Factor Network Autoregressions

Matteo Barigozzi
University of Bologna, Italy

Giuseppe Cavaliere
University of Bologna, Italy
& Exeter Business School, UK

Graziano Moramarco
University of Bologna, Italy

Abstract

We propose a factor network autoregressive (FNAR) model for time series with complex network structures. The coefficients of the model reflect many different types of connections between economic agents ("multilayer network"), which are summarized into a smaller number of network matrices ("network factors") through a novel tensor-based principal component approach. We provide consistency and asymptotic normality results for the estimation of the factors, their loadings, and the coefficients of the FNAR, as the number of layers, nodes and time points diverges to infinity. Our approach combines two different dimension-reduction techniques and can be applied to high-dimensional datasets. Simulation results show the goodness of our estimators in finite samples. In an empirical application, we use the FNAR to investigate the cross-country interdependence of GDP growth rates based on a variety of international trade and financial linkages. The model provides a rich characterization of macroeconomic network effects as well as good forecasts of GDP growth rates.

Keywords: Networks, factor models, principal components, VAR, tensor decomposition.

1 Introduction

Network models are key tools for analyzing interconnected systems and are increasingly used in many areas of research. Typical applications involve social networks, where a high number of agents are connected along several dimensions (e.g., Zhu et al., 2017), or economic networks, which are usually applied to evaluate spillover effects between agents, geographical regions and economic sectors (e.g., Acemoglu et al., 2015; Diebold and Yilmaz, 2014; Billio et al., 2012). From an empirical perspective, the analysis of networks calls for the development of novel statistical techniques that are able to account for complex interactions and handle large datasets, possibly recorded over time.

Several approaches have been proposed to integrate network structures into well-established time series models; the network autoregression (NAR) by Zhu et al. (2017) or the global vector autoregressive (GVAR) model by Pesaran et al. (2004) are two leading examples. These models are generally designed to deal with one type of connection at a time between the agents or "nodes" of a network, e.g., import/export flows. They characterize networks in the form of a single *adjacency matrix*; i.e., a matrix whose ij-th element represents a link between node i and node j.

The assumption of a single adjacency matrix, however, appears in general restrictive. In fact, modern economies or social networks are complex systems, in which agents interact through many different channels. For instance, countries are simultaneously linked not only by international trade flows, but also by financial linkages, multinational firms' activities, migration flows, etc. Accordingly, models of *multilayer networks*, i.e., networks where nodes are linked through multiple types of connections (Kivelä et al., 2014), have important applications in many different areas, such as international economics and financial (systemic) risk assessment.

To state the problem we consider in this paper, let y_t be an N-dimensional vector of stationary time series, e.g., GDP growth rates for different countries. To model the dynamics of y_t when N is large, we could consider a VAR model in which the dynamic relations between the N elements of y_t are mediated by m time-varying networks, denoted as $W_{k,t}$, $k = 1, \dots, m$, e.g., measuring trade flows of different commodities or various financial linkages. All together these networks, which are $N \times N$ matrices, form the m layers of a multilayer network. We could then consider a multilayer NAR:

$$y_t = b_1 W_{1,t-1} y_{t-1} + \ldots + b_m W_{m,t-1} y_{t-1} + \varrho y_{t-1} + a + \zeta_t, \quad t = 1, \ldots, T, \tag{1.1}$$

where b_k, $k = 1, \ldots, m$, ϱ, and a are unknown scalars, and $\zeta_t := (\zeta_{1t}, \cdots, \zeta_{Nt})'$ is an N-dimensional vector of errors. This model generalizes the NAR by Zhu et al. (2017) to the case of multilayer networks. However, in empirical applications, the number m of layers can be very large (e.g., for trade data one may get to hundreds or even thousands of product categories); hence, least-squares estimation of (1.1) can become quickly infeasible.

In this paper, we assume that the observed layers $W_{k,t}$, $k = 1, \ldots, m$, of the multilayer network have a factor structure, where the common factors are unobserved networks, each represented by an $N \times N$ matrix, capturing the commonality between the layers. We call these latent factors *network factors*; once retrieved, we can use them in (1.1) to model the dynamics of y_t in place of all the m layers, thus reducing the dimensionality of the problem. We call the resulting model a factor network autoregression (FNAR). To estimate the FNAR we first develop a novel, tensor-based, principal component analysis approach, which allows to estimate the network factors. We then use these network factors to estimate an N-dimensional FNAR.

Our model exploits two different types of dimension-reduction techniques. First, by assuming, as in Zhu et al. (2017), that the network effects are homogeneous across nodes (or across a finite number of groups of nodes), the number of unknown parameters in our FNAR is at most linear in the number of networks considered, instead of being proportional to N^2 as is the case for ordinary vector autoregressions. Second, since the FNAR is based on the extraction of just few network factors common across the m network layers, the number of parameters does not grow with the number of layers m, but with the number of network factors which we assume to be finite and independent of N and m.

In terms of theory, we first provide sufficient conditions such that the FNAR admits a (weakly) stationary and causal solution. Second, we provide sufficient conditions such that our estimators of the network factors and of the FNAR coefficients are consistent and asymptotically normal as the sample size T, the number of layers m, and the number of nodes N, diverge. Crucially, the estimator of the FNAR coefficients we propose is consistent even when the FNAR errors admit cross-sectional correlation as well as serial correlation, due to the presence of node specific factors. Finally, the number of network factors can be determined by an eigenvalue ratio approach, by adapting to the present context those recently proposed by Han et al. (2022) and Barigozzi et al. (2023).

In an empirical application, we use the FNAR model to study the dynamics of GDP growth rates in a multilayer network of $N = 24$ countries with $m = 25$ different layers, reflecting international trade flows for different categories of goods and services, and a variety of cross-border financial linkages. In a first step we retrieve 6 network factors common across the m layers, which, once identified, have a clear economic meaning. These are then used to construct the regressors which are employed in a FNAR to study and predict GDP growth rates for all countries. A similar idea was explored by Chen and Chen (2022) where, however, the tensor data of trade flows is compressed also along the node (country) dimension, thus making the interpretation of the underlying factors less clear.

The paper is organized as follows. Section 2 gives a summary of the related literature. Section 3 illustrates the model. Section 4 presents the estimation approach. Section 5 introduces the assumptions and the asymptotic properties of the estimators. Section 6 evaluates the finite sample performance of our estimators by means of Monte Carlo simulations. Section 7 presents the empirical application and Section 8 concludes. All proofs are in the online Supplementary Material, which also contains additional possible estimators, further simulation results and details on the empirical application.

NOTATION. The *order* of a tensor is the number of dimensions, also known as *modes*. The *fibers* of a given mode are defined by fixing every index but one. We deal only with order-3 tensors. The mode-q *matricization*, or unfolding, of a tensor $\mathcal{T}$, denoted as $\text{mat}_{(q)}(\mathcal{T})$ or, equivalently, $\mathcal{T}_{(q)}$, is a matrix having as columns its mode-q fibers. For a generic order-3 tensor $\mathcal{T}$ of dimensions $d_1 \times d_2 \times d_3$ with generic element $\mathcal{T}_{ijk}$, the mode-1 matricization is a $d_1 \times (d_2 d_3)$ matrix having as columns the d_1-dimensional vectors $\mathcal{T}_{\cdot jk}$, the mode-2 matricization is a $d_2 \times (d_1 d_3)$ matrix having as columns the d_2-dimensional vectors $\mathcal{T}_{i \cdot k}$, and the mode-3 matricization is a $d_3 \times (d_1 d_2)$ matrix having as columns the d_3-dimensional vectors $\mathcal{T}_{ij \cdot}$.

Finally, we denote the mode-q multiplication of tensor $\mathcal{T}$ by a matrix X of size $p \times d_q$ as $\mathcal{T} \times_q X$, which is a tensor of size p in the q-th dimension and the same size as $\mathcal{T}$ in the other dimensions.

2 Related literature

Our modelling approach is especially related to the network autoregression (NAR) model by Zhu et al. (2017), the community network autoregression (CNAR) model by Chen et al. (2023) and the group network autoregression (GNAR) model by Zhu et al. (2023). Differently from these papers, where just one network is considered, our main contribution is to integrate a large multilayer network into a VAR model by representing multilayer networks as a superposition of common network factors.

Our methodology for extracting the network factors is related to the recent statistical literature on factor analysis of tensor time series building on the tensor Tucker decomposition, which assumes the existence of low-rank tensor factors (Chen et al., 2022; Han et al., 2020, 2022; Barigozzi et al., 2023; Zhang et al., 2022; Chen and Lam, 2024). This approach differs from the canonical polyadic (CP) decomposition which always assumes the factors to be vectors (Chang et al., 2023; Han et al., 2024).

Differently from all the above cited works, since our aim is to reduce the dimensionality of a multilayer network, our setting is a special case of the Tucker decomposition, where the factors are low-rank only along the layers' mode, while they retain all information along the nodes' two modes, i.e., they are $N \times N$ matrices which we can directly interpret as network factors. Moreover, differently from some of the above cited works, we allow for serial dependence in the idiosyncratic tensor, reflecting the assumption, commonly made in the literature on factor modeling of economic data, that factors account for the cross-sectional (for us, cross-layer) variation of the data (see, e.g., Bai, 2003, in the vector case). In Appendix F.5 we provide a simulation-based comparison between our estimates of the network factor loadings and those obtained by means of the approaches by Chen et al. (2022) where no serial idiosyncratic correlation is allowed for.

Our paper is also related to a growing literature that uses tensor decomposition methods to estimate the coefficients of high-dimensional time series models. First, Wang et al. (2022) apply tensor decomposition to the order-3 tensor whose slices are the matrices of (unknown) VAR coefficients at different lags, which makes this approach an alternative to ours. However, unlike our approach, Wang et al. (2022) do not exploit data on observed networks to estimate the VAR. We refer to Appendix G.2.4 for an empirical comparison between this approach and our FNAR. Second, Wang et al. (2024) consider an autoregressive model for tensor-valued time series and use a Tucker decomposition to estimate the autoregressive coefficients, while Chang et al. (2023) model matrix-valued time series using a tensor CP decomposition. Both these last two approaches are more distant from ours, since we do not aim to estimate an AR model for a multilayer network itself, but we aim to use a multilayer network to estimate a vector autoregression in the same spirit as Zhu et al. (2017).

Finally, we also contribute to two further strands of literature. First, to the literature on factor and factor-augmented models (e.g., Stock and Watson, 2002, Bai and Ng, 2006) by developing a new framework where factors are matrices rather than vectors, and enter a factor-augmented autoregression by multiplying (weighting) the lagged vector of endogenous variables, rather than being included directly as regressors. Second, to two important streams of the literature on network econometrics, that is: (i) works that investigate the properties of observed networks, such as production networks (see, e.g., Acemoglu et al., 2012), trading networks in financial and interbank markets (Denbee et al., 2021) and social networks (Zhu et al., 2017), and (ii) works concerned with the estimation of network links from the data (e.g., Diebold and Yilmaz, 2014, Billio et al., 2012, Barigozzi and Brownlees, 2019). Our approach combines these two streams of research: on the one hand, we use data on a large number of observed economic networks; on the other hand, we estimate unobserved common network factors driving them.

3 Model

We assume to observe m networks, each represented by a matrix $W_{k,t}$, $k = 1, \dots, m$, of dimension $N \times N$. The networks have no self-loops, so the diagonal elements of the matrices are zero, and are weighted and directed, so in general the matrices have real entries and are not symmetric. In the present high-dimensional setting, it is convenient to assume that the weights are normalized in such a way that the elements in each row of $W_{k,t}$ sum to N. This can be done without loss of generality (see Section 4 for specific comments on this aspect).

We then assume that each observed network can be written as a linear combination of r network factors, $F_{k,t}$, $k = 1, \ldots, r$, each of them being an $N \times N$ matrix and with $r \ll m$, plus a network $\mathcal{E}_{t \cdot \cdot k}$ idiosyncratic to the k-th layer, which is an $N \times N$ matrix. Specifically, we assume

$$W_{k,t} = u_{k1} F_{1,t} + \ldots + u_{kr} F_{r,t} + \mathcal{E}_{t \cdot \cdot k}, \qquad k = 1, \ldots, m, \ t = 1, \ldots, T, \tag{3.1}$$

where u_{kh}, $h = 1, \ldots, r$, are the scalar factor loadings for network k.

If the factors are sufficiently "pervasive", we can think of replacing the multilayer NAR in (1.1) with a FNAR model, which is given by:

$$y_t = \beta_1 \frac{F_{1,t-1}}{N} y_{t-1} + \ldots + \beta_r \frac{F_{r,t-1}}{N} y_{t-1} + \rho y_{t-1} + \alpha + \nu_t, \quad t = 1, \ldots, T, \tag{3.2}$$

where β_k, $k = 1, \ldots, r$, ρ, and α are unknown scalars, and $\nu_t := (\nu_{1t}, \cdots, \nu_{Nt})'$ is the N-dimensional vector of FNAR errors. Notice that the network factors are rescaled by N in agreement with our normalization assumption on the observed networks.

Furthermore, in order to allow also for factors common across the N nodes, we assume each element of the FNAR errors ν_{it}, $i = 1, \ldots, N$, to have a factor structure:

$$\nu_{it} = \lambda_{i1} G_{1t} + \ldots + \lambda_{iq} G_{qt} + \epsilon_{it}, \qquad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{3.3}$$

where G_{kt}, $k = 1, \ldots, q$, are node factors, λ_{ik}, $k = 1, \ldots, q$, are the scalar factor loadings for node i, and ϵ_{it} is the node idiosyncratic component.

We call the term $\sum_{k=1}^{r} \beta_k N^{-1} F_{k,t-1} y_{t-1}$ the *network effect*, where the coefficients capture the strength of dynamic network effects between nodes, exerted through different types of relationships, which are summarized by means of few network factors. We call the term ρy_{t-1} the *momentum effect*, which captures the direct dynamic interaction of a node with itself. Last, we call the term α the *nodal effect*. This can be generalized to include a random effect by adding a term $Z_{t-1}\delta$, where δ is a K-dimensional parameter vector, common to all nodes, and Z_t is a $N \times K$ matrix of node-specific exogenous variables.

Because of the factor structure in the FNAR errors, the network nodes are correlated with each other not only through the network relationships, but also through the common factors which characterize the cross-sectional dependence at a global level.

By comparing the FNAR in (3.2) with the multilayer NAR in (1.1) and a standard VAR(1) for y_t we see that, while the latter requires estimating N^2+N parameters, the multilayer NAR requires estimating $m+2$ parameters, and the FNAR further reduces the number of parameters to $r+2$. Hence, the FNAR model provides two forms of dimension reduction, both along the layer direction, by means of the network factors in (3.1) and along the node direction by means of the network mediated interactions in (3.2).

4 Estimation

Estimation of the FNAR model requires two distinct steps. The first one involves estimation of the r network factors. The second step involves fitting the FNAR equation (3.2). Furthermore, it is necessary to determine the number of network and node factors.

Network factors. We start by discussing estimation of the network factors. To this aim, it is convenient to collect the m weight matrices or layers of the multinetwork into a *weight tensor* of order 3, which we denote as $\mathcal{W}_t$ and has size $N\times N\times m$. In our notation the m network matrices are the *frontal slices* of the tensor. The factor network structure in (3.1) is then rewritten as:

$$\mathcal{W}_t = \mathcal{F}_t \times_3 U + \mathcal{E}_t, \quad t = 1, \dots, T, \tag{4.1}$$

where $\mathcal{F}_t$ is a $N \times N \times r$ tensor containing, as frontal slices, the r *network factors* $F_{k,t}$, $k = 1, \dots, r$, each of dimensions $N\times N$, which are common across all layers, U is a $m\times r$ matrix of factor loadings, with entries u_{kh}, $k = 1, \dots, m$, $h = 1, \dots, r$, determining how each layer of the original network loads on the network factors, and $\mathcal{E}_t$ is an $N\times N\times m$ tensor containing, as frontal slices, the idiosyncratic networks, $\mathcal{E}_{t\cdot\cdot k}$, $k = 1, \dots, m$. The elements of $\mathcal{E}_t$ are allowed to be (i) weakly correlated in all three modes, and (ii) autocorrelated over time (see Section 5 for the specific assumptions). Finally, from (4.1) the mode-3 matricization of $\mathcal{W}_t$, denoted as $\mathcal{W}_{(3)t}$, is a $m \times N^2$ matrix such that:

$$\mathcal{W}_{(3)t} = U\mathcal{F}_{(3)t} + \mathcal{E}_{(3)t} \quad t = 1, \dots, T. \tag{4.2}$$

where $\mathcal{F}_{(3)t}$ is $r \times N^2$ and $\mathcal{E}_{(3)t}$ is $m \times N^2$. This expression resembles a conventional factor model, with the major difference that each of the r factors is no longer a scalar but a vector of size N^2, containing up to $N(N-1)$ non-zero elements (as there are no self-interactions in the network).

To estimate $\mathcal{F}_t$, or, equivalently, $\mathcal{F}_{(3)t}$, first, we compute the sample $m \times m$ outer product of $\mathcal{W}_{(3)t}$:

$$\widehat{\Gamma}^{\mathcal{W}} := \frac{1}{T}\sum_{t=1}^{T} \mathcal{W}_{(3)t}\mathcal{W}'_{(3)t}. \tag{4.3}$$

Second, letting $\widehat{V}^{\mathcal{W}}$ be the $m \times r$ matrix whose j-th column is the normalized eigenvector corresponding to the j-th largest eigenvalue, $\widehat{\mu}_j^{\mathcal{W}}$, of $\widehat{\Gamma}^{\mathcal{W}}$, and $\widehat{M}^{\mathcal{W}}$ be the $r \times r$ diagonal matrix with $\widehat{\mu}_j^{\mathcal{W}}$ as its j-th diagonal entry, we estimate the $m \times r$ loadings matrix U as:

$$\widehat{U} := \frac{1}{N}\widehat{V}^{\mathcal{W}}(\widehat{M}^{\mathcal{W}})^{1/2}. \tag{4.4}$$

Third, we estimate the mode-3 matricization of the network factors $\mathcal{F}_{(3)t}$ as the principal components (PCs) of $\mathcal{W}_{(3)t}$, i.e., by linear projection of $\mathcal{W}_{(3)t}$ onto $\widehat{U}$: $\widehat{\mathcal{F}}_{(3)t} := (\widehat{U}'\widehat{U})^{-1}\widehat{U}'\mathcal{W}_{(3)t} = N(\widehat{M}^{\mathcal{W}})^{-1/2}\widehat{V}^{\mathcal{W}'}\mathcal{W}_{(3)t}$, which is an $r \times N^2$ matrix; once folded into an order-3 tensor, it gives the estimated $N \times N \times r$ tensor of network factors:

$$\widehat{\mathcal{F}}_t := \mathcal{W}_t \times_3 [(\widehat{U}'\widehat{U})^{-1}\widehat{U}'] = \mathcal{W}_t \times_3 [N(\widehat{M}^{\mathcal{W}})^{-1/2}\widehat{V}^{\mathcal{W}'}], \tag{4.5}$$

having as layers $\widehat{F}_{k,t} := \text{mat}_{(1)}(\widehat{\mathcal{F}}_{\cdot,\cdot,k,t})$, $k = 1, \dots, r$, which are the $N \times N$ matrices of estimated network factors.

Like in ordinary principal component analysis (PCA), the key intuition is that by exploiting the cross-sectional variation we can estimate the space spanned by the factors. In this case, the relevant cross-sectional dimension is the dimension m of the layers in a network; i.e., the third dimension of the weight tensor $\mathcal{W}_t$. The rescaling by N in estimating the loadings reflects the fact that no dimension reduction is applied to the first and second modes of $\mathcal{W}_t$.

Notice that the estimated loadings and factor tensor are such that they satisfy the identifying conditions: $m^{-1}\widehat{U}'\widehat{U} = m^{-1}N^{-2}\widehat{M}^{\mathcal{W}}$, which is a diagonal matrix by construction, and

$$\frac{1}{N^2T}\sum_{t=1}^{T}\widehat{\mathcal{F}}_{(3)t}\widehat{\mathcal{F}}'_{(3)t} = (\widehat{M}^{\mathcal{W}})^{-1/2}\widehat{V}^{\mathcal{W}\prime}\widehat{\Gamma}^{\mathcal{W}}\widehat{V}^{\mathcal{W}}(\widehat{M}^{\mathcal{W}})^{-1/2} = I_r, \tag{4.6}$$

so that the r rows of $\widehat{\mathcal{F}}_{(3)t}$ are the classical normalized PCs of the m rows of $\mathcal{W}_{(3)t}$.

There are two main differences between our approach and the one proposed by Chen et al. (2022) for estimating common tensor factors from tensor times series admitting a Tucker decomposition. First, we estimate factors using only contemporaneous sample second moments, thus allowing the idiosyncratic tensor to be also autocorrelated. The second difference is that we extract factors along a single dimension of the tensor, namely the dimension of the network layers. This implies that the extracted factors still have a network interpretation; i.e., they are common network factors. Indeed, by construction, the estimated factor matrices $\widehat{F}_{k,t}$, $k = 1, \dots, r$, have zeros along the main diagonal, as the observed network matrices, and have a scale fixed by means of (4.6).

To preserve the network properties of the factors, we do not standardize nor demean the elements in $\mathcal{W}_t$ along the time dimension. Indeed, our variables are all expressed in the same unit of measurement and standardization would eliminate the fundamental interpretation of $\mathcal{W}_t$ as a network, as centering would affect the zero diagonal entries. The same strategy is adopted by Chen et al. (2022, Remark 11).

Our estimators in (4.4) and (4.5) are defined consistently with the assumption that the observed networks have rows summing to N. This implies that the estimated network factors in (4.5) have variance growing with N, as shown in (4.6), and they must be rescaled before being used in the FNAR defined in (3.2) to ensure that the scale of the estimated network coefficients β_k does not depend on N. Clearly, we could equivalently work with row-normalized observed networks and then no rescaling by N would be needed anywhere, although this would imply that, as the number of nodes N grows, the entries of $\mathcal{W}_t$ would have to become smaller and smaller.

FNAR COEFFICIENTS. To describe the estimation of the FNAR, it is convenient to introduce some further notation. Hereafter, with reference to (3.2), let $y := (y_1, \cdots, y_T)' = (y_1, \cdots, y_N)$ be the $T \times N$ matrix of observed data (with y_t, $t = 1, \dots, T$, being N-dimensional, and y_i, $i = 1, \dots, N$, being

T-dimensional), and define also the r-dimensional vector $\beta := (\beta_1, \cdots, \beta_r)'$ and the $(r+2)$-dimensional vector $\theta := (\beta', \rho, \alpha)'$ of the FNAR coefficients. Moreover, with reference to the node factor structure in (3.3), we let $\epsilon := (\epsilon_1, \cdots, \epsilon_T)' = (\epsilon_1, \cdots, \epsilon_N)$ be the $T \times N$ matrix of node idiosyncratic components (with ϵ_t, $t = 1, \ldots, T$, being N-dimensional, and ϵ_i, $i = 1, \ldots, N$, being T-dimensional), and we let $G_t := (G_{1t}, \cdots, G_{qt})'$ and $\Lambda_i = (\lambda_{i1}, \cdots, \lambda_{iq})'$, so that $G := (G_1, \cdots, G_T)'$ is the $T \times q$ matrix of node factors, and $\Lambda := (\Lambda_1, \cdots, \Lambda_N)'$ is the $N \times q$ matrix of loadings. Finally, define the order-3 tensor $\mathcal{X}$ of dimensions $T \times N \times (r+2)$ having in each of the first r layers one of the r matrices $\mathsf{F}_k := (N^{-1}F_{k,0}y_0, \cdots, N^{-1}F_{k,T-1}y_{T-1})'$, $k = 1, \ldots, r$, each of size $T \times N$, in layer $(r+1)$ the $T \times N$ matrix $y_{(-1)} := (y_0, \cdots, y_{T-1})'$, and in layer $(r+2)$ a $T \times N$ matrix of ones, denoted as $1_{T\times N}$. It follows that $\mathcal{X}_{(1)} := \text{mat}_{(1)}(\mathcal{X})$ is a $T \times N(r+2)$ matrix having as t-th row $\text{vec}(X_t)'$ and $\mathcal{X}_{(2)} := \text{mat}_{(2)}(\mathcal{X})$ is a $N \times T(r+2)$ matrix having as i-th row $\text{vec}(X_i)'$, where we define $X_t := (\mathsf{F}'_{1,t-1,\cdot}, \cdots, \mathsf{F}'_{r,t-1,\cdot}, y_{t-1}, \iota_N)$ and $X_i := (\mathsf{F}_{1,\cdot,i}, \cdots, \mathsf{F}_{r,\cdot,i}, y_i, \iota_T)$, with ι_N and ι_T being N- and T-dimensional vectors of ones, respectively, with $\mathsf{F}_{k,t-1,\cdot}$ and $\mathsf{F}_{k,\cdot,i}$ being the tth row and ith column of F_k, respectively.

According to the above notation, the FNAR in (3.2), jointly with the node factor structure in its errors given in (3.3), can be equivalently rewritten as:

$$y_t = X_t\theta + \Lambda G_t + \epsilon_t, \quad t = 1, \ldots, T, \quad \text{or} \quad y_i = X_i\theta + G\Lambda_i + \epsilon_i, \quad i = 1, \ldots, N, \tag{4.7}$$

which, by stacking all T or N equations, can also be written as:

$$y = \text{mat}_{(1)}(\mathcal{X} \times_3 \theta') + G\Lambda' + \epsilon = \mathcal{X}_{(1)}(\theta \otimes I_N) + G\Lambda' + \epsilon = (\theta' \otimes I_T)\mathcal{X}'_{(2)} + G\Lambda' + \epsilon. \tag{4.8}$$

Since both the network factors (contained in $\mathcal{X}$) and the node factors are unobserved, estimation of (4.8) is infeasible. To make estimation feasible, we start by substituting $\mathcal{X}$ with $\widehat{\mathcal{X}}$, which, in turn, is obtained by replacing the network factors $F_{k,t}$ with their estimates $\widehat{F}_{k,t}$, $k = 1, \ldots, r$, given by (4.5).

An infeasible OLS estimator of θ would then be obtained by applying the Frisch-Waugh theorem to partial out the effect of either G or Λ:

$$\widehat{\theta}^{\text{FW,G}} \otimes I_N = \left(\widehat{\mathcal{X}}'_{(1)} M_G \widehat{\mathcal{X}}_{(1)}\right)^{-1} \left(\widehat{\mathcal{X}}'_{(1)} M_G\, y\right) \quad \text{or} \quad \widehat{\theta}^{\text{FW},\Lambda} \otimes I_T = \left(\widehat{\mathcal{X}}'_{(2)} M_\Lambda \widehat{\mathcal{X}}_{(2)}\right)^{-1} \left(\widehat{\mathcal{X}}'_{(2)} M_\Lambda\, y'\right),$$

where $M_G := I_T - G(G'G)^{-1}G'$ and $M_\Lambda := I_N - \Lambda(\Lambda'\Lambda)^{-1}\Lambda'$ are the $T \times T$ and $N \times N$ linear projectors onto the spaces orthogonal to the factors and loadings spaces, respectively. Likewise, given θ, the estimators of Λ and G would be the usual PC estimators applied to the FNAR residuals $\widehat{\nu} := y - \text{mat}_{(1)}(\widehat{\mathcal{X}} \times_3 \theta')$.

Following this reasoning, two equivalent estimators of θ are given by:

$$\begin{aligned} \widehat{\theta}^\dagger &:= \left(\sum_{i=1}^N \widehat{X}_i' M_{\widehat{G}^\dagger} \widehat{X}_i\right)^{-1} \left(\sum_{i=1}^N \widehat{X}_i' M_{\widehat{G}^\dagger} y_i\right), \\ \widehat{\theta}^* &:= \left(\sum_{t=1}^T \widehat{X}_t' M_{\widehat{\Lambda}^*} \widehat{X}_t\right)^{-1} \left(\sum_{t=1}^T \widehat{X}_t' M_{\widehat{\Lambda}^*} y_t\right), \end{aligned} \tag{4.9}$$

where, letting $\widehat{\nu}^\dagger := y - \text{mat}_{(1)}(\widehat{\mathcal{X}} \times_3 \widehat{\theta}^{\dagger'})$ and $\widehat{\nu}^* := y - \text{mat}_{(1)}(\widehat{\mathcal{X}} \times_3 \widehat{\theta}^{*'})$, we defined (see also Appendix A)

$$\widehat{G}^\dagger := \widehat{V}^{\widehat{\nu}^\dagger}\sqrt{T}, \quad \widehat{\Lambda}^* := \widehat{V}^{\widehat{\nu}^*}(\widehat{M}^{\widehat{\nu}^*})^{1/2}, \tag{4.10}$$

with $\widehat{V}^{\widehat{\nu}^\dagger}$ being the $T \times q$ matrix of normalized eigenvectors of $N^{-1}\widehat{\nu}^\dagger\widehat{\nu}^{\dagger'}$, and $\widehat{M}^{\widehat{\nu}^*}$ being the $q \times q$ diagonal matrix of eigenvalues of $T^{-1}\widehat{\nu}^{*'}\widehat{\nu}^*$ with corresponding normalized eigenvectors given by the columns of the $N \times q$ matrix $\widehat{V}^{\widehat{\nu}^*}$. By iterating between (4.9) and (4.10), we solve such minimization and compute the two final estimators of θ. As explained at the end of Section 5.3, the two are asymptotically equivalent and they might differ numerically just because of the iterative approaches used to compute them.

Finally, notice that, obviously, once we have computed $\widehat{\theta}^\dagger$ we can also estimate the loadings Λ by linear projection of $\widehat{G}^\dagger$ onto $\widehat{\nu}^\dagger$, and once we have computed $\widehat{\theta}^*$ we can also estimate the node factors G by linear projection of $\widehat{\Lambda}^*$ onto $\widehat{\nu}^*$. These estimates, however, are not needed for estimating θ.

The outlined estimators are robust to the presence of autocorrelated node factors G_t, and, as a consequence, the FNAR errors ν_t are allowed to be both cross-sectionally and serially correlated. The adopted algorithm is similar to the one proposed by Chen et al. (2023), who in turn adapted the approach by Bai (2009) to the NAR setting. Under the assumption of no autocorrelation in the node factors, the OLS and GLS estimators are also valid estimators of θ (see Appendices B and C, respectively).

Number of factors. Letting $\widehat{\mu}_j^{\mathcal{W}}$, $j = 1, \ldots, m$, be the j-th largest eigenvalue of $\widehat{\Gamma}^{\mathcal{W}}$ defined in (4.3), we estimate the number of network factors, r, by means of the eigenvalue ratio criterion:

$$\widehat{r} := \underset{j=1,\ldots,r_{\max}}{\operatorname{argmax}} (\widehat{\mu}_j^{\mathcal{W}}/\widehat{\mu}_{j+1}^{\mathcal{W}}), \tag{4.11}$$

where $r_{\max}$ is a predefined maximum number of network factors such that $r_{\max} < \min\{m, T, N^2\}$. This is the criterion proposed by Barigozzi et al. (2023), which generalizes the approach proposed by Han et al. (2022) to the tensor factor model with autocorrelated idiosyncratic components.

Likewise, letting $\widehat{\mu}_j^{\widehat{\nu}}$, $j = 1, \ldots, N$, be the j-th largest eigenvalue of the sample covariance matrix of the FNAR residuals, the number of node factors, q, is determined by means of the criterion:

$$\widehat{q} := \operatorname*{argmax}_{j=1,\ldots,q_{\max}} (\widehat{\mu}_j^{\widehat{\nu}} / \widehat{\mu}_{j+1}^{\widehat{\nu}}), \tag{4.12}$$

where $q_{\max}$ is a predefined maximum number of node factors such that $q_{\max} < \min\{T, N\}$; see Ahn and Horenstein (2013). In practice, since the FNAR residuals $\widehat{\nu}_t$ depend on the chosen value of q, we can adopt an iterative procedure, which starts by over-estimating q in the first stage, identical to the one described by Bai (2009, Section C.3 in the Supplementary Material).

Alternative approaches to estimating r and q are possible; see, e.g., the information criterion by Bai and Ng (2002), or the randomized test by Trapani (2018) which is based on the divergence rates of the eigenvalues. Importantly, the latter could also be used to test for no factors.

5 Theory

5.1 Assumptions

In the following, we allow for the number of layers m to grow to infinity. Therefore, all our assumptions are stated for the infinite sequence of $N \times N$ networks $W_{i,t} := \text{mat}_{(1)}(\mathcal{W}_{\cdot,\cdot,i,t})$ with $i \in \mathbb{N}$. Equivalently, we could state the assumptions for $i = 1, \ldots, m$ with $m \in \mathbb{N}$. Moreover, all assumptions are stated contemplating the possibility that also the number of nodes N and the sample size T grow to infinity.

Assumption 1 (Common component of multilayer network).

(i) $\lim_{m\to\infty} \|m^{-1}U'U - \Sigma_U\| = 0$ *where* Σ_U *is* $r \times r$ *finite and positive definite, and, for all* $k \in \mathbb{N}$, $\|U_{k\cdot}\| \leq M_U$ *for some finite* M_U *independent of* k.

(ii) For all $t \in \mathbb{Z}$ *and all* $N \in \mathbb{N}$, $\Gamma^{\mathcal{F}} := \mathbb{E}[\mathcal{F}_{(3)t}\mathcal{F}'_{(3)t}]$ *is* $r \times r$ *positive definite, and such that* $\left\|N^{-2}\Gamma^{\mathcal{F}}\right\| \leq M_{\mathcal{F}}$ *for some finite* $M_{\mathcal{F}}$ *independent of* N.

(iii) *For all $N \in \mathbb{N}$ and all $t \in \mathbb{Z}$, $\mathbb{E}\left[\left\|N^{-1}\mathcal{F}_{(3)t}\right\|^4\right] \leq K_{\mathcal{F}}$ for some finite $K_{\mathcal{F}}$ independent of t and N.*

(iv) *For all $i, j = 1, \ldots, r$ and all $T, N \in \mathbb{N}$, $\mathbb{E}\left[\left|\frac{1}{\sqrt{TN}}\sum_{t=1}^T \left\{\mathcal{F}_{(3)ti\cdot}\mathcal{F}'_{(3)tj\cdot} - \mathbb{E}\left[\mathcal{F}_{(3)ti\cdot}\mathcal{F}'_{(3)tj\cdot}\right]\right\}\right|^2\right] \leq C_{\mathcal{F}}$ for some finite $C_{\mathcal{F}}$ independent of $i, j, T,$ and N.*

(v) *There exists an integer $\overline{M}$ such that for all $m > \overline{M}$, r is a finite positive integer, independent of m.*

(vi) *For all $h \in \mathbb{N}$, all $s \in \mathbb{Z}$, and all $l = 1, \ldots, r$, $\mathbb{E}\left[\left|\mathcal{F}_{(3)slh}\right|\right] \leq C'_{\mathcal{F}}$, for some finite $C'_{\mathcal{F}}$ independent of $h, s,$ and l.*

Assumptions 1(i) and 1(ii) imply that we consider only pervasive, or strong, factors. In other words, the network factors are loaded by most or all the network layers. Notice that the factor tensor has dimension $N \times N \times r$, hence the rescaling introduced in Assumption 1(ii), which accounts for its first two modes having diverging dimensions. Furthermore, under Assumption 1(ii), the 2nd order moments of the process $\{N^{-1}\text{vec}(\mathcal{F}_{(3)t}), t \in \mathbb{Z}\}$ are finite and independent of time, for all $N \in \mathbb{N}$.

Assumptions 1(iii) and 1(iv) imply that, given the factors, we can consistently estimate $\Gamma^{\mathcal{F}} := \mathbb{E}[\mathcal{F}_{(3)t}\mathcal{F}'_{(3)t}]$, as proved in Lemma E.5(i). Therefore, given the loadings, we can also consistently estimate $\Gamma^{\mathcal{X}} := U\mathbb{E}[\mathcal{F}_{(3)t}\mathcal{F}'_{(3)t}]U'$. Assumption 1(v) simply states that the number of factors is finite for all $m, N \in \mathbb{N}$ and that in order to find such factors we need m to be large enough. Finally, Assumption 1(vi) is very mild as it simply requires each element of the network factors to have finite first moment.

Assumption 2 (Idiosyncratic component of multilayer network).

(i) *For all $m, N \in \mathbb{N}$ and all $t \in \mathbb{Z}$, $\mathbb{E}\left[\mathcal{E}_{(3)t}\right] = 0_{m\times N^2}$ and $\Gamma^{\mathcal{E}} := \mathbb{E}[\mathcal{E}_{(3)t}\mathcal{E}'_{(3)t}]$ is $m\times m$ positive definite.*

(ii) *For all $N \in \mathbb{N}$, all $t, s \in \mathbb{Z}$, and all $i, j = 1, \ldots N^2$, $N^{-2}\sum_{h=1}^{N^2}\sum_{k=1}^{N^2}\left|\mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sjk}\right]\right| \leq \rho_{\mathcal{E}}^{|t-s|}M_{ij}$ and, for all $N \in \mathbb{N}$, all $t, s \in \mathbb{Z}$, and all $i, j, k = 1, \ldots, N^2$, $N^{-2}\sum_{h=1}^{N^2}\left|\mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sjk}\right]\right| \leq \rho_{\mathcal{E}}^{|t-s|}M_{ij}$ for some finite $\rho_{\mathcal{E}}$ and M_{ij} independent of t, s, k and N such that $0 \leq \rho_{\mathcal{E}} < 1$, $\sum_{i=1, i\neq j}^m M_{ij} \leq M_{\mathcal{E}}$ and $\sum_{j=1, j\neq i}^m M_{ij} \leq M_{\mathcal{E}}$, for some finite $M_{\mathcal{E}}$ independent of i, j and m.*

(iii) *For all $i, j \in \mathbb{N}$ and all $t \in \mathbb{Z}$, $\mathbb{E}[|\mathcal{E}_{(3)tij}|^4] \leq K_{\mathcal{E}}$ for some finite $K_{\mathcal{E}}$ independent of i, j, t.*

(iv) *For all $m, T, N \in \mathbb{N}$ and all $j = 1, \ldots, N^2$ and all $s = 1, \ldots, T$,*

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{mT}N^2}\sum_{i=1}^m\sum_{t=1}^T\sum_{h=1}^{N^2}\sum_{k=1}^{N^2}\left\{\mathcal{E}_{(3)tih}\mathcal{E}_{(3)tjk} - \mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)tjk}\right]\right\}\right|^2\right] \leq C_{\mathcal{E}}$$

and

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{mT}N^2}\sum_{i=1}^{m}\sum_{t=1}^{T}\sum_{h=1}^{N^2}\sum_{k=1}^{N^2}\left\{\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sik}-\mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sik}\right]\right\}\right|^2\right]\leq C_{\mathcal{E}}$$

for some finite $C_{\mathcal{E}}$ independent of j, s, m, T, N.

(v) For all $m, T, N \in \mathbb{N}$ and all $j = 1, \dots, N^2$ and all $s = 1, \dots, T$,

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{mT}N}\sum_{i=1}^{m}\sum_{t=1}^{T}\sum_{h=1}^{N^2}\left\{\mathcal{E}_{(3)tih}\mathcal{E}_{(3)tjh}-\mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)tjh}\right]\right\}\right|^2\right]\leq C_{\mathcal{E}}$$

and

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{mT}N}\sum_{i=1}^{m}\sum_{t=1}^{T}\sum_{h=1}^{N^2}\left\{\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sih}-\mathbb{E}\left[\mathcal{E}_{(3)tih}\mathcal{E}_{(3)sih}\right]\right\}\right|^2\right]\leq C_{\mathcal{E}}$$

for some finite $C_{\mathcal{E}}$ independent of j, s, m, T, N.

This assumption controls the serial and cross-sectional dependence of the entries of the idiosyncratic tensor. In particular, according to Assumption 2(ii), the covariance across time and layers is controlled in a standard way; hence, the classical summability conditions hold (see Lemma E.1 and Bai, 2003, Assumption C). The covariances across the elements of the N^2-dimensional vector $\mathcal{E}_{(3)tj\cdot}$ are of order N^4 for any given $t = 1, \dots, T$ and $j = 1, \dots, m$, and we require to scale their sum by N^2, thus assuming a standard summability condition. Assumptions 2(iii), 2(iv), and 2(v) imply that, given the idiosyncratic tensor, we can consistently estimate $m^{-1}N^{-2}\Gamma^{\mathcal{E}}$, for any $m, N \in \mathbb{N}$, as proved in Lemma E.5(ii).

Assumption 3 (Moment conditions - part 1). *For all $i, j \in \mathbb{N}$, all $k = 1, \dots, r$, and all $t \in \mathbb{Z}$, $\mathbb{E}[\mathcal{F}_{(3)tkj}\mathcal{E}_{(3)tij}] = 0$, and, for all $m, N, T \in \mathbb{N}$ and all $t = 1, \dots, T$,*

$$\mathbb{E}\left[\frac{1}{mN^2}\sum_{i=1}^{m}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{F}_{(3)t}\mathcal{E}'_{(3)ti\cdot}\right\|_F^2\right]\leq C_{\mathcal{F}\mathcal{E}},$$

$$\mathbb{E}\left[\frac{1}{N^4}\left\|\frac{1}{\sqrt{mT}}\sum_{i=1}^{m}\sum_{s=1}^{T}\mathcal{F}_{(3)s}\left\{\mathcal{E}'_{(3)si\cdot}\mathcal{E}_{(3)ti\cdot}-\mathbb{E}[\mathcal{E}'_{(3)si\cdot}\mathcal{E}_{(3)ti\cdot}]\right\}\right\|_F^2\right]\leq C'_{\mathcal{F}\mathcal{E}},$$

for some finite $C_{\mathcal{F}\mathcal{E}}$ and $C'_{\mathcal{F}\mathcal{E}}$ independent of t, m, N, and T.

Uncorrelatedness of the processes $\{\mathcal{F}_{(3)t}\}$ and $\{\mathcal{E}_{(3)t}\}$ is a natural assumption, while the moment conditions controlling higher-order dependence are standard and are direct extensions of what typically assumed in the vector case (Bai, 2003, Assumptions D and F1).

Define the $m \times m$ matrix $\Gamma^{\mathcal{W}} := \mathbb{E}[\mathcal{W}_{(3)t}\mathcal{W}'_{(3)t}]$ and let Γ^{χ} be as previously defined. Denote the eigenvalues of Γ^{χ} as μ_j^{χ}, $j = 1, \ldots, r$, in decreasing order. Then, as established by Lemmas E.1(i) and E.1(iv), we have that $\mu_j^{\chi} \asymp mN^2$ and $\|N^{-2}\Gamma^{\mathcal{E}}\|$ is finite for all $N \in \mathbb{N}$. As a consequence, by Assumption 3 and Weyl's inequality, the matrix $\Gamma^{\mathcal{W}} = \Gamma^{\chi} + \Gamma^{\mathcal{E}}$ is characterized by an eigengap between the r-th and the $(r+1)$-th largest eigenvalues which widens as $m \to \infty$. This property allows us to identify the number of network factors r and it is the rationale for the eigenvalue ratio criterion for estimating r defined in (4.11). This also means that the network factor model in (4.2) is always identified as long as $m \to \infty$.

In general, the network factors and their loadings are not separately identified unless we impose further restrictions. To this end we make the following assumption.

Assumption 4 (Identification of network factors and loadings).

(i) For all $m \in \mathbb{N}$, $m^{-1}U'U$ is diagonal with distinct entries.

(ii) For all $N, T \in \mathbb{N}$, $N^{-2}T^{-1}\sum_{t=1}^{T}\mathcal{F}_{(3)t}\mathcal{F}'_{(3)t} = I_r$.

Under Assumption 4 the columns of the loadings matrix U and the layers of the tensor factor $\mathcal{F}_t$ are identified up to a sign multiplication. This identification scheme is a classical one adopted for example by Bai (2009) in the vector factor model case.

Assumption 5 (CLTs).

(i) For any given $i = 1, \ldots, m$, as $N, T \to \infty$, $\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\mathcal{F}_{(3)t}\mathcal{E}'_{(3)ti\cdot} \xrightarrow{d} \mathcal{N}(0_r, \Phi_i)$, where $\Phi_i := \lim_{N,T\to\infty}\mathbb{E}\left[\left(\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\mathcal{F}_{(3)t}\mathcal{E}'_{(3)ti\cdot}\right)\left(\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\mathcal{F}_{(3)t}\mathcal{E}'_{(3)ti\cdot}\right)'\right]$.

(ii) For any given $t = 1, \ldots, T$ and $j = 1, \ldots, N^2$, as $m \to \infty$, $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}u_i\mathcal{E}_{(3)tij} \xrightarrow{d} \mathcal{N}(0_r, \Pi_{tj})$, where $\Pi_{tj} := \lim_{m\to\infty}\mathbb{E}\left[\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}u_i\mathcal{E}_{(3)tij}\right)\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}u_i\mathcal{E}_{(3)tij}\right)'\right]$ and u_i' is the i-th row of U.

Assumption 5(i) is standard in the vector factor model case; i.e., when $N = 1$, where it is satisfied for example by strong-mixing processes (Bai, 2003, Assumption F4). In Assumption 5(ii) we directly assume a cross-sectional CLT, which is a standard approach in the vector factor model case (Bai, 2003, Assumption F3).

The FNAR errors ν_t follow a factor model given in (3.3), characterized by the following assumption.

Assumption 6 (FNAR errors).

(i) $\lim_{N\to\infty} \|N^{-1}\Lambda'\Lambda - \Sigma_\Lambda\| = 0$ *where* Σ_Λ *is* $q \times q$ *finite and positive definite, and, for all* $i \in \mathbb{N}$, $\|\Lambda_{i\cdot}\| \leq M_\Lambda$ *for some finite* M_Λ *independent of* i.

(ii) *For all* $t \in \mathbb{Z}$, $\mathbb{E}[G_t] = 0_q$, $\Gamma^G := \mathbb{E}\left[G_t G_t'\right]$ *is* $q \times q$ *positive definite, and such that* $\left\|\Gamma^G\right\| \leq M_G$ *for some finite* M_G *independent of* t.

(iii) *For all* $t \in \mathbb{Z}$, $\mathbb{E}[\|G_t\|^4] \leq K_G$ *for some finite* K_G *independent of* t.

(iv) *For all* $i, j = 1, \ldots, q$ *and all* $T \in \mathbb{N}$, $\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \left\{G_{it}G_{jt} - \mathbb{E}\left[G_{it}G_{jt}\right]\right\}\right|^2\right] \leq C_G$ *for some finite* C_G *independent of* i, j, *and* T.

(v) *There exists an integer* $\underline{N}$ *such that for all* $N > \underline{N}$, q *is a finite positive integer, independent of* N.

(vi) *For all* $i \in \mathbb{N}$ *and all* $t \in \mathbb{Z}$, $\mathbb{E}[\epsilon_{it}] = 0$, $\mathbb{E}[\epsilon_{it}^2] = \sigma_i^2$ *such that* $\sigma_i^2 \geq \underline{M}_\epsilon$ *and* $\sigma_i^2 \leq \overline{M}_\epsilon$ *for some finite* $\underline{M}_\epsilon$ *and* $\overline{M}_\epsilon$ *independent of* i *and* t.

(vii) *For all* $i, j \in \mathbb{N}$ *and all* $t, s \in \mathbb{Z}$, $\mathbb{E}\left[\epsilon_{it}\epsilon_{js}\right] = 0$ *if* $i \neq j$ *and* $\mathbb{E}\left[\epsilon_{it}\epsilon_{is}\right] = 0$ *if* $t \neq s$.

(viii) *For all* $i \in \mathbb{N}$ *and all* $t \in \mathbb{Z}$, $\mathbb{E}[\epsilon_{it}^4] \leq K_\epsilon$ *for some finite* K_ϵ *independent of* i *and* t.

(ix) *For all* $N, T \in \mathbb{N}$, $\mathbb{E}\left[\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \left\{\epsilon_{it}^2 - \mathbb{E}\left[\epsilon_{it}^2\right]\right\}\right|^2\right] \leq C_\epsilon$ *for some finite* C_ϵ *independent of* N *and* T.

This assumption is similar to the usual set of assumptions for the vector factor model (Bai, 2003). The comments to these assumptions are analogous to those previously made for the network factors and are therefore omitted. Concerning the idiosyncratic components, we follow Chen et al. (2023) and assume zero correlations both in time and across units. Given that we are considering a factor structure for the FNAR errors, this assumption is not very restrictive, as most of the correlations are likely to be already captured by the lagged terms in the FNAR and by the common factors G_t. Nevertheless, it is possible to develop the following asymptotic theory by allowing for the usual kind of weak cross- and autocorrelations between the components of ϵ_t.

Assumption 7 (Independence of network and node factors). *For all* $N \in \mathbb{N}$, *the processes* $\{\mathcal{F}_t,\, t \in \mathbb{Z}\}$, $\{\epsilon_t,\, t \in \mathbb{Z}\}$, *and* $\{G_t,\, t \in \mathbb{Z}\}$ *are mutually independent.*

Assumption 7 is taken from Bai (2009, Assumption D) and is made just to simplify the proof. In principle it could be relaxed to allow for weak dependence between $\{G_t\}$ and $\{\epsilon_t\}$ by means of a condition similar to those required in Assumption 3, which, in turn, derives from Bai (2003, Assumption D).

Because of Assumptions 6(i), 6(ii), 6(vi), 6(vii), and 7, the FNAR errors have covariance matrix $V = \Lambda\Gamma^G\Lambda' + S$, with $S = \mathbb{E}[\epsilon_t\epsilon_t']$, which is positive definite for all $N \in \mathbb{N}$. This also implies that V^{-1} is finite for all $N \in \mathbb{N}$. Moreover, V has the usual eigengap property; i.e., its largest q eigenvalues diverge at rate N, while the remaining $N-q$ stay bounded for all $N \in \mathbb{N}$. This implies that the factor model in (3.3), and therefore the number of factors q, is always identified as $N \to \infty$. This is the rationale for the eigenvalue ratio criterion for estimating q defined in (4.12).

Assumption 8 (Identification of node factors and loadings).

(i) For all $N \in \mathbb{N}$, $N^{-1}\Lambda'\Lambda$ is diagonal with distinct entries.

(ii) For all $T \in \mathbb{N}$, $T^{-1}\sum_{t=1}^T G_tG_t' = I_q$.

Under Assumption 8 the columns of the loadings matrix Λ and the factors G_t are identified up to a sign multiplication.

Turning to the FNAR defined in (3.2), we make the following assumption.

Assumption 9 (Stability of FNAR).

(i) For all $t \in \mathbb{Z}$ and all $N \in \mathbb{N}$, $\det\Big(I_N - \rho I_N - N^{-1}\sum_{j=1}^r \beta_j \mathbb{E}[F_{j,t}]\Big) \neq 0$.

(ii) For all $t \in \mathbb{Z}$ and all $N \in \mathbb{N}$, $\det\Big(\rho^2 I_{N^2} + N^{-2}\sum_{j=1}^r \beta_j^2 \mathbb{E}[F_{j,t} \otimes F_{j,t}] - zI_{N^2}\Big) = 0$ has roots $z_j^* \in \mathbb{C}$, $j = 1, \dots, N^2$, such that $|z_j^*| \leq C_S$ for some finite $C_S < 1$ independent of j, t, and N.

This assumption is a generalization to the case of random multivariate AR models of the usual stability conditions for a VAR. As shown below it implies, together with Assumptions 1(ii) and 7, that the FNAR has a stationary solution for all $N \in \mathbb{N}$. Notice that Assumption 9(i) is stated for the general case in which $\mathbb{E}[F_{j,t}] \neq 0$, otherwise the condition needed to ensure the existence of the mean is simply $|\rho| < 1$.

Assumption 10 (Moment conditions - part 2). *For all $m, N, T \in \mathbb{N}$,*

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{mT}N^2}\sum_{t=1}^T\sum_{i=1}^m u_i \mathcal{E}_{(3)ti\cdot}(y_{t-1} \otimes X_t)\right\|^2\right] \leq \mathfrak{K}_1,$$

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{mT}N^2}\sum_{t=1}^{T}\sum_{i=1}^{m}u_i\mathcal{E}_{(3)ti\cdot}(y_{t-1}\otimes\nu_t)\right\|^2\right]\le\mathfrak{K}_2,$$

for some finite $\mathfrak{K}_1$ and $\mathfrak{K}_2$ independent of $m, N,$ and T.

To get an intuition of this assumption, consider the $m \times (r+2)$ matrix process $\{\mathcal{E}_{(3)t}(y_{t-1} \otimes X_t)\}$. We are saying that this process is weakly correlated along the time dimension, which is a standard requirement, but it is also weakly correlated across its m rows. The latter requirement is fulfilled by the idiosyncratic terms $\mathcal{E}_{(3)t}$ via Assumption 2(ii), and here is extended to the case in which $\mathcal{E}_{(3)t}$ is multiplied by $y_{t-1} \otimes X_t$ which is weakly dependent of $\mathcal{E}_{(3)t}$ because of Assumption 3.

Assumption 11 (CLT for FNAR). *Let $Z_i := M_G X_i - \frac{1}{N}\sum_{k=1}^{N}(\Lambda_i'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\Lambda_k)\, M_G X_k$ such that $Z_i := (Z_{i1}\cdots Z_{iT})'$ is $T \times (r+2)$ and $W_t := M_\Lambda X_t - \frac{1}{T}\sum_{s=1}^{T}(G_t'G_s)M_\Lambda X_s$ such that $W_t := (W_{1t}\cdots W_{Nt})'$ is $N \times (r+2)$. Then, as $N, T \to \infty$,*

(i) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} Z_i'\epsilon_i \xrightarrow{d} N(0, D_1)$, *where* $D_1 := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Z_{it}Z_{it}'\right]\sigma_i^2$ *is an* $(r+2)\times(r+2)$ *positive definite matrix.*

(ii) $\frac{1}{NT}\sum_{i=1}^{N} Z_i'Z_i \xrightarrow{p} \Sigma_{ZZ}$, *where* $\Sigma_{ZZ} := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Z_{it}Z_{it}']$ *is an* $(r+2)\times(r+2)$ *positive definite matrix.*

(iii) $\frac{1}{\sqrt{NT}}\sum_{t=1}^{T} W_t'\epsilon_t \xrightarrow{d} N(0, D_2)$, *where* $D_2 := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[W_{it}W_{it}'\right]\sigma_i^2$ *is an* $(r+2)\times(r+2)$ *positive definite matrix.*

(iv) $\frac{1}{NT}\sum_{t=1}^{T} W_t'W_t \xrightarrow{p} \Sigma_{WW}$, *where* $\Sigma_{WW} := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[W_{it}W_{it}']$ *is an* $(r+2)\times(r+2)$ *positive definite matrix.*

Notice that we do not rule out autocorrelation in the node factors, as only the node specific idiosyncratic components are required to have no correlation for the above CLTs to hold, and this is ensured by Assumption 6(vii). These assumptions are similar to the conditions in Bai (2009, Assumptions A and E).

5.2 Stationarity

In order to develop the theory for the FNAR, we first discuss under which conditions equation (3.2) admits a stationary causal solution. Given the difficulty of the problem we limit ourselves to consider weakly stationary solutions. This poses two issues. First, the FNAR is defined for an N-dimensional vector y_t where we allow $N \to \infty$. Second, the FNAR is an autoregressive model with stochastic time-varying coefficients. Regarding the former issue, we adopt the definition proposed by Zhu et al. (2017).

Definition 5.1. *Let $\{y_t\}$ be an N-dimensional stochastic process with $N \in \mathbb{N}$. Let $W := \{\omega := (\omega_1 \cdots \omega_N)' \in \mathbb{R}^N : \sum_{i=1}^N |\omega_i| < \infty, N \in \mathbb{N}\}$. Then, $\{y_t\}$ is weakly stationary if for all $N \in \mathbb{N}$ and any given $\omega \in W$, $y_t^\omega := \lim_{N\to\infty} \omega' y_t$ exists almost surely and $\{y_t^w\}$ is weakly stationary and causal.*

Turning to the second problem, we have the following result.

Proposition 5.1 (Stationarity of FNAR). *Under Assumptions 1(ii), 6(vi), 7, and 9, for all $N \in \mathbb{N}$, the FNAR has a unique weakly stationary and causal solution.*

In general, one might object that when $N \to \infty$ a meaningful concept of stationarity cannot be stated, as no causal solution to the FNAR can exist since Assumption 9 will break down, see, e.g., the remark by Zhou et al. (2020), in a similar context. What we mean by Definition 5.1 and Proposition 5.1 is that we are implicitly assuming that there exists a space where the causal solution is well-defined even when $N \to \infty$. This is the same approach adopted by Zhu et al. (2017). An interesting implication of this definition is that, under our assumptions, we can ensure that any finite linear combination of the elements of $\{y_t\}$ satisfies a finite dimensional FNAR with a causal solution.

5.3 Asymptotic properties of network factors and FNAR coefficients

Consistency and asymptotic normality of the estimated network factor loadings are given next.

Theorem 5.1 (Consistency and asymptotic normality of loadings).

(i) Under Assumptions 1-4, as $m, N, T \to \infty$,

$$\left\| \frac{\widehat{U} - UJ}{\sqrt{m}} \right\| = O_p\left(\max\left(\frac{1}{N\sqrt{T}}, \frac{1}{m} \right) \right),$$

where J is a $r \times r$ diagonal matrix whose diagonal entries are equal to ± 1.

(ii) Under Assumptions 1-5, for any given $i = 1, \dots, m$, as $m, N, T \to \infty$, if $N\sqrt{T}/m \to 0$,

$$N\sqrt{T}\left(\widehat{u}_i' - u_i' J\right) \xrightarrow{d} \mathcal{N}\left(0_r, \Phi_i\right),$$

where $\widehat{u}_i'$ and u_i' are the i-th rows of $\widehat{U}$ and U, respectively, Φ_i is defined in Assumption 5(i), and J is defined in part (i).

Theorem 5.1 shows that, when applying PCA to a given mode of the tensor $\mathcal{W}_t$, the dimensions of all other modes contribute to a faster convergence rate, hence allowing for more degrees of freedom. This is an advantage with respect to the vector case, since even for moderately small values of T we can still have good estimates of the loadings matrix and therefore of the network factors. In particular, we see that the estimated loadings vector $\widehat{u}_i$ has a consistency rate $\min(m, N\sqrt{T})$ and is asymptotically normal if $N\sqrt{T}/m \to 0$. This is the generalization to the multilayer network case (i.e., to order-3 tensors) of the usual vector case, which corresponds to setting $N = 1$ (see Bai, 2003, Theorem 2).

Next we prove consistency and asymptotic normality of the estimated network factors.

Theorem 5.2 (Consistency and asymptotic normality of network factors).

(i) Under Assumptions 1-4, for any given $t = 1, \ldots, T$, as $m, N, T \to \infty$,

$$\left\| \frac{\widehat{\mathcal{F}}_{(3)t} - J\mathcal{F}_{(3)t}}{N} \right\| = O_p\left(\max\left(\frac{1}{N^2 T}, \frac{1}{\sqrt{m}}\right)\right),$$

where J is defined in Theorem 5.1(i).

(ii) Under Assumptions 1-5, for any given $t = 1, \ldots, T$ and $j = 1, \ldots, N^2$, as $m, N, T \to \infty$, if $\sqrt{m}/(N^2 T) \to 0$,

$$\sqrt{m}\left(\widehat{\mathcal{F}}_{(3)t\cdot j} - J\mathcal{F}_{(3)t\cdot j}\right) \xrightarrow{d} \mathcal{N}\left(0_r, \Sigma_U^{-1}\Pi_{tj}\Sigma_U^{-1}\right),$$

where Π_{tj} is defined in Assumption 5(ii) and J is defined in Theorem 5.1(i).

Theorem 5.2(i) proves consistency of the whole network factor tensor. Theorem 5.2(ii) proves asymptotic normality of any given column of $\widehat{\mathcal{F}}_{(3)t}$, which is equivalent to asymptotic normality of any of the N^2 entries of each of the r layers of the multilayer network factor $\widehat{\mathcal{F}}_t$. This is the natural generalization to the multinetwork case of the usual vector case, i.e., when $N = 1$ (see Bai, 2003, Theorem 1).

We then turn to the asymptotic properties of the estimated FNAR coefficients. Hereafter, let

$$\bar{J} := \begin{bmatrix} J & 0_{r\times 2} \\ 0_{2\times r} & I_2 \end{bmatrix}, \tag{5.1}$$

with J as in Theorem 5.1(i). Then, we analyze the properties of the estimators $\widehat{\theta}^{\dagger}$ and $\widehat{\theta}^{*}$ by noticing that

$$\begin{aligned} \widehat{\theta}^{\dagger}-\bar{J}\theta &= \left(\tfrac{1}{NT}\textstyle\sum_{i=1}^{N}\widehat{X}_i' M_{\widehat{G}^{\dagger}}\widehat{X}_i\right)^{-1}\left(\tfrac{1}{NT}\textstyle\sum_{i=1}^{N}\widehat{X}_i' M_{\widehat{G}^{\dagger}}(G\Lambda_i+\epsilon_i)+\tfrac{1}{NT}\textstyle\sum_{i=1}^{N}\widehat{X}_i' M_{\widehat{G}^{\dagger}}u_i\right) \\ \widehat{\theta}^{*}-\bar{J}\theta &= \left(\tfrac{1}{NT}\textstyle\sum_{t=1}^{T}\widehat{X}_t' M_{\widehat{\Lambda}^{*}}\widehat{X}_t\right)^{-1}\left(\tfrac{1}{NT}\textstyle\sum_{t=1}^{T}\widehat{X}_t' M_{\widehat{\Lambda}^{*}}(\Lambda G_t+\epsilon_t)+\tfrac{1}{NT}\textstyle\sum_{t=1}^{T}\widehat{X}_t' M_{\widehat{\Lambda}^{*}}u_t\right) \end{aligned} \tag{5.2}$$

where $u_i := (X_i\bar{J}-\widehat{X}_i)\bar{J}\theta$ and $u_t := (X_t\bar{J}-\widehat{X}_t)\bar{J}\theta$, and recall that $\theta := (\beta', \rho, \alpha)'$.

The following theorem holds.

Theorem 5.3 (CLT for FNAR coefficients estimated by iterative OLS). *Under Assumptions 1-10 and 11, and if $\sqrt{NT}/m \to 0$ and $N/m \to 0$, as $m, N, T \to \infty$, then:*

(i) if $T/N \to 0$ and $\sqrt{N}/T \to 0$, we have

$$\sqrt{NT}(\widehat{\theta}^{\dagger}-\bar{J}\theta) \xrightarrow{d} \mathcal{N}(0_{r+2}, \Sigma_{ZZ}^{-1}D_1\Sigma_{ZZ}^{-1}),$$

where D_1 and Σ_{ZZ} are defined in Assumptions 11(i) and 11(ii), and $\bar{J}$ is defined in (5.1);

(ii) if $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$ and $\sigma_i^2 = \sigma^2$ for all $i = 1, \dots, N$, we have

$$\sqrt{NT}(\widehat{\theta}^{\dagger}-\bar{J}\theta) \xrightarrow{d} \mathcal{N}(0_{r+2}, \sigma^2\Sigma_{ZZ}^{-1}),$$

where Σ_{ZZ} is defined in Assumption 11(ii), and $\bar{J}$ is defined in (5.1);

(iii) if $T/N \to 0$ and $\sqrt{N}/T \to 0$, we have

$$\sqrt{NT}(\widehat{\theta}^{*}-\bar{J}\theta) \xrightarrow{d} \mathcal{N}(0_{r+2}, \Sigma_{WW}^{-1}D_2\Sigma_{WW}^{-1}),$$

where D_2 and Σ_{WW} are defined in Assumptions 11(iii) and 11(iv), and $\bar{J}$ is defined in (5.1);

(iv) if $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$ and $\sigma_i^2 = \sigma^2$ for all $i = 1, \dots, N$, we have

$$\sqrt{NT}(\widehat{\theta}^{*}-\bar{J}\theta) \xrightarrow{d} \mathcal{N}(0_{r+2}, \sigma^2\Sigma_{WW}^{-1}),$$

where Σ_{WW} is defined in Assumption 11(iv), and $\bar{J}$ is defined in (5.1);

Parts (i) and (ii) extend Theorem 2 in Bai (2009) to the FNAR case. The interesting cases are parts (i) and (iii), where we do not impose homoskedastic idiosyncratic components in the FNAR errors.

Notice that the network coefficients, β_j, $j = 1, \ldots, r$, which are the first r elements of θ, are consistently estimated only up to a sign, due to the indeterminacy in the identification of the network factors.

Estimators of the asymptotic variance-covariance matrix of $\widehat{\theta}^{\dagger}$ and $\widehat{\theta}^{*}$ under the assumptions in parts (i) and (iii) are (see also Bai, 2009):

$$\widehat{\text{Avar}}\left[\sqrt{NT}(\widehat{\theta}^{\dagger} - \bar{J}\theta)\right] = \left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{Z}_i'\widehat{Z}_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{Z}_i'\widehat{Z}_i\frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_{it}^2\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{Z}_i'\widehat{Z}_i\right)^{-1},$$

$$\widehat{\text{Avar}}\left[\sqrt{NT}(\widehat{\theta}^{*} - \bar{J}\theta)\right] = \left(\frac{1}{NT}\sum_{t=1}^{T}\widehat{W}_t'\widehat{W}_t\right)^{-1}\left(\frac{1}{NT}\sum_{t=1}^{T}\widehat{W}_t'\widehat{W}_t\frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_{it}^2\right)\left(\frac{1}{NT}\sum_{t=1}^{T}\widehat{W}_t'\widehat{W}_t\right)^{-1}, \quad (5.3)$$

where $\widehat{Z}_i := M_{\widehat{G}^{\dagger}}\widehat{X}_i - \frac{1}{N}\sum_{k=1}^{N}\left(\widehat{\Lambda}_i^{\dagger\prime}\left(\frac{\widehat{\Lambda}^{\dagger\prime}\widehat{\Lambda}^{\dagger}}{N}\right)^{-1}\widehat{\Lambda}_k^{\dagger}\right)M_{\widehat{G}^{\dagger}}\widehat{X}_k$, and $\widehat{W}_t := M_{\widehat{\Lambda}^*}\widehat{X}_t - \frac{1}{T}\sum_{s=1}^{T}(\widehat{G}_t^{*\prime}\widehat{G}_s^{*})M_{\widehat{\Lambda}^*}\widehat{X}_s$.

Four important comments about this result follow. First, the proof is based on showing that if $\sqrt{NT}/m \to 0$ and $N/m \to 0$, as $m, N, T \to \infty$, then the network factors can be treated as observed; i.e., the generated regressors bias is asymptotically negligible (see Proposition C.1). Under these conditions, and if also $T/N \to 0$ and $\sqrt{N}/T \to 0$, the iterative estimators are $\sqrt{NT}$-consistent.

Second, if the node factors are not autocorrelated we can also compare the iterative estimators with the OLS and the GLS estimators studied in Appendix B and C, respectively. In this case, the GLS is also $\sqrt{NT}$-consistent since, similarly to the iterative estimator, it rescales X_t by the FNAR error covariance matrix V, which is $O(N)$ by Assumption 6. For the same reason the OLS estimator is just $\sqrt{T}$-consistent since it does not control for the FNAR error covariance, and it would be $\sqrt{NT}$-consistent only if V were a diagonal matrix; i.e., when no node factor is present, as assumed by Zhu et al. (2017). The same results on OLS and GLS are obtained by Chen et al. (2023) for the case of observed networks.

Third, the two estimators are asymptotically equivalent and thus also equally efficient. To see this notice that $\widehat{\theta}^{\dagger}$ and $\widehat{\theta}^{*}$ are such that they solve

$$\widehat{\theta}^{\dagger} = \arg\min_{\theta}\frac{1}{NT}\sum_{i=1}^{N}(y_i - X_i\theta - G\Lambda_i)'(y_i - X_i\theta - G\Lambda_i) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - X_{it}\theta - \Lambda_i'G_t)^2, \quad (5.4)$$

$$\widehat{\theta}^{*} = \arg\min_{\theta}\frac{1}{NT}\sum_{t=1}^{T}(y_t - X_t\theta - \Lambda G_t)'(y_t - X_t\theta - \Lambda G_t) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - X_{it}\theta - \Lambda_i'G_t)^2, \quad (5.5)$$

so the two losses are identical and must have the same minimum. Once we fix the identification constraints as in Assumption 8, the only difference is then about the implementation of the minimizations. For $\widehat{\theta}^{\dagger}$, by replacing $\Lambda_i = (G'G)^{-1}G'(y_i - X_i\theta) = T^{-1}G'(y_i - X_i\theta)$ in (5.4), we can solve for G and θ only. For $\widehat{\theta}^{*}$, by replacing $G_t = (\Lambda'\Lambda)^{-1}\Lambda'(y_t - X_t\theta)$ in (5.5), we can solve for Λ and θ only. These two procedures lead to the two solutions given in (4.9). However, since in practice the solutions are obtained by iteration, the two estimates might not coincide exactly, although they will be very similar. And, as expected, the estimated standard errors will coincide (see the results in Section 7 and Appendix F.4).

Fourth, and last, we should view Theorem 5.3 as giving the asymptotic distribution of the theoretical estimator minimizing (5.4) or (5.5). This is the same point of view adopted by Bai (2009). In practice, it might be important to investigate how the initialization of the algorithm affects such convergence. In the simpler case of a panel regression having errors with a factor structure, Jiang et al. (2021, Theorem 3) show that any initial estimator could still lead to a consistent iterated estimator, depending on the structure of the regressors which can be quite general. However, the estimator computed in practice might have a slower convergence rate if the initial estimator is not consistent. We do not explore this aspect further here, but we limit ourselves to notice that in our numerical exercises of Sections 6 and 7, convergence is always achieved in few steps and the iterated estimator works well even in presence of weak serial correlation of the node factors.

6 Monte Carlo simulations

To evaluate the finite sample performance of the proposed estimators, we generate artificial time series of y_t and $\mathcal{W}_t$, for $t = 1, \dots, T$, according to the model equations (3.1), (3.2), and (3.3). We fix $r = 1$ and $q = 1$; i.e., one network factor and one node-specific factor. We also fix the values of FNAR parameters $\beta = 0.5$, $\rho = 0.3$, and $\alpha = 0.2$. We consider $N \in \{10, 20, 50, 100, 200\}$ nodes, $m \in \{20, 50, 100\}$ layers, and $T \in \{50, 100\}$ time periods. Also, for each value of the pair (N, m), we randomly generate the entries of the loading vectors U and Λ once (and independently) from $\mathcal{N}(1, 1)$, and then we keep them fixed across Monte Carlo (MC) iterations (see Chen et al., 2023). All other quantities are generated at each MC iteration. All results are based on 500 iterations. Full details on the data generating process are in

Appendix F.1.

Table 1: Monte Carlo RMSEs - $T = 50$, case II: dependent $\mathcal{E}_t$

		$N=10$		$N=20$		$N=50$		$N=100$		$N=200$	
		RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE
	β	0.084	16.9%	0.078	15.6%	0.077	15.4%	0.075	15.0%	0.078	15.6%
	ρ	0.045	14.9%	0.031	10.2%	0.021	7.1%	0.015	5.0%	0.011	3.6%
$m=20$	α	0.084	41.8%	0.043	21.6%	0.029	14.4%	0.020	10.1%	0.014	7.2%
	$\mathcal{F}$	0.206	21.7%	0.211	21.7%	0.214	21.7%	0.216	21.7%	0.216	21.7%
	U	0.085	3.9%	0.079	3.6%	0.078	3.5%	0.078	3.5%	0.078	3.5%
	β	0.090	17.9%	0.078	15.6%	0.073	14.6%	0.080	16.0%	0.077	15.4%
	ρ	0.047	15.8%	0.032	10.7%	0.020	6.7%	0.014	4.8%	0.010	3.5%
$m=50$	α	0.068	34.1%	0.045	22.4%	0.030	15.2%	0.019	9.4%	0.015	7.4%
	$\mathcal{F}$	0.132	13.9%	0.135	13.9%	0.137	13.9%	0.138	13.9%	0.138	13.9%
	U	0.037	2.1%	0.029	1.6%	0.027	1.5%	0.026	1.5%	0.026	1.5%
	β	0.086	17.1%	0.081	16.1%	0.078	15.6%	0.078	15.7%	0.078	15.7%
	ρ	0.047	15.6%	0.031	10.3%	0.021	7.0%	0.015	5.0%	0.011	3.6%
$m=100$	α	0.087	43.6%	0.044	22.1%	0.027	13.6%	0.019	9.6%	0.014	7.0%
	$\mathcal{F}$	0.093	9.8%	0.096	9.8%	0.097	9.8%	0.098	9.8%	0.098	9.8%
	U	0.026	1.6%	0.016	1.0%	0.013	0.8%	0.012	0.7%	0.012	0.7%

Table 2: Monte Carlo RMSEs - $T = 100$, case II: dependent $\mathcal{E}_t$

		$N=10$		$N=20$		$N=50$		$N=100$		$N=200$	
		RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE	RMSE	ReRMSE
	β	0.056	11.1%	0.056	11.2%	0.055	11.0%	0.053	10.6%	0.056	11.2%
	ρ	0.031	10.2%	0.022	7.2%	0.014	4.6%	0.010	3.3%	0.008	2.5%
$m=20$	α	0.056	28.0%	0.029	14.5%	0.020	10.1%	0.014	7.2%	0.010	5.0%
	$\mathcal{F}$	0.206	21.7%	0.211	21.7%	0.214	21.7%	0.216	21.7%	0.216	21.7%
	U	0.081	3.7%	0.078	3.6%	0.078	3.5%	0.078	3.5%	0.078	3.5%
	β	0.064	12.9%	0.055	11.0%	0.054	10.9%	0.053	10.7%	0.056	11.2%
	ρ	0.031	10.3%	0.021	7.0%	0.014	4.7%	0.010	3.3%	0.008	2.5%
$m=50$	α	0.048	24.0%	0.032	15.8%	0.019	9.6%	0.014	6.9%	0.010	5.2%
	$\mathcal{F}$	0.132	13.9%	0.135	13.9%	0.137	13.9%	0.138	13.9%	0.138	13.9%
	U	0.032	1.8%	0.027	1.5%	0.026	1.5%	0.026	1.5%	0.026	1.5%
	β	0.060	12.1%	0.056	11.3%	0.054	10.8%	0.055	10.9%	0.055	11.0%
	ρ	0.030	10.0%	0.022	7.4%	0.014	4.5%	0.010	3.4%	0.007	2.5%
$m=100$	α	0.059	29.4%	0.030	15.1%	0.018	8.8%	0.014	7.1%	0.010	4.9%
	$\mathcal{F}$	0.093	9.8%	0.096	9.8%	0.097	9.8%	0.098	9.8%	0.098	9.8%
	U	0.020	1.3%	0.014	0.9%	0.012	0.8%	0.012	0.7%	0.012	0.7%

Tables 1-2 report the RMSE and Relative RMSE (ReRMSE) of the estimates. Here we report results under case II, which corresponds to idiosyncratic terms $\mathcal{E}_t$ having serial and cross-layer correlation, and using the iterative estimator $\widehat{\theta}^{\dagger}$ in (4.9). Additional results for case I of uncorrelated idiosyncratic terms are in Appendix F.2. As predicted by the theory, the accuracy of estimates for β, ρ and α improves with both N and T, and the RMSE of network factors and loadings decreases when the number of layers m increases. Furthermore, as shown in Appendix F.3, the MC distributions of the estimated network coefficient are all strongly centered around the true value $\beta = 0.5$ and become narrower as T and N increase.

Last, in Appendix F.5 we compare our estimates of the loadings U with those obtained using the TOPUP and TIPUP estimation methods proposed by Chen et al. (2022). As expected our approach improves over those estimators in presence of serial idiosyncratic correlation.

7 Empirical application

In this section, we present an application of the FNAR for studying cross-country macroeconomic interdependence determined by global trade flows and cross-border financial relationships.

Data. For a sample of $N = 24$ countries, we use $m = 25$ networks constructed using bilateral import/export flows for different good (layers 1–9) and services (10–19) categories, bilateral financial positions for different types of financial claims (20–23) and cross-border mergers and acquisitions classified by sector of economic activity (24–25). The list of countries and network layers, including details on how the networks are built, are given in Appendix G.1.

Network factors. Due to data limitations in the time series of financial positions, we collect data for the networks at the annual frequency from 2001 to 2019, so the factor analysis is conducted on a sample of length $T_1 = 19$.[1] Although this is a short time span, we recall that in tensor PCA the effective sample size when estimating the loadings space is N^2T_1 (see Theorem 5.1).

We then extract the common network factors from the 25 layers of the network, and we set $\widehat{r} = 6$ network factors, as in Chen et al. (2022). From Figures G.9-G.12 in Appendix G.2 we can interpret the six network factors as follows. The first network factor conveys approximately the average country weights across all layers of the network. In particular, the factor values are very close to the average

[1] We have few missing values over this period. In these cases, we use the previous year's value or the closest available year's value.

weights (scaled by a constant), and the loading coefficients are almost the same for all layers. The countries with the largest factor weights for the US are its major economic partners: Canada, UK, Mexico, China, Japan and Germany.

The second factor captures a difference between financial relationships and trade in goods. The factor loadings for financial layers have opposite sign (positive) compared to the loadings for trade-in-goods layers (negative). Recall that factors are identified up to a sign. The largest positive weights are assigned to economies having relatively large financial sectors with global reach: UK, US, and Hong Kong. In the case of the US connections, a large positive weight is assigned to the UK, whose tight economic links with the US are mostly concentrated in the financial sector, and large negative weights are assigned to Canada, Mexico, and China, i.e., the US biggest trade partners.

The third factor distinguishes between equity and debt relationships, being the only factor where equity, on the one hand, and debt, on the other hand, show loadings with opposite signs. The fourth factor is strongly associated with M&A relationships. The fifth factor is mainly driven by agricultural/extractive goods (positive weights, especially for vegetable fuels, oils, fats, and waxes). It also loads on trade in manufacturing goods (negative weights). Positive weights are assigned to countries with strong trade links with the US in non-manufacturing sectors, such as Canada, Saudi Arabia, and Italy, while negative weights are associated with large manufacturing partners, like China. This factor also distinguishes between stocks of portfolio holdings and flows associated with M&A deals and banking. Finally, the sixth factor captures a distinction between goods-sector M&A integration and services-sector integration.

Next, in line with conventional PCA, we evaluate the fraction of variance in network layers explained by each factor, denoted as $v^{(k)}$, $k = 1, \ldots, 6$ (computed as in Appendix G.2). We have $v^{(1)} = 0.68$, $v^{(2)} = 0.07$, $v^{(3)} = 0.03$, $v^{(4)} = 0.03$, $v^{(5)} = 0.02$, and $v^{(6)} = 0.02$. Thus, overall the 6 factors explain about 85% of the total variance of $\mathcal{W}$. However, the importance of different factors varies greatly across countries; see Table G.11 in Appendix G.2.

FNAR COEFFICIENTS. The endogenous vector y_t, $t = 1, \ldots, T$, collects (quarterly) real GDP growth rates for all N considered countries and for the sample 2001Q1-2019Q4; i.e., $T_2 = 76$. To address heterogeneity of nodal and momentum effects, we split the countries into two groups: (1) advanced economies ($N_1 = 15$), and (2) emerging economies ($N_2 = 9$) and the vector y_t is partitioned accordingly as $y_t = (y_t^{(1)'}; y_t^{(2)'})'$. We consider the following FNAR, for $t = 1, \dots, T$,

$$y_t = \sum_{j=1}^{r} \beta_j \frac{\widetilde{F}_{j,t-1}}{N} y_{t-1} + \rho^{(1)} \begin{pmatrix} y_{t-1}^{(1)} \\ 0_{N_2} \end{pmatrix} + \rho^{(2)} \begin{pmatrix} 0_{N_1} \\ y_{t-1}^{(2)} \end{pmatrix} + \alpha^{(1)} \begin{pmatrix} \iota_{N_1} \\ 0_{N_2} \end{pmatrix} + \alpha^{(2)} \begin{pmatrix} 0_{N_1} \\ \iota_{N_2} \end{pmatrix} + \nu_t, \tag{7.1}$$

where $\widetilde{F}_{j,t} = \widehat{F}_{j,\tau}$ for $4(\tau - 1) + 1 \le t \le 4\tau$, $\tau = 1, \dots, T_1$. In other words, the network factors $F_{k,t}$, $t = 1, \dots, T_1$, which are computed on a yearly basis, are treated as constant throughout all quarters of a given year. Hereafter, we let $\theta := (\beta', \rho^{(1)}, \rho^{(2)}, \alpha^{(1)}, \alpha^{(2)})'$.

By means of the criterion defined in (4.12), we find evidence of one common node factor, i.e., $\widehat{q} = 1$. We then estimate the model by GLS as described in Appendix C. Last, we consider the iterative estimators $\widehat{\theta}^{\dagger}$ or $\widehat{\theta}^{*}$ defined in (4.9) and we initialize the algorithm by using the GLS estimator and the estimated node loadings, $\widehat{\Lambda}$, and factor, $\widehat{G}_t$, computed by PCA on the GLS residuals as described in Appendix A. Since these residuals do not display significant autocorrelation, we are confident that the GLS estimator is $\sqrt{NT}$-consistent and, based on the results of Jiang et al. (2021) we conjecture that Theorem 5.3 holds for our iterated estimators. Convergence is reached in 8 or 4 iterations for $\widehat{\theta}^{\dagger}$ or $\widehat{\theta}^{*}$, respectively.

Table 3 reports the estimated coefficients and their standard errors with significance reported according to the usual Z-test. The coefficients on $N^{-1}\widehat{F}_{1,t-1}y_{t-1}$ and $N^{-1}\widehat{F}_{5,t-1}y_{t-1}$ are always strongly significant, while there is mixed evidence regarding the coefficients on $N^{-1}\widehat{F}_{2,t-1}y_{t-1}$, $N^{-1}\widehat{F}_{4,t-1}y_{t-1}$, and $N^{-1}\widehat{F}_{6,t-1}y_{t-1}$ which are mildly significant and not for all estimates.

Based on the interpretation of the first network factor, the coefficient on $N^{-1}\widehat{F}_{1,t-1}y_{t-1}$ captures a general network effect operating through aggregate economic linkages. Given the loadings of factor 5 in Figure G.10, the coefficient on $N^{-1}\widehat{F}_{5,t-1}y_{t-1}$ indicates that trade in mineral fuels and in animal and vegetable oils (major inputs of chemical industry) has the main impact on GDP growth. Apart from this, trade linkages tend to generate larger spillovers in manufacturing sectors (layers 6-9) than in

Table 3: FNAR coefficient estimates and standard errors.

	$\widehat{\theta}^{\mathrm{OLS}}$	$\widehat{\theta}^{\mathrm{GLS}}$	$\widehat{\theta}^{\dagger}$	$\widehat{\theta}^{*}$
network effects				
$\widehat{\beta}_1 N^{-1}\widehat{F}_{1,t-1}y_{t-1}$	1.1641***	1.0805***	1.0005***	1.0012***
	(0.1220)	(0.0877)	(0.1088)	(0.1088)
$\widehat{\beta}_2 N^{-1}\widehat{F}_{2,t-1}y_{t-1}$	-0.3529**	0.0348	0.1156	0.1158
	(0.1889)	(0.1163)	(0.1133)	(0.1133)
$\widehat{\beta}_3 N^{-1}\widehat{F}_{3,t-1}y_{t-1}$	-0.1217	0.0385	0.0164	0.0162
	(0.1919)	(0.1369)	(0.1442)	(0.1442)
$\widehat{\beta}_4 N^{-1}\widehat{F}_{4,t-1}y_{t-1}$	-0.0597	0.0977	0.1605*	0.1607*
	(0.1911)	(0.1093)	(0.1098)	(0.1098)
$\widehat{\beta}_5 N^{-1}\widehat{F}_{5,t-1}y_{t-1}$	0.8145***	0.3700***	0.5506***	0.5506***
	(0.1893)	(0.1154)	(0.1128)	(0.1128)
$\widehat{\beta}_6 N^{-1}\widehat{F}_{6,t-1}y_{t-1}$	0.2669**	0.0042	0.1050	0.1051
	(0.1552)	(0.0974)	(0.1118)	(0.1118)
momentum effects				
$\widehat{\rho}^{(1)}y^{(1)}_{t-1}$	0.0658*	0.1504***	0.0802**	0.0802**
	(0.0424)	(0.0306)	(0.0350)	(0.0350)
$\widehat{\rho}^{(2)}y^{(2)}_{t-1}$	0.2407***	0.3365***	0.3119***	0.3121***
	(0.0348)	(0.0328)	(0.0277)	(0.0277)
nodal effects				
$\widehat{\alpha}^{(1)}$	0.0023***	0.0875*	0.0021***	0.0021***
	(0.0007)	(0.0548)	(0.0004)	(0.0004)
$\widehat{\alpha}^{(2)}$	0.0054***	0.2741***	0.0049***	0.0049***
	(0.0010)	(0.0729)	(0.0006)	(0.0006)

non-manufacturing sectors (layers 1-3), and financial linkages tend to generate larger spillovers when they take the form of M&A or flows of banking assets (rather than portfolio holdings).

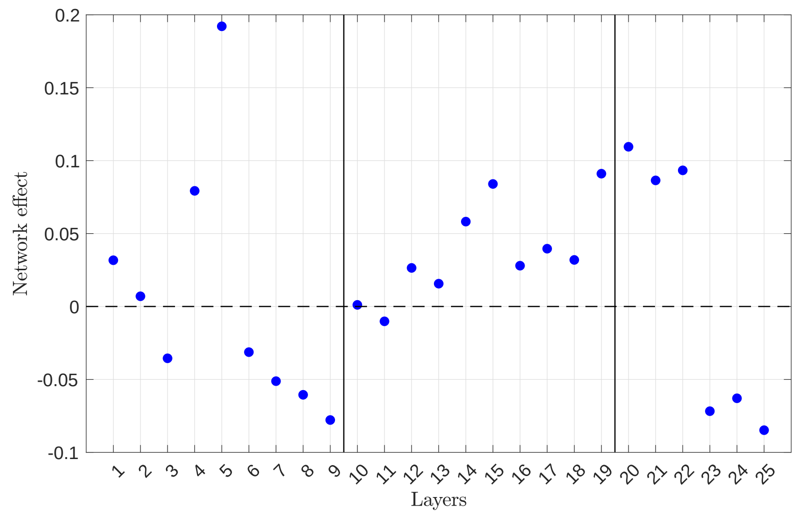
Finally, based on these estimates, we can approximate the network effects associated with the original layers of the network, by appropriately rescaling the estimated network effect coefficients β. Specifically, given the definition of estimated loadings in (4.4) and the properties of tensor multiplication, and letting $\widehat{\mathcal{W}}_t := \widehat{\mathcal{F}}_t \times_3 \widehat{U}$, for a given estimate $\widehat{\beta}$ we have that:

$$\frac{\widehat{\mathcal{F}}_{t-1}}{N} \times_2 y'_{t-1} \times_3 \widehat{\beta}' = \frac{\widehat{\mathcal{F}}_{t-1}}{N} \times_2 y'_{t-1} \times_3 \left(\widehat{\beta}'(\widehat{M}^{\mathcal{W}})^{-1}\widehat{U}'\widehat{U}N^2\right) = \frac{\widehat{\mathcal{W}}_{t-1}}{N} \times_2 y'_{t-1} \times_3 \left(N^2\widehat{U}(\widehat{M}^{\mathcal{W}})^{-1}\widehat{\beta}\right)'.$$

Thus, $N^2\widehat{U}(\widehat{M}^{\mathcal{W}})^{-1}\widehat{\beta}$ is the vector of network effects in terms of the row-normalized tensor $\widehat{\mathcal{W}}_{t-1}/N$ and its entries are shown in Figure 1, when computed using the iterated estimator $\widehat{\theta}^*$. The figure confirms a substantial heterogeneity of effects across layers, reflecting their different loadings on the network factors.

Forecasting. We conclude by studying the performance of our FNAR when producing 1-quarter-ahead forecasts of GDP growth rates based on a recursive window exercise from 2002Q1 to 2019Q4. We consider the following competitors: the tensor-based estimators MLR and SHORR by Wang et al. (2022); a NAR estimated either using the m layers of the common component (denoted as TUCKER COMMON and further regularized via LASSO or Ridge due to collinearity of the regressors) or just r network factors

Figure 1: Network effects by original layer.



The figure plots the estimated network effects associated with the original (and row-normalized) layers of the network computed as $N^2\widehat{U}(\widehat{M}^{\mathcal{W}})^{-1}\widehat{\beta}$, when using the iterated estimator. The vertical lines separate layers related to trade in goods (1-9), layers related to trade in services (10-19) and financial layers (20-25). See Table G.9 for the complete list of layers. The estimation sample is 2001Q1-2019Q4.

(denoted as TUCKER FACTORS) both obtained from a full Tucker factor decomposition, estimated via TOPUP as in Chen et al. (2022) (see Appendix H for more details); a multilayer NAR estimated via LASSO or Ridge; and an ordinary VAR. Details on the adopted forecasting scheme and on the implementation of the alternative estimation methods are in Appendix G.2.4.

Table 4 reports the root mean squared forecast errors (RMSFE) in terms of percentage points of GDP growth, for each country considered. In the last two rows of Table 4, we report the average (across all countries) RMFSE and relative RMSFE (ReRMSFE) with respect to the FNAR, for all forecasting methods (values larger than one indicate a better performance of the FNAR). For most countries, and on average, our approach delivers the forecasts with smallest RMSFEs. In particular, we outperform the approaches based on a full Tucker decomposition which are our most natural competitors. Indeed, contrary to the case of factors extracted by means of a full Tucker decomposition, our network factors still contain terms which are idiosyncratic to the network nodes, i.e., countries, which are potentially relevant for predicting country specific GDP growth rates.

Table 4: RMSFEs for GDP growth

	FNAR	MLR	SHORR	TUCKER COMMON +LASSO	TUCKER COMMON +RIDGE	TUCKER FACTORS	LASSO	RIDGE	VAR
AUS	0.49%	0.46%	0.46%	0.43%	0.45%	0.62%	0.44%	0.45%	0.56%
BEL	0.45%	0.48%	0.50%	0.62%	0.66%	0.48%	0.58%	0.67%	0.58%
CAN	0.46%	0.47%	0.49%	0.61%	0.64%	0.60%	0.58%	0.64%	0.48%
FRA	0.46%	0.48%	0.48%	0.64%	0.69%	0.49%	0.60%	0.70%	0.43%
GER	0.80%	0.83%	0.84%	0.93%	0.96%	0.80%	0.91%	0.96%	0.83%
ITA	0.73%	0.69%	0.70%	0.94%	1.00%	0.72%	0.90%	1.01%	0.57%
JAP	1.05%	1.06%	1.06%	1.16%	1.19%	1.06%	1.13%	1.20%	1.14%
KOR	0.92%	1.09%	1.10%	0.91%	0.91%	1.02%	0.92%	0.91%	1.08%
NLD	0.54%	0.56%	0.58%	0.76%	0.80%	0.56%	0.72%	0.80%	0.56%
NOR	1.27%	1.26%	1.23%	1.25%	1.27%	1.28%	1.25%	1.27%	1.21%
ESP	0.46%	0.56%	0.56%	0.66%	0.71%	0.50%	0.62%	0.72%	0.34%
SWE	1.03%	1.06%	1.05%	1.08%	1.09%	1.02%	1.07%	1.09%	1.16%
CHE	0.78%	0.83%	0.83%	0.85%	0.88%	0.82%	0.83%	0.88%	0.96%
GBR	0.68%	0.72%	0.72%	0.78%	0.82%	0.70%	0.76%	0.82%	0.71%
USA	0.50%	0.55%	0.55%	0.60%	0.63%	0.54%	0.58%	0.63%	0.57%
avg. advanced	*0.71%*	*0.74%*	*0.74%*	*0.81%*	*0.85%*	*0.75%*	*0.79%*	*0.85%*	*0.74%*
BRA	1.27%	1.35%	1.40%	1.33%	1.28%	1.28%	1.36%	1.27%	1.51%
CHN	1.61%	1.15%	1.16%	1.88%	1.95%	1.80%	1.82%	1.96%	1.22%
HKG	1.43%	1.74%	1.82%	1.41%	1.37%	1.39%	1.40%	1.37%	1.69%
IND	1.36%	1.19%	1.18%	1.42%	1.49%	1.55%	1.39%	1.50%	1.51%
IDN	0.48%	0.65%	0.67%	0.54%	0.64%	1.80%	0.50%	0.65%	0.87%
MEX	1.03%	1.01%	1.02%	1.18%	1.13%	0.99%	1.18%	1.13%	1.15%
SAU	1.03%	1.27%	1.23%	1.13%	1.13%	1.07%	1.11%	1.14%	1.24%
ZAF	0.60%	0.58%	0.58%	0.77%	0.67%	0.55%	0.77%	0.66%	0.59%
TUR	2.35%	2.39%	2.41%	2.32%	2.33%	2.38%	2.33%	2.33%	2.30%
avg. emerging	*1.24%*	*1.26%*	*1.28%*	*1.33%*	*1.33%*	*1.31%*	*1.32%*	*1.33%*	*1.34%*
avg. all	*0.91%*	*0.93%*	*0.94%*	*1.01%*	*1.03%*	*0.96%*	*0.99%*	*1.03%*	*0.97%*
avg. ReRMSFE	**1.00**	**1.05**	**1.06**	**1.15**	**1.19**	**1.10**	**1.12**	**1.19**	**1.09**

The table reports the root mean squared forecast errors (RMSFE) of 1-quarter-ahead forecasts, based on a recursive forecasting scheme over the period 2002Q1-2019Q4. Advanced economies are in the upper panel; emerging economies are in the lower panel. See Appendix G.1 for the complete list of countries and their acronyms.

8 Conclusions

In this paper, we have introduced a factor network autoregression (FNAR) for time series characterized by multiple network effects. Estimation is based on two steps. First, we extract few network factors common across the layers of the underlying multilayer network. Second, we estimate a factor-augmented NAR or FNAR where the network effects are determined by the latent network factors. The FNAR errors are allowed to have an underlying factor structure capturing common correlations across the network nodes. We prove consistency and asymptotic normality of the proposed estimators as the number of layers, nodes and time observations diverges to infinity.

The results of an empirical application show that, by accounting for cross-country economic and financial linkages, the model provides a rich description of the dynamics of GDP growth rates and produces accurate forecasts.

We outline three possible extensions of this work, which we leave for further research. First, by adapting the works by Chen et al. (2023) and Zhu et al. (2023) to the FNAR framework, we could consider a FNAR with momentum and nodal coefficients which are group specific, where the group structure is unknown and the number of groups K can grow with the number of nodes N. Second, by generalizing to the tensor setting the three-pass regression filter by Kelly and Pruitt (2015), we could improve the performance of our estimator by accounting also for the information contained in the vector of dependent variables y_t when extracting the network factors. Third, by extending to the tensor case an approach similar to the one proposed by Wu and Zhou (2024), we could allow for time-varying network factor loadings under the standard assumption of local stationarity.

Supplementary Materials

The supplementary materials contain all proofs, additional possible estimators, further simulation results and details on the empirical application.

Funding

All authors gratefully acknowledge financial support from the Italian Ministry of Education, University and Research (PRIN 2017, Grant 2017TA7TYC). Graziano Moramarco gratefully acknowledges funding from the European Union and the Italian Ministry of Education, University and Research (ESF REACT-EU, PON Ricerca e Innovazione 2014-2020).

Disclosure Statement

There are no competing interests to declare.

References

Acemoglu, D., Akcigit, U. and Kerr, W. (2015). 'Networks and the Macroeconomy: An Empirical Exploration', NBER Working Paper 21344.

Acemoglu, D., Carvalho, V.M., Ozdaglar, A. and Tahbaz-Salehi, A. (2012). 'The Network Origins of Aggregate Fluctuations', *Econometrica*, Vol. 80, pp. 1977–2016.

Ahn, S.C. and Horenstein, A.R. (2013). 'Eigenvalue ratio test for the number of factors', *Econometrica*, Vol. 81, pp. 1203–1227.

Bai, J. (2003). 'Inferential theory for factor models of large dimensions', *Econometrica*, Vol. 71, pp. 135–171.

Bai, J. (2009). 'Panel data models with interactive fixed effects', *Econometrica*, Vol. 77, pp. 1229–1279.

Bai, J. and Ng, S. (2002). 'Determining the number of factors in approximate factor models', *Econometrica*, Vol. 70, pp. 191–221.

Bai, J. and Ng, S. (2006). 'Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions', *Econometrica*, Vol. 74, pp. 1133–1150.

Barigozzi, M. and Brownlees, C. (2019). 'NETS: Network Estimation for Time Series', *Journal of Applied Econometrics*, Vol. 34, pp. 347–364.

Barigozzi, M., He, Y., Li, L. and Trapani, L. (2023). 'Statistical Inference for Large-dimensional Tensor Factor Model by Iterative Projections', arXiv:2206.09800.

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*, Vol. 104, pp. 535–559.

Chang, J., He, J., Yang, L. and Yao, Q. (2023). 'Modelling matrix time series via a tensor CP-decomposition', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 85, pp. 127–148.

Chen, E.Y. and Chen, R. (2022). 'Modeling dynamic transport network with matrix factor models: an application to international trade flow', *Journal of Data Science*, Vol. 21, pp. 490–507.

Chen, E.Y., Fan, J. and Zhu, X. (2023). 'Community Network Auto-Regression for High-Dimensional Time Series', *Journal of Econometrics*, Vol. 235, pp. 1239–1256.

Chen, R., Yang, D. and Zhang, C.H. (2022). 'Factor Models for High-Dimensional Tensor Time Series', *Journal of the American Statistical Association*, Vol. 117, pp. 94–116.

Chen, W. and Lam, C. (2024). 'Rank and factor loadings estimation in time series tensor factor model by pre-averaging', *The Annals of Statistics*, Vol. 52, pp. 364–391.

Denbee, E., Julliard, C., Li, Y. and Yuan, K. (2021). 'Network Risk and Key Players: A Structural Analysis of Interbank Liquidity', *Journal of Financial Economics*, Vol. 141, pp. 831–859.

Diebold, F.X. and Yilmaz, K. (2014). 'On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms', *Journal of Econometrics*, Vol. 182, pp. 119–134.

Han, Y., Chen, R., Yang, D. and Zhang, C.H. (2020). 'Tensor Factor Model Estimation by Iterative Projection', arXiv:2006.02611.

Han, Y., Chen, R. and Zhang, C.H. (2022). 'Rank determination in tensor factor model', *Electronic Journal of Statistics*, Vol. 16, pp. 1726–1803.

Han, Y., Yang, D., Zhang, C.H. and Chen, R. (2024). 'CP factor model for dynamic tensors', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 86, pp. 1383–1413.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2021). 'Recursive estimation in large panel data models: Theory and practice', *Journal of Econometrics*, Vol. 224, pp. 439–465.

Kelly, B. and Pruitt, S. (2015). 'The three-pass regression filter: A new approach to forecasting using many predictors', *Journal of Econometrics*, Vol. 186, pp. 294–316.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y. and Porter, M.A. (2014). 'Multilayer Networks', *Journal of Complex Networks*, Vol. 2, pp. 203–271.

Pesaran, M.H., Schuermann, T. and Weiner, S.M. (2004). 'Modeling Regional Interdependencies Using a Global Error Correcting Macroeconometric Model', *Journal of Business and Economic Statistics*, Vol. 22, pp. 129–62.

Stock, J.H. and Watson, M.W. (2002). 'Forecasting Using Principal Components From a Large Number of Predictors', *Journal of the American Statistical Association*, Vol. 97, pp. 1167–1179.

Trapani, L. (2018). 'A randomized sequential procedure to determine the number of factors', *Journal of the American Statistical Association*, Vol. 113, pp. 1341–1349.

Wang, D., Zheng, Y. and Li, G. (2024). 'High-dimensional low-rank tensor autoregressive time series modeling', *Journal of Econometrics*, Vol. 238, p. 105544.

Wang, D., Zheng, Y., Lian, H. and Li, G. (2022). 'High-Dimensional Vector Autoregressive Time Series Modeling via Tensor Decomposition', *Journal of the American Statistical Association*, Vol. 117, pp. 1593–1636.

Wu, W. and Zhou, Z. (2024). 'Adaptive estimation for non-stationary factor models and a test for static factor loadings', arXiv:2012.14708.

Zhang, X., Li, G. and Liu, C.C. (2022). 'Tucker Tensor Factor Models for High-Dimensional Higher-Order Tensor Observations', arXiv:2206.02508.

Zhou, J., Li, D., Pan, R. and Wang, H. (2020). 'Network GARCH model', *Statistica Sinica*, Vol. 30, pp. 1723–1740.

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017). 'Network Vector Autoregression', *The Annals of Statistics*, Vol. 45, pp. 1096–1123.

Zhu, X., Xu, G. and Fan, J. (2023). 'Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes', *Journal of Econometrics*, p. 105564.