one embedding their contemporaneous relationships and finally, an undirected network that summarizes both lead-lag and contemporaneous linkages. In addition, FNETS provides a suite of methods for forecasting the factor-driven and the idiosyncratic VAR processes. Under general conditions permitting tails heavier than the Gaussian one, we derive uniform consistency rates for the estimators in both network estimation and forecasting, which hold as the dimension of the panel and the sample size diverge. Simulation studies and real data application confirm the good performance of FNETS.

1. Introduction

Vector autoregressive (VAR) models are popularly adopted for time series analysis in economics and finance. Fitting a VAR model to the data enables inferring dynamic interdependence between the variables as well as forecasting the future. VAR models are particularly appealing for network analysis since estimating the nonzero elements of the VAR parameter matrices, a.k.a. transition matrices, recovers directed edges between the components of vector time series in a Granger causality network. In addition, by estimating a precision matrix (inverse of the covariance matrix) of the VAR innovations, we can also define a network capturing contemporaneous linear dependencies. For the network interpretation of VAR modeling, see for example, Dahlhaus (2000), Eichler (2007), Billio et al. (2012), Ahelegbey, Billio, and Casarin (2016), Barigozzi and Brownlees (2019), Guðmundsson and Brownlees (2021), and Uematsu and Yamagata (2023).

Estimation of VAR models quickly becomes a high-dimensional problem as the number of parameters grows quadratically with the dimensionality. There is a mature literature on estimation of high-dimensional VAR models under the sparsity (Hsu, Hung, and Chang 2008; Basu and Michailidis 2015; Han, Lu, and Liu 2015; Kock and Callot 2015; Nicholson et al. 2020; Krampe and Paparoditis 2021; Masini, Medeiros, and Mendes 2022; Adamek, Smeekes, and Wilms 2023) and low-rank plus sparsity (Basu, Li,

and Michailidis 2019) assumptions, see also BańburaBańbura, Giannone, and Reichlin (2010) for a Bayesian approach. In all above, either explicitly or implicitly, the spectral density of the time series is required to have eigenvalues which are uniformly bounded over frequencies. Indeed, this condition is crucial for controlling the deviation bounds involved in theoretical investigation of regularized estimators.

Lin and Michailidis (2020) observe that for VAR processes, this assumption restricts the parameters to be either dense but small in their magnitude (which makes their estimation using the shrinkage-based methods challenging) or highly sparse, while Giannone, Lenza, and Primiceri (2021) note the difficulty of identifying sparse predictive representations in many economic applications. Moreover, some datasets typically exhibit strong serial and cross-sectional correlations and violate the bounded spectrum assumption. The left panel of Figure 1 provides an illustration of this phenomenon; with the increase of dimensionality, a volatility panel dataset (see Section 5.3 for its description) exhibits a linear increase in the leading eigenvalue of the estimate of its spectral density matrix at frequency 0 (i.e., long-run covariance). The right panel visualizes the outcome from fitting a VAR(5) model to the same dataset without making any adjustment of the strong correlations (see the caption for further details), from which we cannot infer meaningful, sparse pairwise relationship.
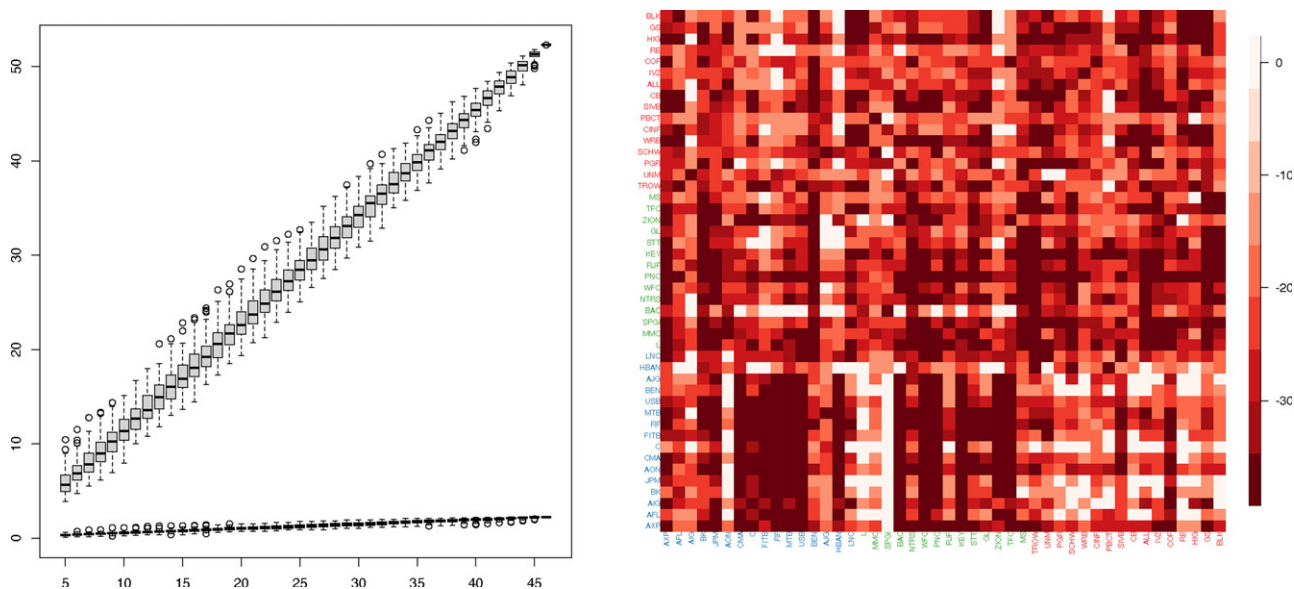

Figure 1. Left: The two largest eigenvalues (*y*-axis) of the long-run covariance matrix estimated from the volatility panel analyzed in Section 5.3 (March 2008 to March 2009, $n = 252$) with subsets of cross-sections randomly sampled 100 times for each given dimension $p \in \{5, \ldots, 46\}$ (*x*-axis). Right: logged and truncated *p*-values (truncation level chosen by Bonferroni correction with the significance level 0.1) from fitting a VAR(5) model to the same dataset using ridge regression and generating *p*-values corresponding to each coefficient as described in Cule, Vineis, and De Iorio (2011). For each pair of variables (corresponding tickers given in *x*- and *y*-axes), the minimum *p*-value over the five lags is reported.

In this article, we propose to model high-dimensional time series by means of a factor-adjusted VAR approach, which simultaneously accounts for strong serial and cross-sectional correlations attributed to factors, as well as sparse, idiosyncratic correlations among the variables that remain after factor adjustment. We take the most general approach to factor modeling based on the generalized dynamic factor model, where factors are dynamic in the sense that they are allowed to have not only contemporaneous but also lagged effects on the variables (Forni et al. 2000). We propose FNETS, a suite of tools accompanying the model for estimation and forecasting with a particular focus on network analysis, which addresses the challenges arising from the latency of the VAR process as well as high dimensionality.

We make the following methodological and theoretical contributions.

(a) We propose an ℓ_1-regularized Yule-Walker estimation method for estimating the factor-adjusted, idiosyncratic VAR, while permitting the number of nonzero parameters to slowly grow with the dimensionality. Estimating the VAR parameters and the inverse of the innovation covariance, and then combining them allow us to define three networks underlying the latent VAR process, namely a direct network representing Granger causal linkages, an undirected one underpinning their contemporaneous relationships, as well as an undirected network summarizing both. Under general conditions permitting weak factors and heavier tails than the sub-Gaussian one, we show the consistency of FNETS in estimating the edge sets of these networks, which holds uniformly over all p^2 entries of the networks (Propositions 3.3 and 3.5).

(b) We provide new consistency rates for the estimation and forecasting approaches considered by Forni et al. (2005, 2017), which hold uniformly for the entire cross-sections

of *p*-dimensional time series (Propositions 4.1 and B.2). In doing so, we establish uniform consistency of the estimators of high-dimensional spectral density matrices of the factor-driven and the idiosyncratic components, extending the results of Zhang and Wu (2021) to the presence of latent factors.

Our approach differs from the existing ones for factor-adjusted regression problems (Fan, Ke, and Wang 2020; Fan, Masini, and Medeiros 2021; Fan, Lou, and Yu 2023; Krampe and Margaritella 2021), as (i) it allows for the presence of dynamic factors, thus including all possible dynamic linear co-dependencies, and (ii) it relies only on the estimators of the autocovariances of the latent idiosyncratic process, and avoids estimating the entire latent process and controlling the errors arising from such a step, which increase with the sample size. The price to pay for the generality of the factor modeling in (i), is an extra term appearing in the rate of consistency which represents the bandwidth for spectral density estimation required for factor-adjustment in the frequency domain. We make explicit the role played by this bandwidth in the theoretical results, and also present the results under a more restricted static factor model for ease of comparison. We mention two more differences between this article and Fan, Masini, and Medeiros (2021) and Fan, Lou, and Yu (2023). First, they additionally consider the problem of testing hypotheses on the idiosyncratic covariance and the adequacy of factor/sparse regression, while we focus on network estimation. Second, their methods accommodate models for the idiosyncratic component other than VAR.

FNETS is another take the popular low-rank plus sparsity modeling framework in the high-dimensional learning literature, Also, it is in line with a frequently adopted practice in financial time series analysis where factor-driven common components representing the systematic sources of risk, are removed

prior to inferring a network structure via (sparse) regression modeling and identifying the most central nodes representing the systemic sources of risk (Diebold and Yılmaz 2014; Barigozzi and Brownlees 2019). We provide a rigorous theoretical treatment of this empirical approach by accounting for the effect of the factor-adjustment step on the second step regression.

The rest of the article is organized as follows. Section 2 introduces the factor-adjusted VAR model. Sections 3 and 4 describe the network estimation and forecasting methodologies comprising FNETS, respectively, and provide their theoretical consistency. In Section 5, we demonstrate the good estimation and forecasting performance of FNETS on a panel of volatility measures. Section 6 concludes the article, and all the proofs and complete simulation results are presented in Supplementary Appendix. The R software fnets implementing FNETS is available from CRAN (Barigozzi, Cho, and Owens 2023).

Notations. By $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{0}$, we denote an identity matrix, a matrix of zeros, and a vector of zeros whose dimensions depend on the context. For a matrix $\mathbf{A} = [a_{ii'}, 1 \leq i \leq m, 1 \leq i' \leq n]$, we denote by $\mathbf{A}^\top$ its transpose. The element-wise $\ell_\infty, \ell_0, \ell_1$ and ℓ_2-norms are denoted by $|\mathbf{A}|_\infty = \max_{1 \leq i \leq m} \max_{1 \leq i' \leq n} |a_{ii'}|$, $|\mathbf{A}|_0 = \sum_{i=1}^m \sum_{i'=1}^n \mathbb{I}_{\{a_{ii'} \neq 0\}}$, $|\mathbf{A}|_1 = \sum_{i=1}^m \sum_{i'=1}^n |a_{ii'}|$ and $|\mathbf{A}|_2 = \sqrt{\sum_{i=1}^m \sum_{i'=1}^n |a_{ii'}|^2}$. The Frobenius, spectral, induced L_1 and L_∞-norms are denoted by $\|\mathbf{A}\|_F = |\mathbf{A}|_2$, $\|\mathbf{A}\| = \sqrt{\Lambda_{\max}(\mathbf{A}^\top \mathbf{A})}$ (with $\Lambda_{\max}(\mathbf{A})$ and $\Lambda_{\min}(\mathbf{A})$ denoting its largest and smallest eigenvalues in modulus), $\|\mathbf{A}\|_1 = \max_{1 \leq i' \leq n} \sum_{i=1}^m |a_{ii'}|$ and $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq n} \sum_{i'=1}^m |a_{ii'}|$. Let $\mathbf{A}_{i\cdot}$ and $\mathbf{A}_{\cdot k}$ denote the ith row and the kth column of $\mathbf{A}$. For two real numbers, set $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. Given two sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if, for some finite constant $C > 0$ there exists $N \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ such that $|a_n||b_n|^{-1} \leq C$ for all $n \geq N$; we denote by O_P the stochastic boundedness. We write $a_n \asymp b_n$ when $a_n = O(b_n)$ and $b_n = O(a_n)$. Throughout, L denotes the lag operator and $\iota = \sqrt{-1}$. Finally, $\mathbb{I}_{\mathcal{A}} = 1$ if the event $\mathcal{A}$ takes place and 0 otherwise.

2. Factor-Adjusted Vector Autoregressive Model

Consider a zero-mean, second-order stationary p-variate process $\mathbf{X}_t = (X_{1t}, \ldots, X_{pt})^\top$, $1 \leq t \leq n$, which is decomposed into the sum of two latent components: a factor-driven, *common* component $\boldsymbol{\chi}_t = (\chi_{1t}, \ldots, \chi_{pt})^\top$, and an *idiosyncratic* component $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{pt})^\top$ modeled as a VAR process. That is, $\mathbf{X}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$ where

$$\boldsymbol{\chi}_t = \mathcal{B}(L)\mathbf{u}_t = \sum_{\ell=0}^\infty \mathbf{B}_\ell \mathbf{u}_{t-\ell} \quad \text{with } \mathbf{u}_t = (u_{1t}, \ldots, u_{qt})^\top, \quad \text{and} \tag{1}$$

$$\mathcal{A}(L)\boldsymbol{\xi}_t = \boldsymbol{\xi}_t - \sum_{\ell=1}^d \mathbf{A}_\ell \boldsymbol{\xi}_{t-\ell} = \mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t \quad \text{with } \boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{pt})^\top. \tag{2}$$

In (1), the latent random vector $\mathbf{u}_t$, referred to as the vector of *common factors* or *common shocks*, is assumed to satisfy $\mathsf{E}(\mathbf{u}_t) = \mathbf{0}$ and $\mathsf{cov}(\mathbf{u}_t) = \mathbf{I}_q$, and are loaded on each χ_{it} via square summable, one-sided filters $\mathcal{B}_{ij}(L) = \sum_{\ell=0}^\infty B_{\ell,ij}L^\ell$, where $\mathbf{B}_\ell =$

$[B_{\ell,ij}, 1 \leq i \leq p, 1 \leq j \leq q] \in \mathbb{R}^{p \times q}$. This defines the generalized dynamic factor model (GDFM) proposed by Forni et al. (2000) and Forni and Lippi (2001), which provides the most general approach to high-dimensional time series factor modeling.

In (2), the idiosyncratic component $\boldsymbol{\xi}_t$ is modeled as a VAR(d) process for some finite positive integer d, with innovations $\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$ where $\mathbf{\Gamma} \in \mathbb{R}^{p \times p}$ is some positive definite matrix and $\mathbf{\Gamma}^{1/2}$ its symmetric square root matrix, and $\mathsf{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $\mathsf{cov}(\boldsymbol{\varepsilon}_t) = \mathbf{I}_p$. We assume that $\boldsymbol{\xi}_t$ is causal (see Assumption 2.3(i)), that is, it admits the Wold representation:

$$\boldsymbol{\xi}_t = \mathcal{D}(L)\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t = \sum_{\ell=0}^\infty \mathbf{D}_\ell \mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_{t-\ell} \quad \text{with } \mathcal{D}(L) = \mathcal{A}^{-1}(L), \tag{3}$$

such that $\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$ is seen as a vector of *idiosyncratic shocks* loaded on each ξ_{it} via square summable, one-sided filters $\mathcal{D}_{ik}(L) = \sum_{\ell=0}^\infty D_{\ell,ik}L^\ell$ where $\mathbf{D}_\ell = [D_{\ell,ik}, 1 \leq i, k \leq p]$. After accounting for the dominant cross-sectional dependence in the data (both contemporaneous and lagged) by factors, it is reasonable to assume that the dependences left in $\boldsymbol{\xi}_t$ are weak and, therefore, that the VAR structure is sufficiently sparse. Discussion on the precise requirement on the sparsity of $\mathbf{A}_\ell$, $1 \leq \ell \leq d$, and $\mathbf{\Gamma}^{-1}$ is deferred to Section 3.

Remark 2.1. A special case of the GDFM is the popularly adopted static factor model where the factors are loaded only contemporaneously (see e.g., Stock and Watson 2002; Bai 2003; Fan, Liao, and Mincheva 2013). This is formalized in Assumption 4.1, where we consider forecasting under a static representation. A sufficient condition to obtain a static representation from the GDFM in (1), is to assume $\mathcal{B}(L) = \sum_{\ell=0}^s \mathbf{B}_\ell L^\ell$ for some finite integer $s \geq 0$. For example, if $s = 0$, the model reduces to $\boldsymbol{\chi}_t = \mathbf{B}_0 \mathbf{u}_t$ while if $s > 0$, it can be written as $\boldsymbol{\chi}_t = \mathbf{\Lambda}\mathbf{F}_t$ with $\mathbf{\Lambda} = [\mathbf{B}_\ell, 0 \leq \ell \leq s]$ and $\mathbf{F}_t = (\mathbf{u}_t^\top, \ldots, \mathbf{u}_{t-s}^\top)^\top$. Under the static factor model, $\mathbf{X}_t$ admits a factor-augmented VAR representation (see Remark 4.1).

In the remainder of this section, we list the assumptions required for identification and estimation of (1)–(2). Since $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ are latent, some assumptions are required to ensure their (asymptotic) identifiability which are made in the frequency domain. Denote by $\mathbf{\Sigma}_x(\omega)$ the spectral density matrix of $\mathbf{X}_t$ at frequency $\omega \in [-\pi, \pi]$, and $\mu_{x,j}(\omega)$ its dynamic eigenvalues which are real-valued and ordered in the decreasing order. We similarly define $\mathbf{\Sigma}_\chi(\omega)$, $\mu_{\chi,j}(\omega)$, $\mathbf{\Sigma}_\xi(\omega)$ and $\mu_{\xi,j}(\omega)$.

Assumption 2.1. There exist a positive integer $p_0 \geq 1$, constants $\rho_j \in (3/4, 1]$ with $\rho_1 \geq \cdots \geq \rho_q$, and pairs of continuous functions $\omega \mapsto \alpha_{\chi,j}(\omega)$ and $\omega \mapsto \beta_{\chi,j}(\omega)$ for $\omega \in [-\pi, \pi]$ and $1 \leq j \leq q$, such that for all $p \geq p_0$,

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

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

Under the assumption, if $\rho_j = 1$ for all $1 \leq j \leq q$, then we are in presence of q factors that are equally pervasive for the

whole cross-section. The left panel of Figure 1 depicts the case when $\rho_1 = 1$. If $\rho_j < 1$ for some j, we permit the presence of "weak" factors and our theoretical analysis explicitly reflects this; see, for example, Onatski (2012) and Freyaldenhoven (2021) for static factor models permitting weak factors. When weak factors are present, the ordering of the variables becomes important as $p \to \infty$, whereas the case of linearly diverging factor strengths is compatible with completely arbitrary cross-sectional ordering. The requirement that $\rho_j > 3/4$ is a minimal one, and generally larger values of ρ_j are required as the dimensionality increases and heavier tails are permitted as discussed later.

Assumptions 2.2 and 2.3 are made to control the serial dependence in $\mathbf{X}_t$.

Assumption 2.2. There exist some constants $\Xi > 0$ and $\varsigma > 2$ such that for all $\ell \geq 0$,

$$\max_{1 \leq i \leq p} |\mathbf{B}_{\ell,i\cdot}|_2 \leq \Xi(1+\ell)^{-\varsigma} \quad \text{and}$$

$$\left(\sum_{j=1}^{q} |\mathbf{B}_{\ell,\cdot j}|_\infty^2 \right)^{1/2} \leq \Xi(1+\ell)^{-\varsigma}.$$

Assumption 2.3. (i) $\quad d$ is a finite positive integer and $\det(\mathcal{A}(z)) \neq 0$ for all $|z| \leq 1$.
(ii) There exist some constants $0 < m_\varepsilon \leq M_\varepsilon$ such that $\|\mathbf{\Gamma}\| \leq M_\varepsilon$ and $\Lambda_{\min}(\mathbf{\Gamma}) \geq m_\varepsilon$.
(iii) There exist a constant $m_\xi > 0$ such that $\inf_{\omega \in [-\pi,\pi]} \mu_{\xi,p}(\omega) \geq m_\xi$.
(iv) There exist some constants $\Xi > 0$ and $\varsigma > 2$ such that for all $\ell \geq 0$,

$$|D_{\ell,ik}| \leq C_{ik}(1+\ell)^{-\varsigma} \text{ with}$$

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

Assumption 2.3 (i) and (ii) are standard in the literature (Lütkepohl 2005) and imply that $\boldsymbol{\xi}_t$ is causal and has finite and nonzero covariance. Under Assumptions 2.2 and 2.3 (iv) (imposed on the Wold decomposition of $\boldsymbol{\xi}_t$ in (3)), the serial dependence in $\mathbf{X}_t$ decays at an algebraic rate. Further, we obtain a uniform bound for $\mu_{\xi,j}(\omega)$ under Assumption 2.3 (iv):

Proposition 2.1. Under Assumption 2.3, uniformly over all $\omega \in [-\pi, \pi]$, there exists some constant $B_\xi > 0$ depending only on M_ε, Ξ and ς, defined in Assumption 2.3 (iii) and (iv), such that $\sup_{\omega \in [-\pi,\pi]} \mu_{\xi,1}(\omega) \leq B_\xi$.

Remark 2.2. Proposition 2.1 and Assumption 2.3 (iii) jointly establish the uniform boundedness of $\mu_{\xi,1}(\omega)$ and $\mu_{\xi,p}(\omega)$, which is commonly assumed in the literature on high-dimensional VAR estimation via ℓ_1-regularization. A sufficient condition for Assumption 2.3 (iii) is that $\max \left\{ \max_{1 \leq i \leq p} \sum_{\ell=1}^{d} |\mathbf{A}_{\ell,i\cdot}|_1, \max_{1 \leq j \leq p} \sum_{\ell=1}^{d} |\mathbf{A}_{\ell,\cdot j}|_1 \right\} \leq \Xi$ for some constant $\Xi > 0$ (Basu and Michailidis 2015). Further, when for example, $d = 1$, Assumption 2.3 (iv) follows if $|\mathbf{A}_1|_\infty \leq \gamma < 1$ since $\max(\|\mathbf{D}_\ell\|_1, \|\mathbf{D}_\ell\|_\infty) \leq \Xi \gamma^\ell$ with $\mathbf{D}_\ell = \mathbf{A}_1^\ell$.

The two latent components $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$, and the number of factors q, are identified thanks to the large gap between the eigenvalues of their spectral density matrices, which follows from Assumption 2.1 and Proposition 2.1. Then by Weyl's inequality, the qth dynamic eigenvalue $\mu_{x,q}(\omega)$ diverges almost everywhere in $[-\pi, \pi]$ as $p \to \infty$, whereas $\mu_{x,q+1}(\omega)$ is uniformly bounded for any $p \in \mathbb{N}$ and ω. This property is exploited in the FNETS methodology as later described in Section 3.2. It is worth stressing that Assumption 2.1 and Proposition 2.1 jointly constitute both a necessary and sufficient condition for the process $\mathbf{X}_t$ to admit the dynamic factor representation in (1), see Forni and Lippi (2001).

Finally, we characterize the common and idiosyncratic innovations.

Assumption 2.4. (i) $\{\mathbf{u}_t\}_{t \in \mathbb{Z}}$ is a sequence of zero-mean, q-dimensional martingale difference vectors with $\mathsf{cov}(\mathbf{u}_t) = \mathbf{I}_q$, and u_{it} and u_{jt} are independent for all $1 \leq i, j \leq q$ with $i \neq j$ and all $t \in \mathbb{Z}$.
(ii) $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is a sequence of zero-mean, p-dimensional martingale difference vectors with $\mathsf{cov}(\boldsymbol{\varepsilon}_t) = \mathbf{I}_p$, and ε_{it} and ε_{jt} are independent for all $1 \leq i, j \leq p$ with $i \neq j$ and all $t \in \mathbb{Z}$.
(iii) $\mathsf{E}(u_{jt}\varepsilon_{it'}) = 0$ for all $1 \leq j \leq q$, $1 \leq i \leq p$ and $t, t' \in \mathbb{Z}$.
(iv) There exist some constants $\nu > 4$ and $\mu_\nu > 0$ such that $\max \left\{ \max_{1 \leq j \leq q} \mathsf{E}(|u_{jt}|^\nu), \max_{1 \leq i \leq p} \mathsf{E}(|\varepsilon_{it}|^\nu) \right\} \leq \mu_\nu$.

Assumption 2.4 (i) and (ii) allow the common and idiosyncratic innovations to be sequences of martingale differences, relaxing the assumption of serial independence found in Forni et al. (2017). Condition (iii) is standard in the factor modeling literature. Under (iv), we require that the innovations have $\nu > 4$ moments, which is considerably weaker than Gaussian or sub-Weibull tails assumed in the literature on VAR modeling of high-dimensional time series (Basu and Michailidis 2015; Kock and Callot 2015; Wong, Li, and Tewari 2020; Krampe and Paparoditis 2021; Masini, Medeiros, and Mendes 2022). Similar approaches to ours, based only on moment conditions, are found in Wu and Wu (2016) who investigate the Lasso performance in deterministic designs under functional dependence, and Adamek, Smeekes, and Wilms (2023) who assume instead near-epoch-dependence. In Appendix F, we separately consider the case when $\mathbf{u}_t$ and $\boldsymbol{\varepsilon}_t$ are Gaussian for the sake of comparison.

3. Network Estimation via FNETS

3.1. Networks Underpinning Factor-Adjusted VAR Processes

Under the latent VAR model in (2), we can define three types of networks underpinning the interconnectedness of $\mathbf{X}_t$ after factor adjustment (Barigozzi and Brownlees 2019).

Let $\mathcal{V} = \{1, \ldots, p\}$ denote the set of vertices representing the p time series. First, the transition matrices $\mathbf{A}_\ell = [A_{\ell,ii'}, 1 \leq i, i' \leq p]$, encode the directed network $\mathcal{N}^G = (\mathcal{V}, \mathcal{E}^G)$ representing Granger causal linkages, with

$$\mathcal{E}^G = \left\{ (i, i') \in \mathcal{V} \times \mathcal{V} : A_{\ell,ii'} \neq 0 \text{ for some } 1 \leq \ell \leq d \right\} \quad (4)$$

as the set of edges. Here, the presence of an edge $(i, i') \in \mathcal{E}^G$ indicates that $\xi_{i',t-\ell}$ Granger causes ξ_{it} at some lag $1 \leq \ell \leq d$.

The second network contains undirected edges representing contemporaneous dependence between VAR innovations $\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$, denoted by $\mathcal{N}^{\mathrm{C}} = (\mathcal{V}, \mathcal{E}^{\mathrm{C}})$; we have $(i, i') \in \mathcal{E}^{\mathrm{C}}$ iff the partial correlation between the ith and i'th elements of $\mathbf{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$ is nonzero. Specifically, letting $\mathbf{\Gamma}^{-1} = \mathbf{\Delta} = [\delta_{ii'}, \, 1 \le i, i' \le p]$, the set of edges is given by

$$\mathcal{E}^{\mathrm{C}} = \left\{ (i, i') \in \mathcal{V} \times \mathcal{V} : i \ne i' \text{ and } -\frac{\delta_{ii'}}{\sqrt{\delta_{ii} \cdot \delta_{i'i'}}} \ne 0 \right\}. \quad (5)$$

Finally, we summarize the aforementioned lead-lag and contemporaneous relations between the variables in a single, undirected network $\mathcal{N}^{\mathrm{L}} = (\mathcal{V}, \mathcal{E}^{\mathrm{L}})$ by means of the long-run partial correlations of $\boldsymbol{\xi}_t$. Let $\mathbf{\Omega} = [\omega_{ii'}, \, 1 \le i, i' \le p]$ denote the long-run partial covariance matrix of $\boldsymbol{\xi}_t$, that is $\mathbf{\Omega} = (\mathbf{\Sigma}_{\xi}(0))^{-1} = 2\pi \mathcal{A}^{\top}(1)\mathbf{\Delta}\mathcal{A}(1)$ under (2). Then, the set of edges of $\mathcal{N}^{\mathrm{L}}$ is

$$\mathcal{E}^{\mathrm{L}} = \left\{ (i, i') \in \mathcal{V} \times \mathcal{V} : i \ne i' \text{ and } -\frac{\omega_{ii'}}{\sqrt{\omega_{ii} \cdot \omega_{i'i'}}} \ne 0 \right\}. \quad (6)$$

Generally, $\mathcal{E}^{\mathrm{L}}$ is greater than $\mathcal{E}^{\mathrm{G}} \cup \mathcal{E}^{\mathrm{C}}$, see Appendix C for a sufficient condition for the absence of an edge (i, i') from $\mathcal{N}^{\mathrm{L}}$. In the remainder of Section 3, we describe the network estimation methodology of FNETS which, consisting of three steps, estimates the three networks while fully accounting for the challenges arising from not directly observing the VAR process $\boldsymbol{\xi}_t$, and investigate its theoretical properties.

3.2. Step 1: Factor Adjustment via Dynamic PCA

As described in Section 2, under our model (1)–(2), there exists a large gap in $\mu_{x,j}(\omega)$, the dynamic eigenvalues of the spectral density matrix of $\mathbf{X}_t$, between those attributed to the factors ($j \le q$) and those which are not ($j \ge q + 1$). With the goal of estimating the autocovariance (ACV) matrix of the latent VAR process $\boldsymbol{\xi}_t$, we exploit this gap in the factor-adjustment step based on dynamic principal component analysis (PCA); see Chapter 9 of Brillinger (1981) for the definition of dynamic PCA and Forni et al. (2000) for its use in the estimation of GDFM. Throughout, we treat q as known and refer to Hallin and Liška (2007) for its consistent estimation under (1).

Denote the ACV matrices of $\mathbf{X}_t$ by $\mathbf{\Gamma}_x(\ell) = \mathrm{E}(\mathbf{X}_{t-\ell}\mathbf{X}_t^{\top})$ for $\ell \ge 0$ and $\mathbf{\Gamma}_x(\ell) = \mathbf{\Gamma}_x^{\top}(-\ell)$ for $\ell \le -1$, and analogously define $\mathbf{\Gamma}_\chi(\ell)$ and $\mathbf{\Gamma}_\xi(\ell)$ with $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ replacing $\mathbf{X}_t$, respectively. Then, $\mathbf{\Sigma}_x(\omega)$ and $\mathbf{\Gamma}_x(\ell)$ satisfy $\mathbf{\Sigma}_x(\omega) = (2\pi)^{-1} \sum_{\ell=-\infty}^{\infty} \mathbf{\Gamma}_x(\ell) \exp(-\iota\ell\omega)$ for all $\omega \in [-\pi, \pi]$. Motivated by this, we estimate $\mathbf{\Sigma}_x(\omega)$ by

$$\widehat{\mathbf{\Sigma}}_x(\omega) = \frac{1}{2\pi} \sum_{\ell=-m}^{m} K\left(\frac{\ell}{m}\right) \widehat{\mathbf{\Gamma}}_x(\ell) \exp(-\iota\ell\omega), \quad (7)$$

with the sample ACV $\widehat{\mathbf{\Gamma}}_x(\ell) = n^{-1} \sum_{t=\ell+1}^{n} \mathbf{X}_{t-\ell}\mathbf{X}_t^{\top}$ when $\ell \ge 0$, and $\widehat{\mathbf{\Gamma}}_x(\ell) = \widehat{\mathbf{\Gamma}}_x(-\ell)^{\top}$ for $\ell < 0$, and the kernel bandwidth $m = \lfloor n^\beta \rfloor$ for some $\beta \in (0, 1)$. We adopt the Bartlett kernel as $K(\cdot)$ which ensures positive semi-definiteness of $\widehat{\mathbf{\Sigma}}_x(\omega)$ (see Appendix F.2.4). Then, we evaluate $\widehat{\mathbf{\Sigma}}_x(\omega)$ at the $2m + 1$ Fourier frequencies $\omega_k, \, -m \le k \le m$ ($\omega_k = 2\pi k/(2m + 1)$ for $0 \le k \le m$, and $\omega_k = -\omega_{|k|}$ for $-m \le k \le -1$), and estimate $\mathbf{\Sigma}_\chi(\omega_k)$ by retaining the contribution from the q largest eigenvalues and eigenvectors only. That

is, we obtain $\widehat{\mathbf{\Sigma}}_\chi(\omega_k) = \sum_{j=1}^{q} \widehat{\mu}_{x,j}(\omega_k)\widehat{\mathbf{e}}_{x,j}(\omega_k)(\widehat{\mathbf{e}}_{x,j}(\omega_k))^*$ (with $*$ denoting the transposed complex conjugate), where $\widehat{\mu}_{x,1}(\omega) \ge \cdots \ge \widehat{\mu}_{x,q}(\omega)$, denote the q leading eigenvalues of $\widehat{\mathbf{\Sigma}}_x(\omega)$ and $\widehat{\mathbf{e}}_{x,j}(\omega)$ the associated (normalized) eigenvectors. From this, an estimator of $\mathbf{\Gamma}_\chi(\ell)$ at a given lag $\ell \in \mathbb{N}$, is obtained via inverse Fourier transform as $\widehat{\mathbf{\Gamma}}_\chi(\ell) = 2\pi(2m + 1)^{-1} \sum_{k=-m}^{m} \widehat{\mathbf{\Sigma}}_\chi(\omega_k) \exp(\iota\ell\omega_k)$ and finally, we estimate the ACV matrices of $\boldsymbol{\xi}_t$ with $\widehat{\mathbf{\Gamma}}_\xi(\ell) = \widehat{\mathbf{\Gamma}}_x(\ell) - \widehat{\mathbf{\Gamma}}_\chi(\ell)$, by virtue of Assumption 2.4 (iii).

3.3. Step 2: Estimation of VAR Parameters and $\mathcal{N}^{\mathrm{G}}$

Recalling the VAR(d) model in (2), let $\boldsymbol{\beta} = [\mathbf{A}_\ell, \, 1 \le \ell \le d]^{\top} \in \mathbb{R}^{(pd) \times p}$ denote the matrix collecting all the VAR parameters. When $\boldsymbol{\xi}_t$ is directly observable, ℓ_1-regularized least squares or maximum likelihood estimators have been proposed for $\boldsymbol{\beta}$, see the references given in Introduction. In the context of factor-adjusted regression modeling where the aim is to estimate the regression structure in the latent idiosyncratic process, it has been proposed to apply the ℓ_1-regularization methods after estimating the entire latent process by, say, $\widehat{\boldsymbol{\xi}}_t$ (Fan, Ke, and Wang 2020; Fan, Masini, and Medeiros 2021; Fan, Lou, and Yu 2023; Krampe and Margaritella 2021). However, such an approach possibly suffers from the lack of statistical efficiency due to having to control the estimation errors in $\widehat{\boldsymbol{\xi}}_t$ uniformly for all $1 \le t \le n$. Instead, we make use of the Yule-Walker (YW) equation $\boldsymbol{\beta} = \mathbb{G}^{-1}\mathbb{g}$, where

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

$$\mathbb{g} = \begin{bmatrix} \mathbf{\Gamma}_\xi(1) \\ \vdots \\ \mathbf{\Gamma}_\xi(d) \end{bmatrix},$$

with $\mathbb{G}$ being always invertible since $\Lambda_{\min}(\mathbb{G}) \ge 2\pi m_\xi > 0$ by Assumption 2.3 (iii). We propose to estimate $\boldsymbol{\beta}$ as a regularized YW estimator based on $\widehat{\mathbb{G}}$ and $\widehat{\mathbb{g}}$, which are obtained by replacing $\mathbf{\Gamma}_\xi(\ell)$ with $\widehat{\mathbf{\Gamma}}_\xi(\ell)$ derived in Step 1 of FNETS via dynamic PCA, in the definitions of $\mathbb{G}$ and $\mathbb{g}$, respectively.

To handle the high dimensionality, we consider an ℓ_1-regularised estimator for $\boldsymbol{\beta}$ which solves the following ℓ_1-penalised M-estimation problem

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\mathbf{M} \in \mathbb{R}^{pd \times p}} \mathrm{tr}\left(\mathbf{M}^{\top}\widehat{\mathbb{G}}\mathbf{M} - 2\mathbf{M}^{\top}\widehat{\mathbb{g}}\right) + \lambda|\mathbf{M}|_1 \quad (8)$$

with a tuning parameter $\lambda > 0$. Note that the matrix $\widehat{\mathbb{G}}$ is guaranteed to be positive semi-definite (see Appendix F.2.4), thus the problem in (8) is convex with a global minimizer. We note the similarity between (8) and the Lasso estimator, but our estimator is specifically tailored for the problem of estimating the parameters for the latent VAR process $\boldsymbol{\xi}_t$ by means of second-order moments only, and thus differs fundamentally from the Lasso-type estimators proposed for high-dimensional VAR estimation. In Appendix A, we propose an alternative estimator based on a constrained ℓ_1-minimization approach closely related to the Dantzig selector (Candès and Tao 2007).

Once the VAR parameters are estimated, we propose to estimate the edge set of $\mathcal{N}^G$ in (4) by the set of indices of the nonzero elements of a thresholded version of $\widehat{\boldsymbol{\beta}}$, denoted by $\widehat{\boldsymbol{\beta}}(\mathsf{t}) = [\widehat{\beta}_{ij} \cdot \mathbb{I}_{\{|\widehat{\beta}_{ij}| > \mathsf{t}\}}]$, with some threshold $\mathsf{t} > 0$.

3.4. Step 3: Estimation of $\mathcal{N}^C$ and $\mathcal{N}^L$

Recall that the edge sets of $\mathcal{N}^C$ and $\mathcal{N}^L$ defined in (5)–(6), are given by the supports of $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$. Given $\widehat{\boldsymbol{\beta}}$ in (8) which estimates $\boldsymbol{\beta}$, a natural estimator of $\boldsymbol{\Gamma}$ arises from the YW equation $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_\xi(0) - \sum_{\ell=1}^{d} \mathbf{A}_\ell \boldsymbol{\Gamma}_\xi(\ell) = \boldsymbol{\Gamma}_\xi(0) - \boldsymbol{\beta}^\top \mathsf{g}$, as $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}}_\xi(0) - \widehat{\boldsymbol{\beta}}^\top \widehat{\mathsf{g}}$. Then, we propose to estimate $\boldsymbol{\Delta} = \boldsymbol{\Gamma}^{-1}$ via constrained ℓ_1-minimization as

$$\check{\boldsymbol{\Delta}} = \arg\min_{\mathbf{M} \in \mathbb{R}^{p \times p}} |\mathbf{M}|_1 \quad \text{subject to} \quad \left|\widehat{\boldsymbol{\Gamma}}\mathbf{M} - \mathbf{I}\right|_\infty \leq \eta, \quad (9)$$

where $\eta > 0$ is a tuning parameter. This approach has originally been proposed for estimating the precision matrix of independent data (Cai, Liu, and Luo 2011), which we extend to time series settings. Since $\check{\boldsymbol{\Delta}} = [\check{\delta}_{ii'}, 1 \leq i, j \leq p]$ is not guaranteed to be symmetric, a symmetrization step is performed to obtain $\widehat{\boldsymbol{\Delta}} = [\widehat{\delta}_{ii'}, 1 \leq i, i' \leq p]$ with $\widehat{\delta}_{ii'} = \check{\delta}_{ii'} \cdot \mathbb{I}_{\{|\check{\delta}_{ii'}| \leq |\check{\delta}_{i'i}|\}} + \check{\delta}_{i'i} \cdot \mathbb{I}_{\{|\check{\delta}_{i'i}| < |\check{\delta}_{ii'}|\}}$. Then, the edge set of $\mathcal{N}^C$ in (5) is estimated by the support of the thresholded estimator $\widehat{\boldsymbol{\Delta}}(\mathsf{t}_\delta) = [\widehat{\delta}_{ii'} \cdot \mathbb{I}_{\{|\widehat{\delta}_{ii'}| > \mathsf{t}_\delta\}}, 1 \leq i, i' \leq p]$ with some threshold $\mathsf{t}_\delta > 0$.

Finally, we estimate $\boldsymbol{\Omega} = 2\pi(\mathcal{A}(1))^\top \boldsymbol{\Delta} \mathcal{A}(1)$ by replacing $\mathcal{A}(1)$ and $\boldsymbol{\Delta}$ with their estimators. We adopt the thresholded estimator $\widehat{\boldsymbol{\beta}}(\mathsf{t}) = [\widehat{\mathbf{A}}_1(\mathsf{t}), \ldots, \widehat{\mathbf{A}}_d(\mathsf{t})]^\top$, to obtain $\widehat{\mathcal{A}}(1) = \mathbf{I} - \sum_{\ell=1}^{d} \widehat{\mathbf{A}}_\ell(\mathsf{t})$ and set $\widehat{\boldsymbol{\Omega}} = 2\pi(\widehat{\mathcal{A}}(1))^\top \widehat{\boldsymbol{\Delta}} \widehat{\mathcal{A}}(1)$. Analogously, the edge set of $\mathcal{N}^L$ in (6) is obtained by thresholding $\widehat{\boldsymbol{\Omega}} = [\widehat{\omega}_{ii'}, 1 \leq i, i' \leq p]$ with some threshold $\mathsf{t}_\omega > 0$, as the support of $\widehat{\boldsymbol{\Omega}}(\mathsf{t}_\omega) = [\widehat{\omega}_{ii'} \cdot \mathbb{I}_{\{|\widehat{\omega}_{ii'}| > \mathsf{t}_\omega\}}, 1 \leq i, i' \leq p]$.

3.5. Theoretical Properties

We prove the consistency of FNETS in network estimation by establishing the theoretical properties of each of its three steps in Sections 3.5.1–3.5.3. Then in Section 3.5.4, we present the results for a special case where $\boldsymbol{\chi}_t$ admits a static representation, $n \asymp p$ and $\mathsf{E}(|X_{it}|^\nu) < \infty$ for $\nu > 8$, for ease of comparing our results to the existing ones.

Hereafter, we define

$$\psi_n = \left(\frac{m}{n^{1-2/\nu}} \vee \sqrt{\frac{m \log(m)}{n}} \right) \quad \text{and}$$

$$\vartheta_{n,p} = \left(\frac{m(np)^{2/\nu} \log^{7/2}(p)}{n} \vee \sqrt{\frac{m \log(mp)}{n}} \right), \quad (10)$$

where the dependence of these quantities on ν is omitted for simplicity.

3.5.1. Factor Adjustment via Dynamic PCA
We first establish the consistency of the dynamic PCA-based estimator of $\boldsymbol{\Gamma}_\chi(\ell)$.

Theorem 3.1. Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are met. Then, for any finite positive integer $s \leq d$, as $n, p \to \infty$,

$$\max_{\ell:|\ell| \leq s} \frac{1}{p} \left\| \widehat{\boldsymbol{\Gamma}}_\chi(\ell) - \boldsymbol{\Gamma}_\chi(\ell) \right\|_F = O_P\left(qp^{2(1-\rho_q)} \left(\psi_n \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right),$$

$$\max_{\ell:|\ell| \leq s} \left| \widehat{\boldsymbol{\Gamma}}_\chi(\ell) - \boldsymbol{\Gamma}_\chi(\ell) \right|_\infty = O_P\left(qp^{2(1-\rho_q)} \left(\vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right).$$

Remark 3.1. (a) Theorem 3.1 is complemented by Proposition F.15 in Appendix which establishes the consistency of the spectral density matrix estimator $\widehat{\boldsymbol{\Sigma}}_\chi(\omega)$ uniformly over $\omega \in [-\pi, \pi]$ in both Frobenius and ℓ_∞-norms.

(b) Both ψ_n and $\vartheta_{n,p}$ in (10) increase with the bandwidth m. It is possible to find m that minimizes for example, $(\vartheta_{n,p} \vee m^{-1})$ which, roughly speaking, represents the bias-variance trade-off in the estimation of the spectral density matrix $\boldsymbol{\Sigma}_x(\omega)$. For example, in light-tailed settings with large enough ν, the choice $m \asymp (n \log^{-1}(np))^{1/3}$ leads to the minimal rate in ℓ_∞-norm $(\vartheta_{n,p} \vee m^{-1}) \asymp (\log(np)/n)^{1/3}$ which nearly matches the optimal nonparametric rate when using the Bartlett kernel as in (7) (Priestley 1982, p. 463).

(c) Consistency in Frobenius norm depends on ψ_n which tends to zero as $n \to \infty$ without placing any constraint on the relative rate of divergence between n and p. Consistency in ℓ_∞-norm is determined by $\vartheta_{n,p}$ which depends on the interplay between the dimensionality and the tail behavior. Generally, the estimation error worsens as weaker factors are permitted ($\rho_q < 1$ in Assumption 2.1) and as p grows, and also when ν is small such that heavier tails are permitted. Consider the case when all factors are strong (i.e., $\rho_j = 1$). If $p \asymp n$, then ℓ_∞-consistency holds with an appropriately chosen $m = n^\beta$, $\beta \in (0, 1)$, that leads to $\vartheta_{n,p} = o(1)$, provided that $\nu > 4$. When all moments of u_{jt} and ε_{it} exist, we achieve ℓ_∞-consistency even in the ultra high-dimensional case where $\log(p) = o(n)$.

From Theorem 3.1, the following proposition immediately follows.

Proposition 3.2. Suppose that the conditions in Theorem 3.1 are met and let Assumption 2.1 hold with $\rho_j = 1$, $1 \leq j \leq q$. Then, $\mathsf{P}(\mathcal{E}_{n,p}) \to 1$ as $n, p \to \infty$, where

$$\mathcal{E}_{n,p} = \left\{ \max_{-d \leq \ell \leq d} \left| \widehat{\boldsymbol{\Gamma}}_\xi(\ell) - \boldsymbol{\Gamma}_\xi(\ell) \right|_\infty \leq C_\xi \left(\vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right\}. \quad (11)$$

for some constant $C_\xi > 0$.

From Proposition 3.2, we have ℓ_∞-consistency of $\widehat{\boldsymbol{\Gamma}}_\xi(\ell)$ in the presence of strong factors. Although it is possible to trace the effect of weak factors on the estimation of $\boldsymbol{\Gamma}_\xi(\ell)$ (see Corollary F.17), we make this simplifying assumption to streamline the presentation of the theoretical results of the subsequent Steps 2–3 of FNETS.

Remark 3.2. In Appendix F.2.8, we show that if $\boldsymbol{\chi}_t$ admits the static representation discussed in Remark 2.1, the rate in Proposition 3.2 is further improved as

$$\max_{\ell:\,|\ell|\leq d}\left|\widehat{\mathbf{\Gamma}}_\xi(\ell)-\mathbf{\Gamma}_\xi(\ell)\right|_\infty = O_P\left(\widetilde{\vartheta}_{n,p}\vee\frac{1}{\sqrt{p}}\right)\ \text{with}$$

$$\widetilde{\vartheta}_{n,p}=\left(\frac{p^{2/\nu}\log^3(p)}{n^{1-2/\nu}}\vee\sqrt{\frac{\log(p)}{n}}\right).\tag{12}$$

The term $\widetilde{\vartheta}_{n,p}$ comes from bounding $\max_{\ell:\,|\ell|\leq d}|\widehat{\mathbf{\Gamma}}_x(\ell)-\mathbf{\Gamma}_x(\ell)|_\infty$. Hence, the improved rate in (12) is comparable to the rate attained when we directly observe $\boldsymbol{\xi}_t$ apart from the presence of $p^{-1/2}$, which is due to the presence of latent factors; similar observations are made in Theorem 3.1 of Fan, Liao, and Mincheva (2013).

3.5.2. Estimation of VAR Parameters and $\mathcal{N}^G$
We measure the sparsity of $\boldsymbol{\beta}$ by $s_{0,j}=|\boldsymbol{\beta}_{\cdot j}|_0$, $s_0=\sum_{j=1}^p s_{0,j}$ and $s_{\text{in}}=\max_{1\leq j\leq p}s_{0,j}$. When $d=1$, the quantity s_{in} coincides with the maximum in-degree per node of $\mathcal{N}^G$.

Proposition 3.3. Suppose that $C_\xi s_{\text{in}}(\vartheta_{n,p}\vee m^{-1}\vee p^{-1/2})\leq \pi m_\xi/16$, where m_ξ is defined in Assumption 2.3 (iii). Also, set $\lambda\geq 4C_\xi(\|\boldsymbol{\beta}\|_1+1)(\vartheta_{n,p}\vee m^{-1}\vee p^{-1/2})$ in (8). Then, conditional on $\mathcal{E}_{n,p}$ defined in (11), we have

$$\left|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right|_\infty\leq\left|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right|_2\leq\frac{32\sqrt{s_{\text{in}}}\lambda}{\pi m_\xi}\quad\text{and}\quad\left|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right|_1\leq\frac{8s_0\lambda}{\pi m_\xi}.$$

Following Loh and Wainwright (2012), the proof of Proposition 3.3 proceeds by showing that, conditional on $\mathcal{E}_{n,p}$, the matrix $\widehat{\mathbb{G}}$ meets a restricted eigenvalue condition (Bickel, Ritov, and Tsybakov 2009) and the deviation bound is controlled as $|\widehat{\mathbb{G}}\boldsymbol{\beta}-\widehat{\mathbf{g}}|_\infty\leq\lambda/4$. Then, thanks to Proposition 3.2, as $n,p\to\infty$, the estimation errors of $\widehat{\boldsymbol{\beta}}$ in ℓ_∞- and ℓ_1-norms, are bounded as in Proposition 3.3 with probability tending to one.

Remark 3.3. As noted in Remark 2.2, the boundedness of $\mu_{\xi,j}(\omega)$ follows from that of $\|\boldsymbol{\beta}\|_1=\max_{1\leq j\leq p}\sum_{\ell=1}^d|\mathbf{A}_{\ell,j\cdot}|_1$, in which case $\|\boldsymbol{\beta}\|_1$ appearing in the assumed lower bound on λ, does not inflate the rate of the estimation errors. In the light-tailed situation, with the optimal bandwidth $m\asymp(n\log^{-1}(np))^{1/3}$ as specified in Remark 3.1(b), it is required that $s_{\text{in}}=O((n\log^{-1}(np))^{1/3}\wedge\sqrt{p})$, which still allows the number of nonzero entries in *each* row of $\mathbf{A}_\ell$ to grow with p. Here, the exponent $1/3$ in place of $1/2$ often found in the literature, comes from adopting the most general approach to time series factor modeling which necessitates selecting a bandwidth for frequency domain-based factor adjustment.

For sign consistency of the Lasso estimator, the (almost) necessary and sufficient condition is the so-called irrepresentable condition (Zhao and Yu 2006), which is known to be highly stringent (Tardivel and Bogdan 2022). Alternatively, Medeiros and Mendes (2016) propose an adaptive Lasso estimator with data-driven weights for high-dimensional VAR estimation when $\boldsymbol{\xi}_t$ is directly observed. Instead, we propose to additionally threshold $\widehat{\boldsymbol{\beta}}$ and obtain $\widehat{\boldsymbol{\beta}}(t)$, whose support consistently estimates the edge set of $\mathcal{N}^G$.

Corollary 3.4. Suppose that the conditions of Proposition 3.3 are met. If

$$\min_{(i,j):\,|\beta_{ij}|>0}|\beta_{ij}|>2t\tag{13}$$

with $t=32\sqrt{s_{\text{in}}}\lambda/(\pi m_\xi)$, then we have $\text{sign}(\widehat{\boldsymbol{\beta}}(t))=\text{sign}(\boldsymbol{\beta})$ conditional on $\mathcal{E}_{n,p}$.

3.5.3. Estimation of $\mathcal{N}^C$ and $\mathcal{N}^L$
Let $s_\delta(\varrho)=\max_{1\leq i\leq p}\sum_{i'=1}^p|\delta_{ii'}|^\varrho$, $\varrho\in[0,1)$, denote the (weak) sparsity of $\boldsymbol{\Delta}=[\delta_{ii'},1\leq i,i'\leq p]$. Also, define $s_{\text{out}}=\max_{1\leq j\leq p}\sum_{\ell=1}^d|\mathbf{A}_{\ell,\cdot j}|_0$ which, complementing s_{in}, represents the sparsity of the out-going edges of $\mathcal{N}^G$. Analogously as in Proposition 3.3, we establish deterministic guarantees for $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{\Omega}}$ conditional on $\mathcal{E}_{n,p}$.

Proposition 3.5. Suppose that the conditions in Propositions 3.3 are met, and set $\eta=Cs_{\text{in}}\|\boldsymbol{\Delta}\|_1(\|\boldsymbol{\beta}\|_1+1)(\vartheta_{n,p}\vee m^{-1}\vee p^{-1/2})$ in (9), with C depending only on C_ξ and m_ξ. Then, conditional on $\mathcal{E}_{n,p}$ defined in (11), we have:

(i) $|\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta}|_\infty\leq 4\|\boldsymbol{\Delta}\|_1\eta$ and $\|\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta}\|\leq 12s_\delta(\varrho)(4\|\boldsymbol{\Delta}\|_1\eta)^{1-\varrho}$.
(ii) If also $s_{\text{out}}t\leq\|\mathcal{A}(1)\|_1$ with t chosen as in Corollary 3.4, then,

$$\left|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\right|_\infty\leq 4\pi\|\mathcal{A}(1)\|_1\left(3\|\boldsymbol{\Delta}\|s_{\text{out}}t+16\|\mathcal{A}(1)\|_1\|\boldsymbol{\Delta}\|_1\eta\right).$$

Together with Assumption 2.3 (ii), Proposition 3.5(i) indicates asymptotic positive definiteness of $\widehat{\boldsymbol{\Delta}}$ provided that $\boldsymbol{\Delta}$ is sufficiently sparse, as measured by $\|\boldsymbol{\Delta}\|_1$ and $s_\delta(\varrho)$. By definition, $\mathcal{N}^L$ combines $\mathcal{N}^G$ and $\mathcal{N}^C$ and consequently, its sparsity structure is determined by the sparsity of the other two networks, which is reflected in Proposition 3.5(ii). Specifically, the term $\|\mathcal{A}(1)\|_1$ is related to the out-going property of $\mathcal{N}^G$, and satisfies $\|\mathcal{A}(1)\|_1\leq\max_{1\leq j\leq p}\sum_{\ell=1}^d|\mathbf{A}_{\ell,\cdot j}|_1$, where the boundedness of the right-hand side is sufficient for the boundedness of $\mu_{\xi,j}(\omega)$ (Remark 2.2). Also, $\|\boldsymbol{\Delta}\|_1$ reflects the sparsity of the edge set of $\mathcal{N}^C$, and the tuning parameter η depends on the sparsity of the in-coming edges of $\mathcal{N}^G$ through $\|\boldsymbol{\beta}\|_1=\max_{1\leq j\leq p}\sum_{\ell=1}^d|\mathbf{A}_{\ell,j\cdot}|_1$ and s_{in}.

Similarly as in Corollary 3.4, we can show the consistency of the thresholded estimators $\widehat{\boldsymbol{\Delta}}(t_\delta)$ and $\widehat{\boldsymbol{\Omega}}(t_\omega)$ in estimating the edge sets of $\mathcal{N}^C$ and $\mathcal{N}^L$, respectively.

Corollary 3.6. Suppose that conditions of Proposition 3.5 are met. Conditional on $\mathcal{E}_{n,p}$:

(i) If $\min_{(i,i'):\,|\delta_{ii'}|>0}|\delta_{ii'}|>2t_\delta$ with $t_\delta=4\|\boldsymbol{\Delta}\|_1\eta$, we have $\text{sign}(\widehat{\boldsymbol{\Delta}}(t_\delta))=\text{sign}(\boldsymbol{\Delta})$.
(ii) If $\min_{(i,i'):\,|\omega_{ii'}|>0}|\omega_{ii'}|>2t_\omega$ with $t_\omega=4\pi\|\mathcal{A}(1)\|_1(3\|\boldsymbol{\Delta}\|s_{\text{out}}t+16\|\mathcal{A}(1)\|_1\|\boldsymbol{\Delta}\|_1\eta)$, we have $\text{sign}(\widehat{\boldsymbol{\Omega}}(t_\omega))=\text{sign}(\boldsymbol{\Omega})$.

3.5.4. The Case of the Static Factor Model
For ease of comparing the performance of FNETS with the existing results, we focus on the static factor model setting discussed in Remark 2.1, and assume that $n\asymp p$ and $\max(\|\boldsymbol{\beta}\|_1,\|\mathcal{A}(1)\|_1)=O(1)$. Then, from Remark 3.2 and the proof of Proposition 3.3, we obtain $\max_{1\leq j\leq p}|\widehat{\boldsymbol{\beta}}_{\cdot j}-\boldsymbol{\beta}_{\cdot j}|_2=O_P(\sqrt{s_{\text{in}}\log(n)/n})$ provided that $\nu>8$, such that the condition in (13) is written with $t\asymp\sqrt{s_{\text{in}}\log(n)/n}$. That is, $\widehat{\boldsymbol{\beta}}$ and its thresholded counterpart proposed for the estimation of the latent VAR process, perform as well as the benchmark derived under independence and Gaussianity in the Lasso literature (van de Geer, Bühlmann, and Zhou 2011). In this same

setting, the factor-adjusted regression estimation method of Fan, Masini, and Medeiros (2021), when applied to the problem of VAR parameter estimation, yields an estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{FARM}}$ which attains $\max_{1 \le j \le p} |\widehat{\boldsymbol{\beta}}_{\cdot j}^{\mathrm{FARM}} - \boldsymbol{\beta}_{\cdot j}|_2 = O_P(\sqrt{s_{\mathrm{in}}} n^{-1/2+5/\nu})$ under strong mixingness, see their Theorem 3. Here, the larger O_P-bound compared to ours stems from that $\widehat{\boldsymbol{\beta}}^{\mathrm{FARM}}$ requires the estimation of ξ_{it} for all i and t, the error from which increases with n as well as p. This demonstrates the efficacy of adopting our regularised YW estimator.

Continuing with the same setting, Propositions 3.5 implies that

$$\left|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\right|_\infty = O_P\left(\|\boldsymbol{\Delta}\|_1^2 s_{\mathrm{in}} \sqrt{\frac{\log(n)}{n}}\right) \text{ and}$$

$$\left|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\right|_\infty = O_P\left((s_{\mathrm{out}} \vee \|\boldsymbol{\Delta}\|_1^2 \sqrt{s_{\mathrm{in}}}) \sqrt{\frac{s_{\mathrm{in}} \log(n)}{n}}\right).$$

The former is comparable (up to s_{in}) to the results in Theorem 4 of Cai, Liu, and Luo (2011) derived for estimating a sparse precision matrix of independent random vectors.

4. Forecasting via FNETS

4.1. Forecasting under the Static Factor Model Representation

For given time horizon $a \ge 0$, the best linear predictor of $\boldsymbol{\chi}_{n+a}$ based on $\boldsymbol{\chi}_{n-\ell}$, $\ell \ge 0$, is

$$\boldsymbol{\chi}_{n+a|n} = \sum_{\ell=0}^{\infty} \mathbf{B}_{\ell+a} \mathbf{u}_{n-\ell}. \tag{14}$$

under (1). Following Forni et al. (2005), we consider a forecasting method for the factor-driven component which estimates $\boldsymbol{\chi}_{n+a|n}$ under a *restricted* GDFM that admits a static representation of finite dimension. We formalize the static factor model discussed in Remark 2.1 in the following assumption.

Assumption 4.1. (i) There exist two finite positive integers m_1 and m_2 such that $m_1 + 1 \ge m_2$, $\boldsymbol{\chi}_t = \mathcal{M}^{(1)}(L)\mathbf{f}_t$ and $\mathbf{f}_t = \mathcal{M}^{(2)}(L)\mathbf{u}_t$ where $\mathcal{M}^{(1)}(L) = \sum_{\ell=0}^{m_1} \mathbf{M}_\ell^{(1)} L^\ell$ with $\mathbf{M}^{(1)} \in \mathbb{R}^{p \times q}$, $\mathcal{M}^{(2)}(L) = \sum_{\ell=0}^{m_2} \mathbf{M}_\ell^{(2)} L^\ell$ with $\mathbf{M}^{(2)} \in \mathbb{R}^{q \times q}$ and $\det(\mathcal{M}^{(2)}(z)) \ne 0$ for all $|z| \le 1$.

(ii) Let $\mu_{\chi,j}$, $1 \le j \le r$, denote the jth largest eigenvalue of $\boldsymbol{\Gamma}_\chi(0)$. Then, there exist a positive integer $p_0 \ge 1$, constants $\varrho_j \in (7/8, 1]$ with $\varrho_1 \ge \cdots \ge \varrho_r$, and pairs of positive constants $(\alpha_{\chi,j}, \beta_{\chi,j})$, $1 \le j \le r$, such that for all $p \ge p_0$,

$$\beta_{\chi,1} \ge \frac{\mu_{\chi,1}}{p^{\varrho_1}} \ge \alpha_{\chi,1} > \beta_{\chi,2} \ge \frac{\mu_{\chi,2}}{p^{\varrho_2}}$$

$$\ge \cdots \ge \alpha_{\chi,r-1} > \beta_{\chi,r} \ge \frac{\mu_{\chi,r}}{p^{\varrho_r}} \ge \alpha_{\chi,r} > 0.$$

In part (i), $\boldsymbol{\chi}_t$ admits a static representation with $r = q(m_1 + 1)$ factors: $\boldsymbol{\chi}_t = \boldsymbol{\Lambda}\mathbf{F}_t$, where $\boldsymbol{\Lambda} = [\mathbf{M}_\ell^{(1)}, 0 \le \ell \le m_1]$, $\mathbf{F}_t = (\mathbf{f}_t^\top, \ldots, \mathbf{f}_{t-m_1}^\top)^\top$ and $\mathbf{f}_t = \mathcal{M}^{(2)}(L)\mathbf{u}_t$. The condition that $m_1 + 1 \ge m_2$ is made for convenience, and the proposed estimator of $\boldsymbol{\chi}_{n+a|n}$ can be modified accordingly when it is relaxed.

Remark 4.1. Under Assumption 4.1 (i), the r-vector of static factors, $\mathbf{F}_t$, is driven by the q-dimensional common shocks $\mathbf{u}_t$. If $q < r$, Anderson and Deistler (2008) show that $\mathbf{F}_t$ always admits a VAR(h) representation: $\mathbf{F}_t = \sum_{\ell=1}^h \mathbf{G}_\ell \mathbf{F}_{t-\ell} + \mathbf{H}\mathbf{u}_t$ for some finite positive integer h and $\mathbf{H} \in \mathbb{R}^{r \times q}$. Then, $\mathbf{X}_t$ has a factor-augmented VAR representation:

$$\mathbf{X}_t = \boldsymbol{\Lambda} \sum_{\ell=1}^h \mathbf{G}_\ell \mathbf{F}_{t-\ell} + \boldsymbol{\Lambda}\mathbf{H}\mathbf{u}_t + \sum_{\ell=1}^d \mathbf{A}_\ell \boldsymbol{\xi}_{t-\ell} + \boldsymbol{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$$

$$= \sum_{\ell=1}^{d \vee h} \mathbf{C}_\ell \mathbf{F}_{t-\ell} + \sum_{\ell=1}^d \mathbf{A}_\ell \mathbf{X}_{t-\ell} + \boldsymbol{\nu}_t,$$

with $\mathbf{C}_\ell = \boldsymbol{\Lambda}\mathbf{G}_\ell \mathbb{I}_{\{\ell \le h\}} - \mathbf{A}_\ell \boldsymbol{\Lambda} \mathbb{I}_{\{\ell \le d\}}$ and $\boldsymbol{\nu}_t = \boldsymbol{\Lambda}\mathbf{H}\mathbf{u}_t + \boldsymbol{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t$. This model is a generalization of the factor augmented forecasting model considered by Stock and Watson (2002) where only the factor-driven component is present, and it is also considered by Fan, Masini, and Medeiros (2021).

It immediately follows from Proposition 2.1 that $\|\boldsymbol{\Gamma}_\xi(0)\| \le 2\pi B_\xi$. This, combined with Assumption 4.1 (ii), indicates the presence of a large gap in the eigenvalues of $\boldsymbol{\Gamma}_x(0)$, which allows the asymptotic identification of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ in the time domain, as well as that of the number of static factors r. Throughout, we treat r as known, and refer to for example Bai and Ng (2002) and Ahn and Horenstein (2013), for its estimation.

Let $(\mu_{\chi,j}, \mathbf{e}_{\chi,j})$, $1 \le j \le r$, denote the pairs of eigenvalues and eigenvectors of $\boldsymbol{\Gamma}_\chi(0)$ ordered such that $\mu_{\chi,1} \ge \cdots \ge \mu_{\chi,r}$. Then, $\boldsymbol{\Gamma}_\chi(0) = \mathbf{E}_\chi \mathcal{M}_\chi \mathbf{E}_\chi^\top$ with $\mathcal{M}_\chi = \mathrm{diag}(\mu_{\chi,j}, 1 \le j \le r)$ and $\mathbf{E}_\chi = [\mathbf{e}_{\chi,j}, 1 \le j \le r]$. Under Assumption 4.1 (i), we have $\boldsymbol{\chi}_{n+a|n}$ in (14) satisfy $\boldsymbol{\chi}_{n+a|n} = \mathrm{Proj}(\boldsymbol{\chi}_{n+a}|\mathbf{F}_{n-\ell}, \ell \ge 0) = \mathrm{Proj}(\boldsymbol{\chi}_{n+a}|\mathbf{F}_n) = \boldsymbol{\Gamma}_\chi(-a)\mathbf{E}_\chi \mathcal{M}_\chi^{-1}\mathbf{E}_\chi^\top \boldsymbol{\chi}_n$, where $\mathrm{Proj}(\cdot|\mathbf{z})$ denotes the linear projection operator onto the linear space spanned by $\mathbf{z}$. When $a = 0$, we trivially have $\boldsymbol{\chi}_{t|n} = \boldsymbol{\chi}_t$ for $t \le n$. Then, a natural estimator of $\boldsymbol{\chi}_{n+a|n}$ is

$$\widehat{\boldsymbol{\chi}}_{n+a|n}^{\mathrm{res}} = \widehat{\boldsymbol{\Gamma}}_\chi(-a)\widehat{\mathbf{E}}_\chi \widehat{\mathcal{M}}_\chi^{-1}\widehat{\mathbf{E}}_\chi^\top \mathbf{X}_n, \tag{15}$$

where $(\widehat{\mu}_{\chi,j}, \widehat{\mathbf{e}}_{\chi,j})$, $1 \le j \le r$, denote the pairs of eigenvalues and eigenvectors of $\widehat{\boldsymbol{\Gamma}}_\chi(0)$, and $\widehat{\boldsymbol{\Gamma}}_\chi(\ell)$, $\ell \in \{0, a\}$, are estimated as described in Section 3.2. As a by-product, we obtain the in-sample estimator by setting $a = 0$, as $\widehat{\boldsymbol{\chi}}_t^{\mathrm{res}} = \widehat{\mathbf{E}}_\chi \widehat{\mathbf{E}}_\chi^\top \mathbf{X}_t$ for $1 \le t \le n$.

Remark 4.2. Our proposed estimator $\widehat{\boldsymbol{\chi}}_{n+a|n}^{\mathrm{res}}$ differs from that of Forni et al. (2005), as they estimate the factor space via generalized PCA on $\widehat{\boldsymbol{\Gamma}}_\chi(0)$. This in effect replaces $\widehat{\mathbf{E}}_\chi$ in (15) with the eigenvectors of $\mathbf{W}^{-1}\widehat{\boldsymbol{\Gamma}}_\chi(0)$ where $\mathbf{W}$ is a diagonal matrix containing the estimators of the sample variance of $\boldsymbol{\xi}_t$. Such an approach may gain in efficiency compared to ours in the same way a weighted least squares estimator is more efficient than the ordinary one in the presence of heteroscedasticity. However, since we investigate the consistency of $\widehat{\boldsymbol{\chi}}_{n+a|n}^{\mathrm{res}}$ without deriving its asymptotic distribution, we do not explore such approach in this article.

In Appendix B, we present an alternative forecasting method that operates under an *unrestricted* GDFM, that is it does not require Assumption 4.1. Referred to as $\widehat{\boldsymbol{\chi}}_{n+a|n}^{\mathrm{unr}}$, we compare its performance with that of $\widehat{\boldsymbol{\chi}}_{n+a|n}^{\mathrm{res}}$ in numerical studies.

Once VAR parameters are estimated by $\widehat{\boldsymbol{\beta}} = [\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_d]^\top$ as in (8), we produce a forecast of $\boldsymbol{\xi}_{n+a}$ given $\mathbf{X}_t$, $t \le n$, by estimating the best linear predictor $\boldsymbol{\xi}_{n+a|n} = \sum_{\ell=1}^{d} \mathbf{A}_\ell \boldsymbol{\xi}_{n+1-\ell|n}$ (with $\boldsymbol{\xi}_{t|n} = \boldsymbol{\xi}_t$ for $t \le n$), as

$$\widehat{\boldsymbol{\xi}}_{n+a|n} = \sum_{\ell=1}^{\max(1,a)-1} \widehat{\mathbf{A}}_\ell \widehat{\boldsymbol{\xi}}_{n+a-\ell|n} + \sum_{\ell=\max(1,a)}^{d} \widehat{\mathbf{A}}_\ell \widehat{\boldsymbol{\xi}}_{n+a-\ell}. \quad (16)$$

When $a \le d$, the in-sample estimators appearing in (16) are obtained as $\widehat{\boldsymbol{\xi}}_t = \mathbf{X}_t - \widehat{\boldsymbol{\chi}}_t$, $n + a - d \le t \le n$, with either $\widehat{\boldsymbol{\chi}}_t^{\text{res}}$ or $\widehat{\boldsymbol{\chi}}_t^{\text{unr}}$ as $\widehat{\boldsymbol{\chi}}_t$.

4.2. Theoretical Properties

Proposition 4.1 establishes the consistency of $\widehat{\boldsymbol{\chi}}_{n+a|n}^{\text{res}}$ in estimating the best linear predictor of $\boldsymbol{\chi}_{n+a}$, where we make it explicit the effects of the presence of weak factors, both dynamic (as measured by $\mu_{\chi,j}(\omega)$ in Assumption 2.1) and static (as measured by $\mu_{\chi,j}$ in Assumption 4.1 (ii)), and the tail behavior (through ψ_n and $\vartheta_{n,p}$ defined in (10)).

Proposition 4.1. Suppose that the conditions in Theorem 3.1 are met and, in addition, we assume that Assumption 4.1 holds. Then, for any finite $a \ge 0$, we have

$$\left| \widehat{\boldsymbol{\chi}}_{n+a|n}^{\text{res}} - \boldsymbol{\chi}_{n+a|n} \right|_\infty$$
$$= O_P \left(p^{4-2\rho_q - 2\varrho_r} \left(\psi_n \vee p^{\varrho_r - 1} \vartheta_{n,p} \vee \frac{1}{m} \vee \frac{1}{\sqrt{p}} \right) \right).$$

As noted in Remark 3.1 (c), weaker factors and heavier tails impose a stronger requirement on the dimensionality p. If all factors are strong ($\varrho_r = 1$), the rate becomes $(\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2})$. When $a = 0$, Proposition 4.1 provides in-sample estimation consistency for any given $t \le n$. The next proposition accounts for the irreducible error in $\boldsymbol{\chi}_{n+a|n}$, with which we conclude the analysis of the forecasting error $|\widehat{\boldsymbol{\chi}}_{n+a|n}^{\text{res}} - \boldsymbol{\chi}_{n+a}|_\infty$ when $a \ge 1$.

Proposition 4.2. Suppose that Assumptions 2.2 and 2.4 hold. Then for any finite $a \ge 1$, $|\boldsymbol{\chi}_{n+a|n} - \boldsymbol{\chi}_{n+a}|_\infty = O_P(q^{1/\nu} \mu_\nu^{1/\nu} \log^{1/2}(p))$.

Recall the definition of s_{in} given in Section 3.5. The next proposition investigates the performance of $\widehat{\boldsymbol{\xi}}_{n+a|n}$ when $a = 1$, which can easily be extended to any finite $a \ge 2$.

Proposition 4.3. Suppose that the in-sample estimator of $\boldsymbol{\xi}_t$ and $\widehat{\boldsymbol{\beta}}$ satisfy

$$\left| \widehat{\boldsymbol{\xi}}_{n+1-\ell} - \boldsymbol{\xi}_{n+1-\ell} \right|_\infty = O_P \left(\bar{\zeta}_{n,p} \right) \text{ for } 1 \le \ell \le d \text{ and}$$
$$\left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_1 = O_P(s_{\text{in}} \zeta_{n,p}). \quad (17)$$

Also, let Assumptions 2.3 and 2.4 hold. Then,

$$\left| \widehat{\boldsymbol{\xi}}_{n+1|n} - \boldsymbol{\xi}_{n+1} \right|_\infty = O_P \left(s_{\text{in}} \zeta_{n,p} \left(\log^{1/2}(p) p^{1/\nu} \mu_\nu^{1/\nu} + \bar{\zeta}_{n,p} \right) \right.$$
$$\left. + \|\boldsymbol{\beta}\|_1 \bar{\zeta}_{n,p} + p^{1/\nu} \mu_\nu^{1/\nu} \right).$$

Either of the in-sample estimators $\widehat{\boldsymbol{\chi}}_t^{\text{res}}$ (described in Section 4.1) or $\widehat{\boldsymbol{\chi}}_t^{\text{unr}}$ (Appendix B), can be used in place of $\widehat{\boldsymbol{\chi}}_t$. Accordingly, the rate $\zeta_{n,p}$ in (17) is inherited by that of $\widehat{\boldsymbol{\chi}}_t^{\text{res}}$ (given in Proposition 4.1) or $\widehat{\boldsymbol{\chi}}_t^{\text{unr}}$ (Proposition B.2 (iii)). From the proof of Proposition 3.3, we have $\zeta_{n,p} \asymp (\|\boldsymbol{\beta}\|_1 + 1)(\vartheta_{n,p} \vee m^{-1} \vee p^{-1/2})$ in (17).

5. Numerical Studies

5.1. Tuning Parameter Selection

We briefly discuss the choice of the tuning parameters for FNETS. For full details, see Owens, Cho, and Barigozzi (2023) that accompanies its R implementation available on CRAN (Barigozzi, Cho, and Owens 2023).

Related to $\boldsymbol{\chi}_t$. We set the kernel bandwidth at $m = \lfloor 4(n/\log(n))^{1/3} \rfloor$ based on the case when sufficiently large number of moments exist and $n \asymp p$ (Remark 3.1 (b)). In simulation studies reported in Appendix E, we treat the number of factors q (required for Step 1 of FNETS) known, and also treat the number of static factors r (for generating the forecast) as known if it is finite; when $\boldsymbol{\chi}_t$ does not admit a static factor model (i.e. $r = \infty$), we use the value returned by the ratio-based estimator of Ahn and Horenstein (2013). In real data analysis reported in Section 5.3, we estimate both q and r, the former with the estimator proposed in Hallin and Liška (2007), the latter as in Ahn and Horenstein (2013).

Related to $\boldsymbol{\xi}_t$. We select the tuning parameter λ in (8) jointly with the VAR order d, by adopting cross validation (CV); in time series settings, a similar approach is explored in Wang and Tsay (2022). For this, the data is partitioned into M consecutive folds with indices $\mathcal{I}_l = \{n_l + 1, \ldots, n_{l+1}\}$ where $n_l = \min(l \lceil n/M \rceil, n)$, $0 \le l \le M$, and each fold is split into $\mathcal{I}_l^{\text{train}} = \{n_l + 1, \ldots, \lceil (n_l + n_{l+1})/2 \rceil\}$ and $\mathcal{I}_l^{\text{test}} = \mathcal{I}_l \setminus \mathcal{I}_l^{\text{train}}$. Then with $\widehat{\boldsymbol{\beta}}_l^{\text{train}}(\mu, b)$ obtained from $\{\mathbf{X}_t, t \in \mathcal{I}_l^{\text{train}}\}$ with the tuning parameter μ and the VAR order b, we evaluate

$$\text{CV}(\mu, b) = \sum_{l=1}^{M} \text{tr} \left(\widehat{\boldsymbol{\Gamma}}_{\xi,l}^{\text{test}}(0) - (\widehat{\boldsymbol{\beta}}_l^{\text{train}}(\mu, b))^\top \widehat{\mathbb{g}}_l^{\text{test}}(b) \right.$$
$$- (\widehat{\mathbb{g}}_l^{\text{test}}(b))^\top \widehat{\boldsymbol{\beta}}_l^{\text{train}}(\mu, b)$$
$$\left. + (\widehat{\boldsymbol{\beta}}_l^{\text{train}}(\mu, b))^\top \widehat{\mathbb{G}}_l^{\text{test}}(b) \widehat{\boldsymbol{\beta}}_l^{\text{train}}(\mu, b) \right),$$

where $\widehat{\boldsymbol{\Gamma}}_{\xi,l}^{\text{test}}(\ell)$, $\widehat{\mathbb{G}}_l^{\text{test}}(b)$, and $\widehat{\mathbb{g}}_l^{\text{test}}(b)$ are generated analogously as $\widehat{\boldsymbol{\Gamma}}_\xi(\ell)$, $\widehat{\mathbb{G}}$, and $\widehat{\mathbb{g}}$, respectively, using the test set $\{\mathbf{X}_t, t \in \mathcal{I}_l^{\text{test}}\}$. The measure $\text{CV}(\mu, b)$ approximates the prediction error while accounting for that we do not directly observe $\boldsymbol{\xi}_t$. Minimizing it over varying μ and b, we select λ and d. In simulation studies, we treat d as known while in real data analysis, we select it from the set $\{1, \ldots, 5\}$ via CV. For selecting η in (9), we adopt the Burg matrix divergence-based CV measure:

$$\text{CV}(\mu) = \sum_{l=1}^{M} \text{tr} \left(\widehat{\boldsymbol{\Delta}}_l^{\text{train}}(\mu) \widehat{\boldsymbol{\Gamma}}_l^{\text{test}} \right) - \log \left| \widehat{\boldsymbol{\Delta}}_l^{\text{train}}(\mu) \widehat{\boldsymbol{\Gamma}}_l^{\text{test}} \right| - p.$$

For both CV procedures, we set $M = 1$ in the numerical results reported below. In simulation studies, we compare the

estimators with their thresholded counterparts in estimating the network edge sets with the thresholds $\mathfrak{t}$, $\mathfrak{t}_\delta$ and $\mathfrak{t}_\omega$ selected according to a data-driven approach motivated by Liu, Zhang, and Liu (2021). Details are in Appendix D.

5.2. Simulations

In Appendix E, we investigate the estimation and forecasting performance of FNETS ondatasets simulated under a variety of settings, from Gaussian innovations $\mathbf{u}_t$ and $\boldsymbol{\varepsilon}_t$ with (E1) $\boldsymbol{\Delta} = \mathbf{I}$ and (E2) $\boldsymbol{\Delta} \neq \mathbf{I}$, to (E3) heavy-tailed ($t_5$) innovations with $\boldsymbol{\Delta} = \mathbf{I}$, and when $\boldsymbol{\chi}_t$ is generated from (C1) fully dynamic or (C2) static factor models. In addition, we consider the "oracle" setting (C0) $\boldsymbol{\chi}_t = \mathbf{0}$ where, in the absence of the factor-driven component, the results obtained can serve as a benchmark. For comparison, we consider the factor-adjusted regression method of Fan, Masini, and Medeiros (2021) and present the performance of their estimator of VAR parameters and forecasts.

5.3. Application to a Panel of Volatility Measures

We investigate the interconnectedness in a panel of volatility measures and evaluate its out-of-sample forecasting performance using FNETS. For this purpose, we consider a panel of $p = 46$ stock prices retrieved from the Wharton Research Data Service, of US companies which are all classified as "financials" according to the Global Industry Classification Standard; a list of company names and industry groups are found in Appendix G. The dataset spans the period between January 3, 2000 and December 31, 2012 (3267 trading days). Following Diebold and Yılmaz (2014), we 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} denote, respectively, the maximum and the minimum log-price of stock i on day t, and set $X_{it} = \log(\sigma_{it}^2)$; Brownlees and Gallo (2010) support this choice of volatility measure over more sophisticated alternatives.

5.3.1. Network Analysis

We focus on the period 03/2006–02/2010 corresponding to the Great Financial Crisis. We partition the data into four segments of length $n = 252$ each (corresponding to the number of trading days in a single year) and on each segment, we apply FNETS to estimate the three networks $\mathcal{N}^{\text{G}}$, $\mathcal{N}^{\text{C}}$, and $\mathcal{N}^{\text{L}}$ described in Section 3.1.

Each row of Figure 2 plots the heat maps of the matrices underlying the three networks of interest. From all four segments, the CV-based approach described in Section 5.1 returns $d = 1$ from the candidate VAR order set $\{1, \ldots, 5\}$. Hence in each row, the left panel represents the estimator $\widehat{\mathbf{A}}_1 = \widehat{\boldsymbol{\beta}}^\top$, and the middle and the right show the (long-run) partial correlations from the corresponding $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{\Omega}}$ (with their diagonals set to be zero). Locations of the nonzero elements estimate the edge sets of the corresponding networks, and the hues represent the (signed) edge weights.

Prior to March 2007, all networks exhibit a low degree of interconnectedness but the number of edges increases considerably in 03/2007–02/2008 due mainly to an overall increase in dynamic co-dependencies and a prominent role of banks (blue

Table 1. Mean, median and standard errors of $\text{FE}_{T+1}^{\text{avg}}$ and $\text{FE}_{T+1}^{\text{max}}$ on the trading days in 2012 for $\widehat{X}_{T+1|T}(n)$ in comparison with AR and FarmPredict (Fan, Masini, and Medeiros 2021) forecasts.

		FNETS			
		$\widehat{\chi}_{T+1\|T}^{\text{res}}(n)$	$\widehat{\chi}_{T+1\|T}^{\text{unr}}(n)$	AR	FarmPredict
FE^{avg}	Mean	**0.7258**	0.7466	0.7572	0.7616
	Median	**0.6029**	0.6412	0.6511	0.6243
	SE	0.4929	**0.3748**	0.4162	0.4946
FE^{max}	Mean	**0.8433**	0.8729	0.879	0.8745
	Median	**0.7925**	0.8088	0.8437	0.8259
	SE	0.2331	0.2246	**0.2169**	0.2337

NOTE: The smallest entry in each row is highlighted in bold.

group) not only in $\mathcal{N}^{\text{G}}$ but also in $\mathcal{N}^{\text{C}}$. In 03/2008–02/2009, the companies belonging to the insurance sector (red group) play a central role and in 03/2009–02/2010, the companies become highly interconnected with two particular firms having many outgoing edges in $\mathcal{N}^{\text{G}}$. Also, while most edges in $\mathcal{N}^{\text{L}}$, which captures the overall long-run dependence, have positive weights across time and companies, their weights become negative in this last segment. We highlight that FNETS is able to capture the aforementioned group-specific activities although this information is not supplied to the estimation method.

5.3.2. Forecasting

We perform a rolling window-based forecasting exercise on the trading days in 2012. Starting from $T = 3016$ (the first trading day in 2012), we forecast $\mathbf{X}_{T+1}$ as $\widehat{\mathbf{X}}_{T+1|T}(n) = \widehat{\boldsymbol{\chi}}_{T+1|T}(n) + \widehat{\boldsymbol{\xi}}_{T+1|T}(n)$, where $\widehat{\boldsymbol{\chi}}_{T+1|T}(n)$ (resp. $\widehat{\boldsymbol{\xi}}_{T+1|T}(n)$) denotes the forecast of $\boldsymbol{\chi}_{T+1}$ (resp. $\boldsymbol{\xi}_{T+1}$) using the preceding n data points $\{\mathbf{X}_t, T - n + 1 \leq t \leq T\}$. We set $n = 252$. After the forecast $\widehat{\mathbf{X}}_{T+1|T}(n)$ is generated, we update $T \leftarrow T + 1$ and repeat the above procedure until $T = 3267$ (the last trading day in 2012) is reached.

For $\widehat{\boldsymbol{\chi}}_{T+1|T}(n)$, we consider the forecasting methods derived under the static factor model (Section 4.1, denoted by $\widehat{\boldsymbol{\chi}}_{T+1|T}^{\text{res}}(n)$) and unrestricted GDFM (Appendix B, $\widehat{\boldsymbol{\chi}}_{T+1|T}^{\text{unr}}(n)$). Following the analysis in Section 5.3.1, we set $d = 1$ when producing $\widehat{\boldsymbol{\xi}}_{T+1|T}(n)$. Additionally, we report the forecasting performance of FarmPredict (Fan, Masini, and Medeiros 2021), which first fits an AR model to each of the p series ("AR"), projects the residuals on their principal components, and then fits VAR models to what remains via Lasso. Combining the three steps gives the final forecast $\widehat{\mathbf{X}}_{T+1|T}^{\text{FARM}}(n)$. The forecast produced by the first step univariate AR modeling, denoted by $\widehat{\mathbf{X}}_{T+1|T}^{\text{AR}}(n)$, is also included for comparison.

We evaluate the performance of $\widehat{\mathbf{X}}_{T+1|T}$ using two measures of errors $\text{FE}_{T+1}^{\text{avg}} = |\mathbf{X}_{T+1}|_2^{-2} \cdot |\mathbf{X}_{T+1} - \widehat{\mathbf{X}}_{T+1|T}|_2^2$ and $\text{FE}_{T+1}^{\text{max}} = |\mathbf{X}_{T+1}|_\infty^{-1} \cdot |\mathbf{X}_{T+1} - \widehat{\mathbf{X}}_{T+1|T}|_\infty$, see Table 1 for the summary of the forecasting results. Among the forecasts generated by FNETS, the one based on $\widehat{\boldsymbol{\chi}}_{T+1|T}^{\text{res}}(n)$ performs the best in this exercise, which outperforms $\widehat{\mathbf{X}}_{T+1|T}^{\text{AR}}(n)$ and $\widehat{\mathbf{X}}_{T+1|T}^{\text{FARM}}(n)$ according to both FE^{avg} and FE^{max} on average. As noted in Appendix E.2.2, the forecast based on $\widehat{\boldsymbol{\chi}}_{T+1|T}^{\text{unr}}$ shows instabilities and generally is outperformed by the one based on $\widehat{\boldsymbol{\chi}}_{T+1|T}^{\text{res}}$, but nonetheless performs reasonably well. Given the high level of co-movements and persistence in the data, the good performance of FNETS is mainly attributed to the way we forecast the factor-driven
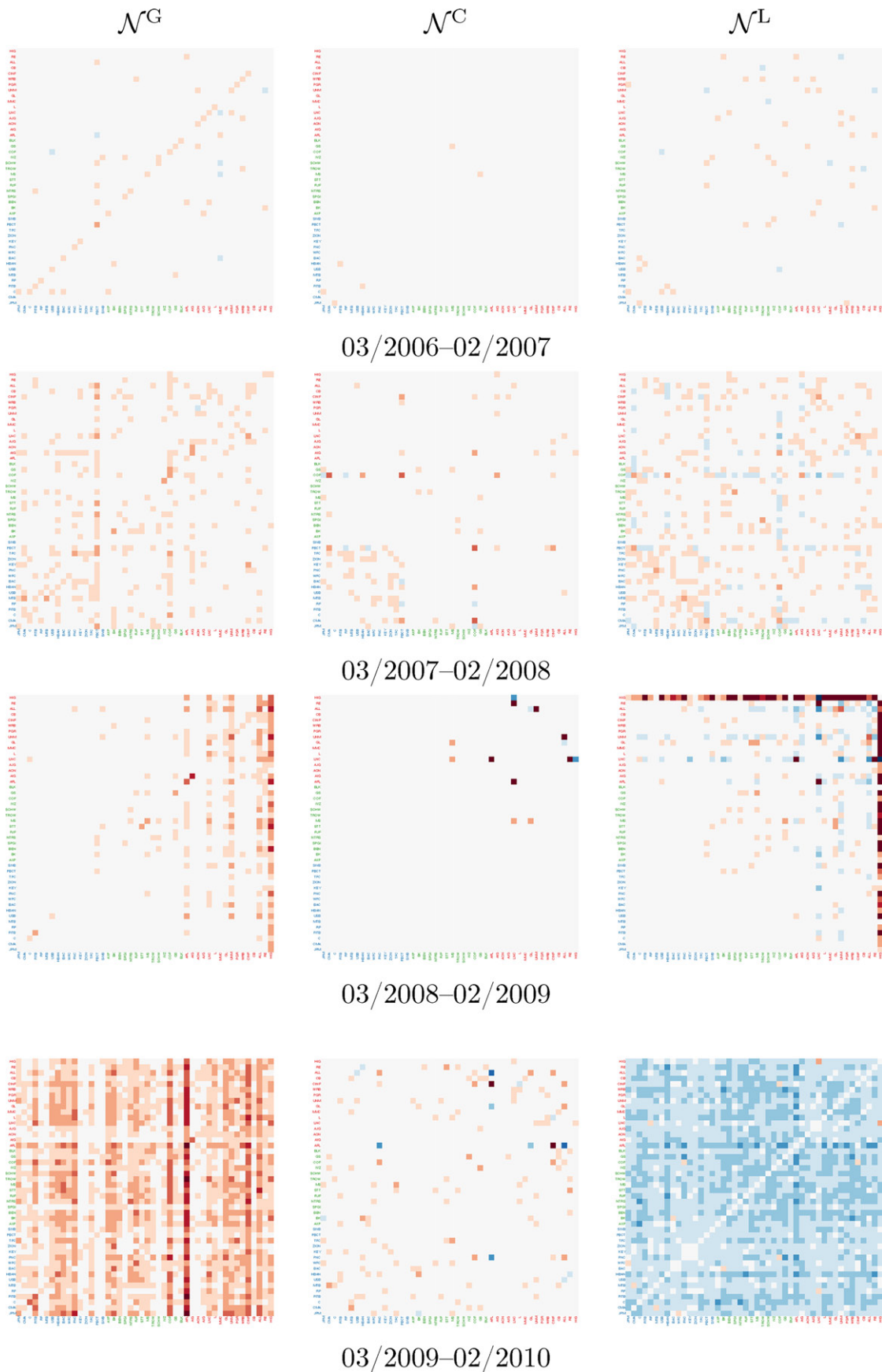
Figure 2. Heat maps of the estimators of the VAR transition matrices $\widehat{\mathbf{A}}_1$, partial correlations from $\widehat{\boldsymbol{\Delta}}$ and long-run partial correlations from $\widehat{\boldsymbol{\Omega}}$ (left to right), which in turn estimate the networks $\mathcal{N}^{\mathrm{G}}$, $\mathcal{N}^{\mathrm{C}}$, and $\mathcal{N}^{\mathrm{L}}$, respectively, over three selected periods. The grouping of the companies according to their industry classifications are indicated by the axis label colors. The heat maps in the left column are in the scale of $[-0.81, 0.81]$ while the others are in the scale of $[-1, 1]$, with red hues denoting large positive values and blue hues large negative values.

component, which is based on the estimators derived under GFDM that fully exploit all the dynamic co-dependencies (see also the results obtained by Barigozzi and Hallin 2017 on a similar dataset).

6. Conclusions

We propose and study the asymptotic properties of FNETS, a network estimation and forecasting methodology for high-dimensional time series under a dynamic factor-adjusted VAR model. Our estimation strategy fully takes into account the latency of the VAR process of interest via regularized YW estimation which, distinguished from the existing approaches, brings in methodological simplicity as well as theoretical benefits. We investigate the theoretical properties of FNETS under general conditions permitting weak factors and heavier tails than sub-Gaussianity commonly imposed in the high-dimensional VAR literature, and provide new insights into the interplay between various quantities determining the sparsity of the networks underpinning VAR processes, factor strength and tail behavior, on the estimation of those networks. Simulation studies and an application to a panel of financial time series show that FNETS is particularly useful for network analysis as it is able to discover group structures as well as producing accurate forecasts for highly co-moving and persistent time series such as log-volatilities. The R software `fnets` implementing FNETS is available from CRAN (Barigozzi, Cho, and Owens 2023).

Supplementary Materials

The supplementary materials contain the descriptions of the alternative estimation and forecasting methods, selection of the tuning parameters and the complete simulation results, in addition to all the proofs of the theoretical results.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Haeran Cho 🄳 http://orcid.org/0000-0002-0725-8704

References

Adamek, R., Smeekes, S., and Wilms, I. (2023), "Lasso Inference for High-Dimensional Time Series," *Journal of Economics*, 235, 1114–1143. [890,893]

Ahelegbey, D. F., Billio, M., and Casarin, R. (2016), "Bayesian Graphical Models for Structural Vector Autoregressive Processes," *Journal of Applied Economics*, 31, 357–386. [890]

Ahn, S. C., and Horenstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [897,898]

Anderson, B. D., and Deistler, M. (2008), "Generalized Linear Dynamic Factor Models – A Structure Theory," in *47th IEEE Conference on Decision and Control*, pp. 1980–1985. [897]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [892]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [897]

Bańbura, M., Giannone, D., and Reichlin, L. (2010), "Large Bayesian Vector Auto Regressions," *Journal of Applied Economics*, 25, 71–92. [890]

Barigozzi, M., and Brownlees, C. (2019), "NETS: Network Estimation for Time Series," *Journal of Applied Economics*, 34, 347–364. [890,892,893]

Barigozzi, M., Cho, H., and Owens, D. (2023), *fnets: Factor-Adjusted Network Estimation and Forecasting for High-Dimensional Time Series*, R package version 0.1.5. [892,898,901]

Barigozzi, M., and Hallin, M. (2017), "Generalized Dynamic Factor Models and Volatilities: Estimation and Forecasting," *Journal of Economics*, 201, 307–321. [901]

Basu, S., Li, X., and Michailidis, G. (2019), "Low Rank and Structured Modeling of High-Dimensional Vector Autoregressions," *IEEE Transactions on Signal Processing*, 67, 1207–1222. [890]

Basu, S., and Michailidis, G. (2015), "Regularized Estimation in Sparse High-Dimensional Time Series Models," *The Annals of Statistics*, 43, 1535–1567. [890,893]

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009), "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, 37, 1705–1732. [896]

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012), "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors," *Journal of Financial Economics*, 104, 535–559. [890]

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, Philadelphia, PA: SIAM. [894]

Brownlees, C., and Gallo, G. (2010), "Comparison of Volatility Measures: A Risk Management Perspective," *Journal of Financial Economics*, 8, 29–56. [899]

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, 594–607. [895,897]

Candès, E. and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation when p is much larger than n," *The Annals of Statistics*, 35, 2313–2351. [894]

Cule, E., Vineis, P., and De Iorio, M. (2011), "Significance Testing in Ridge Regression for Genetic Data," *BMC Bioinformatics*, 12, 1–15. [891]

Dahlhaus, R. (2000), "Graphical Interaction Models for Multivariate Time Series," *Metrika*, 51, 157–172. [890]

Diebold, F. X., and Yılmaz, K. (2014), "On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms," *Journal of Economics*, 182, 119–134. [892,899]

Eichler, M. (2007), "Granger Causality and Path Diagrams for Multivariate Time Series," *Journal of Economics*, 137, 334–353. [890]

Fan, J., Ke, Y., and Wang, K. (2020), "Factor-Adjusted Regularized Model Selection," *Journal of Economics*, 216, 71–85. [891,894]

Fan, J., Liao, Y., and Mincheva, M. (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society*, Series B, 75, 603–680. [892,896]

Fan, J., Lou, Z., and Yu, M. (2023), "Are Latent Factor Regression and Sparse Regression Adequate?," *Journal of the American Statistical Association* (in press). [891,894]

Fan, J., Masini, R., and Medeiros, M. C. (2021), "Bridging Factor and Sparse Models," arXiv preprint arXiv:2102.11341 . [891,894,897,899]

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. [891,892,894]

——— (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840. [891,897]

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017), "Dynamic Factor Models with Infinite-Dimensional Factor Space: Asymptotic Analysis," *Journal of Economics*, 199, 74–92. [891,893]

Forni, M., and Lippi, M. (2001), "The Generalized Dynamic Factor Model: Representation Theory," *Economic Theory*, 17, 1113–1141. [892,893]

Freyaldenhoven, S. (2021), "Factor Models with Local Factors – Determining the Number of Relevant Factors," *Journal of Economics*, 229, 80–102. [893]

Giannone, D., Lenza, M., and Primiceri, G. E. (2021), "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437. [890]

Guðmundsson, G. S., and Brownlees, C. (2021), "Detecting Groups in Large Vector Autoregressions," *Journal of Economics*, 225, 2–26. [890]

Hallin, M., and Liška, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617. [894,898]

Han, F., Lu, H., and Liu, H. (2015), "A Direct Estimation of High Dimensional Stationary Vector Autoregressions," *Journal of Machine Learning Research*, 16, 3115–3150. [890]

Hsu, N.-J., Hung, H.-L., and Chang, Y.-M. (2008), "Subset Selection for Vector Autoregressive Processes Using Lasso," *Computational Statistics & Data Analysis*, 52, 3645–3657. [890]

Kock, A. B., and Callot, L. (2015), "Oracle Inequalities for High Dimensional Vector Autoregressions," *Journal of Economics*, 186, 325–344. [890,893]

Krampe, J., and Margaritella, L. (2021), "Dynamic Factor Models with Sparse VAR Idiosyncratic Components," arXiv preprint arXiv:2112.07149. [891,894]

Krampe, J., and Paparoditis, E. (2021), "Sparsity Concepts and Estimation Procedures for High-Dimensional Vector Autoregressive Models," *Journal of Time Series Analysis*, 42, 554–579. [890,893]

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. [890]

Liu, B., Zhang, X., and Liu, Y. (2021), "Simultaneous Change Point Inference and Structure Recovery for High Dimensional Gaussian Graphical Models," *Journal of Machine Learning Research*, 22, 1–62. [899]

Loh, P.-L., and Wainwright, M. J. (2012), "High-Dimensional Regression with Noisy and Missing Data: Provable Guarantees with Nonconvexity," *The Annals of Statistics*, 40, 1637–1664. [896]

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Berlin: Springer. [893]

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2022), "Regularized Estimation of High-Dimensional Vector Autoregressions with Weakly Dependent Innovations," *Journal of Time Series Analysis*, 43, 532–557. [890,893]

Medeiros, M. C., and Mendes, E. F. (2016), "ℓ_1-regularization of High-Dimensional Time-Series Models with Non-gaussian and Heteroskedastic Errors," *Journal of Economics*, 191, 255–271. [896]

Nicholson, W. B., Wilms, I., Bien, J., and Matteson, D. S. (2020), "High Dimensional Forecasting via Interpretable Vector Autoregression," *Journal of Machine Learning Research*, 21, 1–52. [890]

Onatski, A. (2012), "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors," *Journal of Economics*, 168, 244–258. [893]

Owens, D., Cho, H., and Barigozzi, M. (2023), "fnets: An R Package for Network Estimation and Forecasting via Factor-Adjusted VAR Modelling," *The R Journal* (to appear). [898]

Priestley, M. (1982), *Spectral Analysis and Time Series*, London: Academic Press. [895]

Stock, J. H., and Watson, M. W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [892,897]

Tardivel, P. J., and Bogdan, M. (2022), "On the Sign Recovery by Least Absolute Shrinkage and Selection Operator, Thresholded Least Absolute Shrinkage and Selection Operator, and Thresholded Basis Pursuit Denoising," *Scandinavian Journal of Statistics*, 49, 1636–1668. [896]

Uematsu, Y., and Yamagata, T. (2023), "Discovering the Network Granger Causality in Large Vector Autoregressive Models," arXiv preprint arXiv:2303.15158. [890]

van de Geer, S., Bühlmann, P., and Zhou, S. (2011), "The Adaptive and the Thresholded Lasso for Potentially Misspecified Models," *Electronic Journal of Statistics*, 5, 688–749. [896]

Wang, D., and Tsay, R. S. (2022), "Rate-Optimal Robust Estimation of High-Dimensional Vector Autoregressive Models," arXiv preprint arXiv:2107.11002. [898]

Wong, K. C., Li, Z., and Tewari, A. (2020), "Lasso Guarantees for β-mixing Heavy-Tailed Time Series," *The Annals of Statistics*, 48, 1124–1142. [893]

Wu, W.-B., and Wu, Y. N. (2016), "Performance Bounds for Parameter Estimates of High-Dimensional Linear Models with Correlated Errors," *Electronic Journal of Statistics*, 10, 352–379. [893]

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, 233–254. [891]

Zhao, P., and Yu, B. (2006), "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research*, 7, 2541–2563. [896]