NETS:
NETWORK ESTIMATION FOR TIME SERIES

Matteo Barigozzi[‡] Christian Brownlees[†]

October 15, 2018

Abstract

We model a large panel of time series as a VAR where the autoregressive matrices and the inverse covariance matrix of the system innovations are assumed to be sparse. The system has a network representation in terms of a directed graph representing predictive Granger relations and an undirected graph representing contemporaneous partial correlations. A LASSO algorithm called NETS is introduced to estimate the model. We apply the methodology to analyse a panel of volatility measures of ninety bluechips. The model captures an important fraction of total variability, on top of what is explained by volatility factors, and improves out-of-sample forecasting.

Keywords: Networks, Multivariate Time Series, VAR, LASSO, Forecasting

JEL: C01, C32, C52

[‡] Department of Statistics, London School of Economics and Political Science,
Houghton Street, WC2A 2AE, London, UK, e-mail: `m.barigozzi@lse.ac.uk`;
[†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE,
Ramon Trias Fargas 25-27, 08005, Barcelona, Spain, e-mail: `christian.brownlees@upf.edu`.
The procedures presented in this paper are available in the package `nets` for `R`.
Christian Brownlees acknowledges financial support from Spanish Ministry of Science and Technology (Grant MTM2015-67304-P) and from Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075). We would like to thank Vasco Carvalho, Frank Diebold, Tony Hall, Nikolaus Hautsch, Julien Idier, Clifford Lam, Marco Lippi, Gabor Lugosi, Miguel Morin, Tommaso Proietti, and Kimmo Soramaki for their helpful comments. Special thanks go to Giorgio Calzolari. All mistakes are ours.

1 Introduction

Over the last years, network analysis has become an active topic of research in time series econometrics, with numerous applications in macroeconomics and finance. Example of contributions in the literature include, *inter alia*, Billio et al. (2012), Diebold and Yılmaz (2014, 2015), Hautsch et al. (2014), and Härdle et al. (2016). In a nutshell, network analysis is concerned with representing the interconnections of a large panel as a graph: the vertices of the graph represent the variables in the panel, and the presence of an edge between two vertices denotes the presence of some appropriate measure of dependence between the two variables. From an economic perspective, the interest on networks has been boosted by the research of, *inter alia*, Acemoglu et al. (2012), which shows that individual entities can have a non-negligible effect on the aggregate behavior of the economy when the system has a high degree of interconnectedness.

In this paper we propose network methodology for large panels of time series. We model the panel as a Vector Autoregression (VAR). We work under the assumption that the VAR is sparse, in the sense that the autoregressive matrices and the inverse covariance matrix of the system innovations are assumed to be sparse. Notice that the notion of sparsity used in this work is different from the one used in other papers such as Davis et al. (2015), Kock and Callot (2015), and Medeiros and Mendes (2016) where sparsity assumptions are formulated for the autoregressive matrices only. Sparsity of the autoregressive matrices implies sparsity of the multivariate Granger causality structure of the system whereas sparsity of the inverse covariance matrix implies sparsity of the partial correlation structure (Dempster, 1972; Lauritzen, 1996).

Several network representations can be associated with a VAR system (Dahlhaus, 2000; Eichler, 2007; Diebold and Yılmaz, 2014). In this work we focus on two representations that are natural for the sparse VAR we introduce in this work. The first network representation consists of representing the system as a mixed graph containing both directed and undirected edges: directed edges denote Granger causality linkages among time series while

undirected edges represent contemporaneous partial correlation linkages. The second network representation we introduce is an undirected graph where edges denote long run partial correlation linkages among time series. Long run partial correlation is a partial correlation measure constructed on the basis of the long run covariance matrix of the VAR. It synthesises simultaneously lead/lag and contemporaneous dependence among time series and is a natural generalization for dependent data of the standard partial correlation model used in the statistics graphical literature.

In order to estimate large sparse VARs we introduce a novel LASSO-based algorithm. The highlight of the procedure is that it simultaneously estimates the autoregressive matrices as well as the entries of the concentration matrix, avoiding to split up the estimation of the model parameters in two steps. The large sample properties of the proposed estimator are analysed and we establish conditions for consistent selection and estimation of the VAR parameters. The theory is derived in a high-dimensional setting, allowing the number of series in the system to increase with the sample size. Specifically, the number of series is allowed to be $O(T^{\zeta})$ for $\zeta > 0$ where T denotes the sample size of the panel.

The network methodology we introduce in this work has highlights in terms of interpretation and estimation. Understanding and synthesising the interdependence structure of a large multivariate system can be a daunting task. The network representation of the panel provides a parsimonious synthesis of the data that can bring useful insights on their underlying structure. From an estimation perspective, carrying out inference on the VAR parameters can be challenging when the number of time series is large. The regularized estimation approach based on LASSO put forward in this work can lead to substantial gains in terms of estimation precision and, ultimately, forecasting. The gains of the methodology rely on sparsity assumptions of the underlying system. It is important to emphasize these assumptions may not be always appropriate. In practice, users ought to check whether these are plausible for the data at hand and eventually, when possible, appropriately transform their data.

A natural application of network analysis techniques is the study of interdependence in panels of volatility measures. Detecting the interconnectedness structure of volatility panels

is of interest to understand and monitor the risk transmission channels of these systems. See, for instance, the research of Diebold and Yılmaz (2014) on risk transmission in the 2007–2009 Great financial crisis or Engle et al. (2012) in the 1997–1998 Asian financial crisis. We use the methodology derived in this work to analyze a panel of volatility measures for ninety US bluechips across different industry groups from January 2nd 2004 to December 31st 2015. An important feature of our application is that we study interconnectedness conditional on a market wide and sector specific volatility factors. We show that after conditioning on the factors the sparse VAR captures approximately 10% of the overall variability. The estimated networks connect the vast majority of the series in the panel and the interdependence is positive in the vast majority of cases. Results show that the financial sector is the most interconnected industry in this sample period. In particular, large financial institutions, such as AIG, Bank of America and Citigroup, are some of the most interconnected entities in the panel. An out-of-sample forecasting exercise is used to validate the methodology proposed in our work and shows that the sparse VAR model improves predictive ability over a number of benchmarks.

Our work relates to different strands of literature. First, it is related to the econometric literature on networks, which includes research by Billio et al. (2012), Diebold and Yılmaz (2014, 2015), Hautsch et al. (2014), Anufriev and Panchenko (2015), Härdle et al. (2016), Hagströmer and Menkveld (2016). Early influential work on networks and panels of time series includes Mantegna (1999). This paper is also related to the literature on the estimation of sparse VARs, see Song and Bickel (2011), Davis et al. (2015), Kock and Callot (2015), Medeiros and Mendes (2016) and Basu and Michailidis (2015). Our contribution also relates to the statistical literature on large dimensional network estimation based on LASSO techniques. Contributions in this area include, *inter alia*, Meinshausen and Bühlmann (2006), Friedman et al. (2008), Peng et al. (2009).

The paper is structured as follows. Section 2 introduces the model, the network definitions and the estimation strategy. Section 3 contains a simulation study that analyzes the finite sample properties of the procedure. Section 4 contains the empirical application. Concluding

remarks follow in Section 5. The proofs of all our results and are in a complementary appendix available online.

2 Methodology

2.1 Model

We consider a zero-mean stationary n-dimensional multivariate time series $\mathbf{y}_t = (y_{1\,t}, \ldots, y_{n\,t})'$ generated by a p-th order VAR

$$\mathbf{y}_t = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{y}_{t-k} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim i.i.d.(\mathbf{0}, \mathbf{C}^{-1}), \tag{1}$$

where $\mathbf{A}_k$ and $\mathbf{C}$ are $n \times n$ matrices. Throughout the VAR is assumed to be stable and $\mathbf{C}$ to be positive definite. Notice that for convenience the distribution of the innovation terms is parametrized with the inverse covariance matrix $\mathbf{C}$, also known as concentration matrix, rather than the covariance. The (i, j)-th entries of the matrices $\mathbf{A}_k$ and $\mathbf{C}$ are denoted respectively as $a_{k\,ij}$ and c_{ij}.

In this work we focus on the analysis of sparse VAR systems, in the sense that the autoregressive matrices $\mathbf{A}_k$ and the concentration matrix $\mathbf{C}$ are assumed to be sparse matrices. More specific notions of sparsity are spelled out in Section 2.4, where precise assumptions are required by the estimation theory to establish the results of interest. In general, the sparsity assumption can be interpreted as a sparsity assumption on the lead/lag and contemporaneous dependence structure of the system.

The standard notion of dynamic interdependence used for time series is Granger causality. In this work we rely on a multivariate version of this concept. Formally, we say that $y_{j\,t}$ does not Granger cause $y_{i\,t}$ if adding $y_{j\,t}$ as predictor does not improve the mean square forecast error of $y_{i\,t+k}$ for any $k > 0$, that is

$$\mathsf{E}[(y_{i\,t+k} - \mathsf{E}(y_{i\,t+k}|\{y_{1\,t} \ldots y_{n\,t}\}))^2] = \mathsf{E}[(y_{i\,t+k} - \mathsf{E}(y_{i\,t+k}|\{y_{1\,t} \ldots y_{n\,t}\} \setminus y_{j\,t}))^2]. \tag{2}$$

It is immediate to see that the Granger causality structure of the model is encoded in the sparsity structure of the autoregressive matrices $\mathbf{A}_k$. We have indeed that if $a_{k\,ij} = 0$, for all k, then $y_{j\,t}$ does not Granger cause $y_{i\,t}$.

The classical measure of contemporaneous dependence used in the network literature is partial correlation. In this paper we consider partial correlation between two series conditional on the past realizations of the panel and contemporaneous realizations of the remaining series. This is encoded in the partial correlation between VAR innovations, which is denoted as

$$\rho^{ij} = \mathsf{Cor}(\epsilon_{i\,t}, \epsilon_{j\,t} | \{\epsilon_{k\,t} : k \neq i, j\}). \tag{3}$$

It is well known that partial correlations are related to the entries c_{ij} of the concentration matrix $\mathbf{C}$ by means of the relation (Dempster, 1972)

$$\rho^{ij} = -\frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \tag{4}$$

Thus, the contemporaneous dependence sparsity structure is embedded in the sparsity structure of the concentration matrix $\mathbf{C}$. Indeed, if $c_{ij} = 0$, then series i and j are contemporaneously uncorrelated conditional on all other series in the system.

Networks are a useful tool to represent the interdependence structure of the time series in the panel $\mathbf{y}_t$. A network is defined as a graph $\mathcal{N} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of edges. The set of vertices $\mathcal{V}$ is $\{1, ..., n\}$ where each element corresponds to a component of $\mathbf{y}_t$, while the set of edges $\mathcal{E}$ is a subset of $\mathcal{V} \times \mathcal{V}$ such that the pair (i, j) is in $\mathcal{E}$ if and only if the components i and j are linked by an edge. Several network representations can be associated with a VAR system. In this work we focus on two representations that are natural for the sparse VAR we introduce in this work.

A natural representation of the sparse VAR model is based on the union of two graphs: the first graph contains directed edges denoting Granger causality linkages, whereas the second graph contains undirected edges representing contemporaneous partial correlation

linkages. We label the two networks respectively as the Granger and contemporaneous networks. The Granger network is defined as a directed network $\mathcal{N}_G = (\mathcal{V}, \mathcal{E}_G)$ where the presence of an edge from i to j denotes that i Granger causes j in the sense of (2), that is $\mathcal{E}_G = \{(i,j) \in \mathcal{V} \times \mathcal{V} : a_{k\,ij} \neq 0, \text{ for at least one } k \in \{1, ..., p\}\}$. The contemporaneous network is defined as an undirected network $\mathcal{N}_C = (\mathcal{V}, \mathcal{E}_C)$ where an edge between i and j denotes that i is partially correlated to j, that is $\mathcal{E}_C = \{(i,j) \in \mathcal{V} \times \mathcal{V} : \rho^{ij} \neq 0\}$.

An alternative way to represent the dependence structure of the system using a single graph consists in simultaneously summarising the lead/lag and contemporaneous relations of the process. We do this by introducing a partial correlation measure based on the long run covariance matrix. This is inspired by the HAC literature (Den Haan and Levin, 1996) and is a natural extension of the standard partial correlation network to serially dependent data. The long run covariance matrix of the process $\mathbf{y}_t$ is defined as

$$\mathbf{\Sigma}_L = \lim_{M \to \infty} \frac{1}{M} \mathsf{Cov} \left(\sum_{t=1}^{M} \mathbf{y}_t, \sum_{t=1}^{M} \mathbf{y}_t \right) .$$

Equivalently, the long run covariance may also be defined in terms of the sum of all autocovariance functions of the process $\mathbf{\Sigma}_L = \sum_{h=-\infty}^{+\infty} \mathsf{Cov}(\mathbf{y}_t, \mathbf{y}_{t-h})$, which shows how $\mathbf{\Sigma}_L$ synthesises the linear dependences of $\mathbf{y}_t$ at every lead and lag. Note that since the VAR is assumed to be stationary the sum above is well defined. The long run covariance is therefore the spectral density matrix of $\mathbf{y}_t$ at zero-frequency, which in the case of a VAR is given by

$$\mathbf{\Sigma}_L = \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k \right)^{-1} \mathbf{C}^{-1} \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k' \right)^{-1} .$$

We propose a network definition based on the partial correlations constructed on the basis of the long run concentration matrix which is defined as

$$\mathbf{K}_L = \mathbf{\Sigma}_L^{-1} = \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k' \right) \mathbf{C} \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k \right) .$$

This is also known as the zero-frequency partial spectral coherence (Dahlhaus, 2000; Davis et al., 2015). Notice that $\mathbf{K}_L$ is factorized in a sandwich form determined by the term $\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k$, which captures long run dynamic relations of the system, and the term $\mathbf{C}$, which accounts for the contemporaneous dependence of the system innovations. We can then express long run partial correlation coefficient for series i and j as a function of the entries $k_{L\,ij}$ of the long run concentration matrix $\mathbf{K}_L$

$$\rho_L^{ij} = -\frac{k_{L\,ij}}{\sqrt{k_{L\,ii} k_{L\,jj}}}.$$

The long run partial correlation network is then defined as a undirected network $\mathcal{N}_L = (\mathcal{V}, \mathcal{E}_L)$ where the set of edges $\mathcal{E}_L$ is defined as $\mathcal{E}_L = \left\{(i,j) \in \mathcal{V} \times \mathcal{V} \,:\, \rho_L^{ij} \neq 0\right\}$.

A number of comments on the model and network definitions we propose are in order. First, an important difference between the network modelling approach proposed here and other contributions in the literature is that we focus on representing the partial dependence structure of the panel. On the other hand, the contributions of, *inter alia*, Billio et al. (2012) and Diebold and Yılmaz (2014) propose network definitions that measure the overall degree of dependence between series. The advantage of the approach proposed here is that it is robust to spurious correlation effects among the variables in the system. Moreover, the network definitions we propose can be seen as natural extension for time series data of the popular partial correlation network models used in the statistics and graphical models literatures.

Second, the network representations we propose are useful when the underlying dependence structure of the system is sparse. Sparsity assumptions (that are be spelled out precisely in what follows) are also key to carry out inference in large dimensional systems. It is important to acknowledge however that a number of contributions have criticized these types of assumptions on the grounds that they may not be appropriate for applications in economics and finance. Giannone et al. (2018) analyze different economic datasets and conclude that, in general, sparsity is not a feature of the data and should be assumed only if there is enough a priori evidence in favor of predictive models with a small number of regressors.

Third, and related to the previous point, in this work we consider network analysis as a complement of factor analysis for the purpose of empirical applications. Indeed, an important case in which the assumption of sparsity is violated, is when the components of the panel are a function of a set of common factors (Forni et al., 2000; Stock and Watson, 2002; Bai, 2003). In particular, it is straightforward to see that global common factors induce a fully interconnected network structure among the variables in the panel.[1] Note also that the dependencies we are interested in, being conditional, cannot in general be captured by adding a small number of additional sectoral factors, as proposed for example by Foerster et al. (2011). The presence of common factors in panels of volatilities, similar to those analyzed in this paper, is documented for example in Barigozzi et al. (2014) and Barigozzi and Hallin (2016). On the other hand, Barigozzi and Hallin (2017) show evidence of weak cross-sectional dependence once common factors are removed. As a consequence of these observations, we suggest that the influence of global and sectoral factors ought to be filtered out before carrying out network analysis. Specifically, our analysis in Section 4 shows that, even after controlling for global and sectoral factors in financial markets, network effects, in the form of conditional dependencies, still have an important predictive role.

Last, we point out that an alternative approach for the estimation of the long run partial correlation network consists in estimating the long run concentration matrix and then applying LASSO-type regularization (under the assumption of sparsity of $\mathbf{K}_L$). The advantage of estimating $\mathbf{K}_L$ directly is that one may rely on non-parametric long run covariance estimators, which are consistent for a large class of DGPs. In practice, however, when it is known that two series have non-zero long run (partial) correlation it is natural to ask whether this dependence arises because of lagged or contemporaneous effects (or both). To this extent a VAR is a natural modeling approach that allows to disentangle these different channels.

[1] Consider an n–dimensional panel of time series y_{it} generated by a one factor model $y_{it} = \beta_i f_t + \epsilon_{it}$, where f_t and ϵ_{it} are independent normals with zero mean and unit variance and $\epsilon_{i\,t}$ and $\epsilon_{j\,t}$ are independent for each $i \neq j$. Then the concentration matrix of the system is $\mathbf{K} = \mathbf{I}_n - \frac{1}{1+\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'$, where $\mathbf{I}_n$ is the identity matrix of size $n \times n$ and $\boldsymbol{\beta}$ is a $n \times 1$ vector of factor loadings β_i. If the vector of factor loading does not contain zero entries then $\mathbf{K}$ is not sparse.

2.2 Estimation

We are interested in detecting and estimating the non-zero entries of the autoregressive matrices $\mathbf{A}_k$ and the concentration matrix $\mathbf{C}$. A simple estimation approach for the sparse VAR would consist of using LASSO regression to estimate the autoregressive matrices $\mathbf{A}_k$ (as for example in Kock and Callot, 2015), and then using a LASSO procedure on the residuals to estimate the concentration matrix $\mathbf{C}$ (as for example in Friedman et al., 2008; Peng et al., 2009). The analysis of properties of the second step estimator is however challenging. Moreover, the rate of convergence of the estimator of the concentration matrix $\mathbf{C}$ would depend on the rate of convergence of the estimator of the autoregressive matrices $\mathbf{A}_k$.[2] In this work we propose an estimation approach that avoids these hurdles by estimating both sets of parameters jointly.

For ease of notation, we re-parametrize the VAR as a function of: (i) the coefficients α_{ijk} contained in an n^2p-dimensional vector $\boldsymbol{\alpha}$ which correspond to the autoregressive coefficients $a_{k\,ij}$ in (1), (ii) the partial correlations ρ^{ij} contained in an $n(n-1)/2$-dimensional vector $\boldsymbol{\rho}$ and defined in (3), and (iii) the coefficients c_{ii} contained in an n-dimensional vector $\mathbf{c}$ which correspond to the diagonal of the concentration matrix $\mathbf{C}$. Then, in scalar notation the parameters of our model are given by the VAR equations

$$y_{i\,t} = \sum_{k=1}^{p} \sum_{j=1}^{n} \alpha_{ijk}\, y_{j\,t-k} + \epsilon_{i\,t}, \quad i = 1, \ldots, n, \tag{5}$$

and the contemporaneous equations (see Peng et al., 2009)

$$\epsilon_{i\,t} = \sum_{\substack{h=1 \\ h \neq i}}^{n} \rho^{ih} \sqrt{\frac{c_{hh}}{c_{ii}}} \epsilon_{h\,t} + u_{i\,t}, \quad i = 1, \ldots, n, \tag{6}$$

where $u_{i\,t}$ is an error term uncorrelated with $\epsilon_{h\,t}$ for $i \neq h$.

In this section we define a novel LASSO-based estimator for the parameters of (5) and (6). We call the estimation algorithm `nets` (Network Estimator for Time Series) and we

[2] This is shown in a previous working paper version of this manuscript.

describe it in detail in the next section. The main feature of the proposed procedure is that it estimates the autoregressive parameters, $\boldsymbol{\alpha}$, and partial correlations, $\boldsymbol{\rho}$, simultaneously, conditional on a pilot estimator of $\mathbf{c}$.

Consider the following regression representation of y_{it} as a function of the lags of all series as well as the contemporaneous realizations of all other series in the panel, that is

$$y_{it} = \sum_{k=1}^{p}\sum_{j=1}^{n} \beta_{ijk}\, y_{j\,t-k} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\, y_{ht} + e_{it}, \tag{7}$$

where e_{it} is an error term. It is straightforward to see that (see Lemma A1 in the online appendix) the β_{ijk} and γ_{ih} coefficients can be expressed as a function of the α_{ijk}, ρ^{ih} and c_{ii} parameters. In particular, (7) can be re-written as

$$y_{it} = \sum_{k=1}^{p}\sum_{j=1}^{n} \underbrace{\left(\alpha_{ijk} - \sum_{\substack{l=1\\l\neq i}}^{n} \rho^{il}\sqrt{\frac{c_{ll}}{c_{ii}}}\alpha_{ljk}\right)}_{\beta_{ijk}} y_{j\,t-k} + \sum_{\substack{h=1\\h\neq i}}^{n} \underbrace{\rho^{ih}\sqrt{\frac{c_{hh}}{c_{ii}}}}_{\gamma_{ih}}\, y_{ht} + u_{it}. \tag{8}$$

Notice that this result also shows that the errors e_{it} and u_{it} are the same. We denote by $\boldsymbol{\theta}$ the vector of parameters of interest $(\boldsymbol{\alpha}', \boldsymbol{\rho}')'$ of dimension $m = n^2p + n(n-1)/2$. The regression representation in (8) suggests to associate the following quadratic loss function to the problem of determining $\boldsymbol{\theta}$, conditional on $\mathbf{c}$,

$$\ell(\boldsymbol{\theta}; \mathbf{y}_t, \mathbf{c}) = \sum_{i=1}^{n}\left(y_{it} - \sum_{k=1}^{p}\sum_{j=1}^{n}\left(\alpha_{ijk} - \sum_{\substack{l=1\\l\neq i}}^{n}\rho^{il}\sqrt{\frac{c_{ll}}{c_{ii}}}\alpha_{ljk}\right) y_{j\,t-k} - \sum_{\substack{h=1\\h\neq i}}^{n}\rho^{ih}\sqrt{\frac{c_{hh}}{c_{ii}}}y_{ht}\right)^2. \tag{9}$$

If a sample of T observations of the $\mathbf{y}_t$ process is available for $t = 1, \ldots, T$, then we propose to estimate the model parameters using a LASSO-type estimator

$$\widehat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^m}\left[\frac{1}{T}\sum_{t=1}^{T}\ell(\boldsymbol{\theta}; \mathbf{y}_t, \widehat{\mathbf{c}}_T) + \lambda_T^{\mathbf{G}}\sum_{k=1}^{p}\sum_{i,j=1}^{n}\frac{|\alpha_{ijk}|}{|\widetilde{\alpha}_{T\,ijk}|} + \lambda_T^{\mathbf{C}}\sum_{\substack{l,h=1\\l>h}}^{n}\frac{|\rho^{lh}|}{|\widetilde{\rho}_T^{lh}|}\right], \tag{10}$$

where $\lambda_T^{\mathbf{G}} > 0$ and $\lambda_T^{\mathbf{C}} > 0$ are the LASSO tuning parameters and $\widetilde{\boldsymbol{\alpha}}_T$, $\widetilde{\boldsymbol{\rho}}_T$ and $\widehat{\mathbf{c}}_T$ are pre-estimators of the $\boldsymbol{\alpha}$, $\boldsymbol{\rho}$, and $\mathbf{c}$ coefficients, respectively. Due to the presence of autocorrelation in and across the components of $\mathbf{y}_t$, the regressors in (8) are likely to be dependent, therefore we adopt here an adaptive LASSO penalty as originally proposed by Zou (2006) and then studied in a high-dimensional setting by Huang et al. (2008) for independent data and by Kock (2016) for dependent data. If the sample size is sufficiently large, a natural pre-estimator of $\boldsymbol{\alpha}$ is the least squares estimator of the VAR autoregressive matrices while the pre-estimator of $\boldsymbol{\rho}$ is the partial correlation estimator obtained from the sample covariance of the VAR residuals. Otherwise, if the sample size is not sufficiently large a pre-estimator of $\boldsymbol{\alpha}$ could be obtained by estimating the autoregressive matrices via LASSO or ridge regression, while the pre-estimator of $\boldsymbol{\rho}$ could be obtained from a shrinkage estimator of the residual covariance (Ledoit and Wolf, 2004). A discussion of the implications of these choices for our theory is in Section 2.4. Last, a possible choice for the pre-estimator of $\mathbf{c}$ is the reciprocal of the variance of each series (Peng et al., 2009).

2.3 The `nets` algorithm

In this section we introduce the `nets` algorithm to solve the optimization problem of equation (10). Notice that the loss function in (9) is not the standard quadratic loss function of a linear regression model and the standard LASSO algorithms cannot be applied. However, it is still possible to design a coordinate descent algorithm that can be used to minimize the objective function of (10). The procedure we propose is a generalization of the `space` algorithm proposed by Peng et al. (2009) for the estimation of partial correlation networks, and it is a variation of the shooting algorithm by Fu (1998) typically used for LASSO optimization.

Additional notation is required to describe the algorithm. We begin by introducing the matrix representation of the model of equation (8) obtained by stacking the time series in the panel. Let $\boldsymbol{\mathcal{Y}}$ denote a $nT \times 1$ vector defined as $(y_{1\,1}, ..., y_{1\,T},\ ...,\ y_{i\,1}, ..., y_{i\,T},\ ...,\ y_{n\,1}, ..., y_{n\,T})'$; let $\boldsymbol{\mathcal{X}}_G = (\boldsymbol{x}_{G\,111}, ..., \boldsymbol{x}_{G\,ijk}, ..., \boldsymbol{x}_{G\,nnp})$ be a $nT \times n^2p$ matrix with (i, j, k)-th column defined

as

$$\boldsymbol{x}_{G\,ijk} = (\,0, ..., 0, \underbrace{y_{j\,-k}, ..., y_{j\,t-k}, ..., y_{j\,T-k}}_{i\text{-th block}}, 0, ..., 0)',$$

and let $\boldsymbol{\mathcal{X}}_C = (\boldsymbol{x}_{C\,21}, \boldsymbol{x}_{C\,31}, \boldsymbol{x}_{C\,32}, ..., \boldsymbol{x}_{C\,ij}, ..., \boldsymbol{x}_{C\,n(n-1)})$ be a $nT \times n(n-1)/2$ matrix with (i,j)-th column defined as

$$\boldsymbol{x}_{C\,ij} = \left(0, ..., 0, \underbrace{\sqrt{\frac{c_{jj}}{c_{ii}}}(y_{j\,1}, ..., y_{j\,t}, ..., y_{j\,T})}_{i\text{-th block}}\ 0, ..., 0, \underbrace{\sqrt{\frac{c_{ii}}{c_{jj}}}(y_{i\,1}, ..., y_{i\,t}, ..., y_{i\,T})}_{j\text{-th block}}\ 0, ..., 0 \right)'.$$

Then, it is straightforward to check that model (8) can be represented as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_G\,\boldsymbol{\beta}(\boldsymbol{\alpha}, \boldsymbol{\rho}) + \boldsymbol{\mathcal{X}}_C\,\boldsymbol{\rho} + \boldsymbol{\mathcal{U}},$$

where $\boldsymbol{\mathcal{U}}$ is a $nT \times 1$ vector of residuals and $\boldsymbol{\beta}(\cdot, \cdot)$ denotes the function which maps the $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ parameter vectors onto the $\boldsymbol{\beta}$ parameter vector, whose components are given in (7). Notice that the parameter $\boldsymbol{\beta}$ and the matrix $\boldsymbol{\mathcal{X}}_C$ depend implicitly on the parameter $\mathbf{c}$ and that in the estimation we set this to the pre-estimator $\widehat{\mathbf{c}}_T$. The dependence on $\mathbf{c}$ is suppressed in the notation for simplicity. In what follows it is convenient to introduce shorthand notation for the stacked vectors. Let $\boldsymbol{v}$ be a $nT \times 1$ stacked vector, then we use $v_{[it]}$ to refer to the t-th element of the i-th block of $\boldsymbol{v}$.

The `nets` algorithm is an iterative coordinate descent procedure for the minimization of the objective function of (10). Each iteration s of the algorithm updates one component of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\rho}')'$. The $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ parameter estimates at iteration s are denoted as $\widehat{\boldsymbol{\alpha}}^{(s)}$ and $\widehat{\boldsymbol{\rho}}^{(s)}$ respectively. We define the residual estimate at iteration s as

$$\widehat{\boldsymbol{\mathcal{U}}}^{(s)} = \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}_G\,\boldsymbol{\beta}(\widehat{\boldsymbol{\alpha}}^{(s)}, \widehat{\boldsymbol{\rho}}^{(s)}) - \boldsymbol{\mathcal{X}}_C\,\widehat{\boldsymbol{\rho}}^{(s)}\ .$$

The algorithm iterates until convergence, which is checked at the end of each full cycle of updates of $\boldsymbol{\theta}$. To describe the algorithm, it is useful to use two auxiliary $nT \times 1$ stacked vectors

$\ddot{\boldsymbol{x}}$ and $\ddot{\boldsymbol{y}}$. The $\ddot{\boldsymbol{x}}$ vector denotes the regressors corresponding to the current coefficient being updated, while the $\ddot{\boldsymbol{y}}$ vector is the partial residual of the model with respect to all parameter besides the coefficient being currently updated (either α_{ijk} or ρ^{ij}). The α_{ijk} coefficient is updated as

$$\widehat{\alpha}_{ijk}^{(s)} = \text{sign}\left(\ddot{\boldsymbol{y}}'\ddot{\boldsymbol{x}}\right)\left(\left|\frac{\ddot{\boldsymbol{y}}'\ddot{\boldsymbol{x}}}{\ddot{\boldsymbol{x}}'\ddot{\boldsymbol{x}}}\right| - \frac{\lambda_T^{\mathbf{G}}}{\widetilde{\alpha}_{ijk}}\frac{1}{\ddot{\boldsymbol{x}}'\ddot{\boldsymbol{x}}}\right)_+,$$

where $\ddot{\boldsymbol{x}}$ and $\ddot{\boldsymbol{y}}$ are defined as

$$\begin{aligned}
\ddot{x}_{[lt]} &= \begin{cases} y_{l\,t-k}, & \text{if } l = i \\ -\widehat{\rho}^{\,il\,(s-1)}\sqrt{\frac{\widetilde{c}_{ll}}{\widetilde{c}_{ii}}}\,y_{j\,t-k} & \text{otherwise} \end{cases} \\
\ddot{y}_{[lt]} &= \mathcal{U}_{[lt]}^{(s-1)} + \widehat{\alpha}_{ijk}^{(s-1)}\ddot{x}_{[lt]},
\end{aligned}$$

for each $l = 1, ..., n$ and $t = 1, ..., T$. The ρ^{ij} coefficient is updated as

$$\widehat{\rho}^{\,ij\,(s)} = \text{sign}\left(\ddot{\boldsymbol{y}}'\ddot{\boldsymbol{x}}\right)\left(\left|\frac{\ddot{\boldsymbol{y}}'\ddot{\boldsymbol{x}}}{\ddot{\boldsymbol{x}}'\ddot{\boldsymbol{x}}}\right| - \frac{\lambda_T^{\mathbf{C}}}{\widetilde{\rho}^{ij}}\frac{1}{\ddot{\boldsymbol{x}}'\ddot{\boldsymbol{x}}}\right)_+,$$

where $\ddot{\boldsymbol{x}}$ and $\ddot{\boldsymbol{y}}$ are defined as

$$\begin{aligned}
\ddot{x}_{[lt]} &= \sqrt{\frac{\widetilde{c}_{hh}}{\widetilde{c}_{ll}}}\left(y_{h\,t} - \sum_{j=1}^{n}\sum_{k=1}^{p}\alpha_{hjk}^{(s-1)}y_{j\,t-k}\right) \\
\ddot{y}_{[lt]} &= \mathcal{U}_{[lt]}^{(s-1)} + \widehat{\rho}^{\,ij\,(s-1)}\ddot{x}_{[lt]},
\end{aligned}$$

for (l, h) equal (i, j) or (j, i) and $t = 1, ..., T$, and otherwise $\ddot{y}_{[lt]}$ and $\ddot{x}_{[lt]}$ are set to zero. It is important to stress that the parameter vector $\boldsymbol{\theta}$ contains $n^2p + \frac{n(n-1)}{2}$ elements, whose optimization would require large amounts of memory to be stored when the panel is large. On the other hand, the coordinate wise minimization algorithm is appealing in this context in that it has mild storage requirements and can be applied in large dimensional applications.

As far as the estimation of $\mathbf{c}$ is concerned, we follow the two-step iterative procedure proposed in Peng et al. (2009): (i) given an estimate of $\mathbf{c}$, we estimate the $\boldsymbol{\theta}$ parameter using

`nets`; (ii) given an estimate $\mathbf{c}$ and an estimate of $\boldsymbol{\theta}$ we update the estimate of $\mathbf{c}$. Notice, that c_{ii} is the reciprical of the residual variance of equation (8). These two steps are then iterated until convergence, which typically kicks in within very few iterations.

2.4 Theory

We establish the estimation and selection consistency of our LASSO estimator. In this section we denote by $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\rho}_0')'$ and $\mathbf{c}_0$ the true value of the parameters, while their generic values are denoted as $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\rho}')'$ and $\mathbf{c}$. The estimator defined in (10) can be equivalently formulated as

$$\widehat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^m} \mathcal{L}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T), \tag{11}$$

where

$$\mathcal{L}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T) = \left[\frac{1}{T}\sum_{t=1}^{T} \ell(\boldsymbol{\theta}; \mathbf{y}_t, \widehat{\mathbf{c}}_T) + \lambda_T \sum_{i=1}^{m} w_i |\theta_i|\right], \tag{12}$$

where λ_T is the LASSO tuning parameter and w_i are the adaptive LASSO weights. The specification of the weight is $w_i = C_\bullet/|\widetilde{\theta}_{T\,i}|$ where $\widetilde{\theta}_{T\,i}$ denotes the pre-estimator of the θ_i coefficient and $C_\bullet$ denotes a positive constant that is equal to C_α for the $\boldsymbol{\alpha}$ coefficients and C_ρ for the $\boldsymbol{\rho}$ coefficients. Put differently, in the theoretical analysis of the estimator we assume that $\lambda_T^{\mathbf{G}} = \lambda_T\, C_\alpha$ and $\lambda_T^{\mathbf{C}} = \lambda_T\, C_\rho$ so that $\lambda_T^{\mathbf{G}}/\lambda_T^{\mathbf{C}} = O(1)$. Thus, λ_T controls the overall degree of shrinkage of the parameters of the model. We make the following assumptions.

ASSUMPTION 1. *The n-dimensional random vector process $\mathbf{y}_t$ follows a* VAR*(p)* $\mathbf{y}_t = \sum_{k=1}^{p} \mathbf{A}_{0k}\mathbf{y}_{t-k} + \boldsymbol{\epsilon}_t$, *where* $\boldsymbol{\epsilon}_t \sim i.i.d.(\mathbf{0}, \mathbf{C}_0^{-1})$. *Moreover,*

(a) $\det(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_{0k} z^k) \neq 0$ *for any* $|z| \leq 1$*;*

(b) there exists a constant $c > 0$ *such that* $\mathsf{E}[|\epsilon_{it}|^k] \leq k! c^{k-2} \mathsf{E}[\epsilon_{it}^2] < \infty$*, for any* $i = 1, \ldots, n$*,* $t = 1, \ldots, T$*,* $k = 3, 4, \ldots$*;*

(c) there exist couples of constants $\underline{M}_0, \overline{M}_0$ *such that* $0 < \underline{M}_0 \leq \mu_{\min}(\boldsymbol{\Sigma}(\omega)) \leq \mu_{\max}(\boldsymbol{\Sigma}(\omega)) \leq \overline{M}_0 < \infty$*, where* $\boldsymbol{\Sigma}(\omega)$ *is the spectral density matrix of* $\mathbf{y}_t$*, defined for* $\omega \in [-\pi, \pi]$*, and*

$\underline{M}_1, \overline{M}_1$ *such that* $0 < \underline{M}_1 \leq \mu_{\min}(\mathbf{C}_0) \leq \mu_{\max}(\mathbf{C}_0) \leq \overline{M}_1 < \infty$.

In part (a) we assume $\mathbf{y}_t$ to be generated by a stable VAR. In part (b) we assume a standard Cramér condition for the innovations. It is important to highlight that this implies the existence of all moments of the innovation distribution. This, in turns, implies that the innovation distribution is assumed to be light tailed, as in the case of Gaussian and sub-Gaussian distributions. Then it is straightforward to see that, because of independence of the innovations, part (a) and (b) together imply that $\mathbf{y}_t$ is a strong mixing process with moments also satisfying the Cramér condition (see e.g. the comments in Fan and Yao, 2005, Section 2.6). For such processes suitable Bernstein-type exponential inequalities for dependent processes apply (see e.g. Bosq, 1996; Doukhan and Neumann, 2007; Merlevède et al., 2011). Finally, in part (c) we assume positive definiteness of the spectral density matrix of $\mathbf{y}_t$ and of the precision matrix, $\mathbf{C}_0$, of the VAR innovations $\boldsymbol{\epsilon}_t$. This guarantees that the population Granger and contemporaneous network are both well defined. Notice that by assuming finite eigenvalues of spectral density we implicitly rule out the presence of a strong factor structure as the one assumed for example in Forni et al. (2000). However this is not ruling out neither cross-sectional nor serial dependence and weakly influential factors are still possible (Onatski, 2012; Ando and Bai, 2017). That said, it is important to emphasize that when a high degree of collinearity is present in the data then LASSO estimation can be unstable (De Mol et al., 2008). In these settings ridge estimation may be a better suited estimation strategy for large systems.

Additional notation has to be introduced before we can state the main result. The sets of non-zero parameters are denoted as $\mathcal{A}_G = \{(i,j,k) : \alpha_{0\,ijk} \neq 0\}$, $\mathcal{A}_C = \{(i,j) : \rho_0^{ij} \neq 0\}$ and $\mathcal{A} = \mathcal{A}_G \cup \mathcal{A}_C$. The number of non-zero parameters in the model is $q_T = |\mathcal{A}|$. The set of zero parameters is then $\mathcal{A}^c$.

ASSUMPTION 2. *Consider the pre-estimators* $\widehat{\mathbf{c}}_T$, $\widetilde{\boldsymbol{\theta}}_T = (\widetilde{\boldsymbol{\alpha}}_T, \widetilde{\boldsymbol{\rho}}_T)$. *Then, for* T *sufficiently large there exist constants* $C_1, C_2, C_3 > 0$ *such that, for any* $\eta > 0$, *with probability at least* $1 - O(T^{-\eta})$, *we have (a)* $\max_{1 \leq i \leq n} |\widehat{c}_{T\,ii} - c_{0\,ii}| \leq C_1 \sqrt{\frac{\log T}{T}}$; *(b)* $\max_{i \in \mathcal{A}^c} |\widetilde{\theta}_{T\,i}| \leq C_2 \sqrt{\frac{\log T}{T}}$; *(c)*

$\max_{i\in\mathcal{A}} |\widetilde{\theta}_{T\,i} - \theta_{0\,i}| \leq C_3\sqrt{q_T \frac{\log T}{T}}$.

Condition (a) is also required by Peng et al. (2009). Condition (b) is equivalent to the assumption made by Medeiros and Mendes (2016), while condition (c) is stronger (see also Huang et al., 2008). We discuss possible pre-estimators satisfying these conditions after stating our main result.

The following Proposition establishes the estimation and selection consistency of our proposed estimator.

PROPOSITION 1. (SELECTION CONSISTENCY AND ORACLE PROPERTY). *Suppose that, as* $T \to \infty$, $q_T = o\left(\sqrt{\frac{T}{\log T}}\right)$, $\lambda_T\sqrt{\frac{T}{\log T}} \to \infty$, $\sqrt{q_T}\lambda_T = o(1)$, $\sqrt{q_T}\sqrt{\frac{\log T}{T}} = o\left(\lambda_T\right)$, $n = O(T^{\zeta})$ *for some* $\zeta > 0$. *Let also* $\{s_T\}$ *be a positive signal sequence of real numbers such that for any* $i \in \mathcal{A}$ *we have* $|\theta_{0\,i}| \geq s_T$ *and such that* $\frac{s_T}{\sqrt{q_T}\lambda_T} \to \infty$. *Then, under Assumptions 1 and 2, for any* $\eta > 0$, $\widehat{\boldsymbol{\theta}}_T$ *exists with probability at least* $1 - O(T^{-\eta})$. *Moreover,*
(a) $\Pr\big(\widehat{\theta}_{T\,i} = 0\big) \geq 1 - O(T^{-\eta})$, *for any* $i \in \mathcal{A}^c$;
(b) there exists a constant $\kappa > 0$ *such that*

$$\Pr\left(\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right\| \leq \kappa\sqrt{q_T}\lambda_T\right) \geq 1 - O(T^{-\eta}),$$

and $\Pr\big(\text{sign}(\widehat{\theta}_{T\,i}) = \text{sign}(\theta_{0\,i})\big) \geq 1 - O(T^{-\eta})$, *for any* $i \in \mathcal{A}$.

Notice that, since we assume $q_T = o\left(\sqrt{\frac{T}{\log T}}\right)$, then condition (c) in Assumption 2 requires a vanishing bias for the pre-estimator. When the number of parameters to be estimated is small relative to T, then the least squares estimator of $\mathbf{A}_{0k}$ and the sample covariance estimator of $\mathbf{C}_0$ can be used to build consistent pre-estimators of $\boldsymbol{\theta}_0$, and Assumption 2 is satisfied. When the number of parameters is large relative to T, LASSO and ridge are natural pre-estimators of $\mathbf{A}_{0k}$. These do not automatically satisfy condition (c), as they may have a non-vanishing bias. In this case, additional conditions on the degree of penalization of the pre-estimators and on λ_T and q_T are needed. For LASSO those conditions would be similar to the ones in the work by Medeiros and Mendes (2016) to which we refer for details. For

ridge, if we denote as $\lambda_T^{\mathsf{ridge}}$ the parameter controlling the degree of penalization, it can be seen that the bias is $O(\sqrt{q_T}\lambda_T^{\mathsf{ridge}})$ (Shao and Deng, 2012; Bühlmann, 2013) and in this case, inspection of our proofs shows that, as long as $q_T\lambda_T^{\mathsf{ridge}}\lambda_T = o(1)$, consistent estimators can still be obtained. Similar arguments apply for regularized pre-estimators of $\mathbf{C}_0$.

Proposition 1 implies the consistency of the networks estimators.

COROLLARY 1. *Define as* $\widehat{\alpha}_{T\,ijk}$ *the generic entry of* $\widehat{\boldsymbol{\alpha}}_T$ *and as* $\widehat{\rho}_T^{ij}$ *the generic entry of* $\widehat{\boldsymbol{\rho}}_T$*, and define the estimated edges' sets of the Granger and contemporaneous networks as* $\widehat{\mathcal{E}}_{G\,T} = \{(i,j) \in \mathcal{V} \times \mathcal{V} : \widehat{\alpha}_{T\,ijk} \neq 0,$ *for at least one* $k \in \{1,...,p\}\}$*, and* $\widehat{\mathcal{E}}_{C\,T} = \big\{(i,j) \in \mathcal{V} \times \mathcal{V} : \widehat{\rho}_T^{ij} \neq 0\big\}$*. Then, under the same Assumptions of Proposition 1, for* T *sufficiently large and any* $\eta > 0$*, we have* $\Pr(\widehat{\mathcal{E}}_{G\,T} = \mathcal{E}_G) \geq 1 - O(T^{-\eta})$*, and* $\Pr(\widehat{\mathcal{E}}_{C\,T} = \mathcal{E}_C) \geq 1 - O(T^{-\eta})$*.*

Given the asymptotic conditions on the number of non-zero coefficients, q_T, the worst-case scenario is when it is almost in the order of $\sqrt{\frac{T}{\log T}}$. In that case λ_T needs to be nearly in the order of $T^{-1/4}$ to achieve consistency. On the other hand, for the best-case scenario, that is when $q_T = O(1)$ (for example, when the dimension n is fixed), then the order of λ_T can be nearly as small as $T^{-1/2}$ (within a factor of $\log T$). Consequently, the L_2-norm distance of the estimator from the true parameter is in the order of $\sqrt{\frac{\log T}{T}}$, with probability at least $1 - O(T^{-\eta})$.

3 Simulation Study

In this section we analyse the properties of our estimator using simulated data. The exercise consists in simulating a large sparse VAR process and then using the `nets` algorithm to estimate it.

FIGURE 1 ABOUT HERE

We simulate an $n = 100$ dimensional sparse VAR(1). Notice that the total number of parameters in this system is 15050. The sparse autoregressive matrix $\mathbf{A}_1$ and concentration

matrix $\mathbf{C}$ are obtained from an Erdös-Renyi random graph model. The Erdös-Renyi model is a graph $(\mathcal{V}, \mathcal{E})$ defined over a fixed set of vertices $\mathcal{V}$ and a random set edges $\mathcal{E}$. The existence of an edge between vertices i and j is determined by a Bernoulli trial with probability p that is independent of all other edges. The Erdös-Renyi random graph is said to be directed or undirected depending on whether the underlying edge set $\mathcal{E}$ is directed or not. In our simulation setting we begin by simulating a directed and an undirected Erdös-Renyi random graphs over a set of n vertices $\mathcal{V}$ denoted respectively by $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$. The probability of a link p is set to $1/n$ in both graphs. The $\mathbf{A}_1$ matrix is then constructed on the basis of a directed Erdös-Renyi model $\mathcal{G}_1$ as

$$[\mathbf{A}_1]_{ij} = \begin{cases} 0.275 & \text{if } (i,j) \in \mathcal{E}_1 \\ 0 & \text{otherwise} \end{cases},$$

where $\mathcal{E}_1$ is the set of (directed) edges of $\mathcal{G}_1$. The concentration matrix $\mathbf{C}$ is constructed on the basis of an undirected Erdös-Renyi model $\mathcal{G}_2$

$$[\mathbf{C}]_{ij} = \begin{cases} -\frac{1}{\sqrt{d_i d_j}} & i \neq j \text{ and } (i,j) \in \mathcal{E}_2 \\ 1.5 & i = j \\ 0 & \text{otherwise} \end{cases},$$

where $\mathcal{E}_2$ is the set of (undirected) edges of $\mathcal{G}_2$ and d_i denotes the degree of vertex i in this graph. Note that the simulation is designed in a way such that the sparsity structure of the Granger and contemporaneous networks of the VAR coincide with the one of the two random graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. Also, the specification guarantees that the VAR is stable and that the concentration matrix is positive definite. We report in Figure 1 the plot of the Granger and contemporaneous networks associated with a randomly chosen realization of the model. Notice that despite the networks being sparse (in the sense that the expected number of links is $O(n)$), they are almost fully interconnected. We simulate samples of different sizes

from the sparse VAR(1) we just described (T = 250, 500, 750 and 1000) and then use the `nets` algorithm to estimate the model. For simplicity, the tuning parameters $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ are set equal to a common shrinkage tuning parameter λ_T. Our LASSO estimator requires pre-estimators of the $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ parameters to construct the LASSO penalty weights. The pre-estimator of $\boldsymbol{\alpha}$ is the least squares estimator of the VAR(1) autoregressive matrix, while the pre-estimator of $\boldsymbol{\rho}$ is the partial correlation estimator obtained from the sample covariance of the VAR(1) residuals. Last, we initialize $\mathbf{c}$ using the reciprocal of the sample variances of each series. The model is then estimated over a range of values of the common shrinkage tuning parameter λ_T.

FIGURE 2 AND TABLE 1 ABOUT HERE

The simulation is replicated 1000 times and the quality of the `nets` estimator is measured on the basis of the MSE and the ROC curve, which is the plot of the false discovery rate (FDR) of the estimator versus the true positive rate (TPR). We report in the left panel of Figure 2 the MSE of the `nets` estimator as a function of the tuning parameter λ_T for the sample size T equal to 500, 750 and 1000.[3] The picture displays the typical profile of shrinkage type estimators, that is the MSE is a convex function of the tuning parameter, and as the sample size increases the MSE of the estimator decreases. The right panel of Figure 2 reports the ROC curve associated with the `nets` estimator for the sample size T equal to 250, 500, 750 and 1000. Recall that the FDR is defined as the ratio of incorrectly detected non-zero parameters over the total number of zero parameters, while TPR is defined as the ratio of correctly detected non-zero parameters over the total number of non-zero parameters. Note that the penalization coefficient determines the FDR and TFR properties of the estimator: when λ_T is small (large), the proportion of type 1 errors is high (low) while the proportion of type 2 errors is low (high). The curves show that as the sample size T increases the performance of the `nets` estimator, as measured by the area underneath the ROC curve, increases steadily. In Table 1 we report detailed results on the MSE and TPR of the `nets`

[3]We omit from the picture for $T = 250$ because of scaling issues.

estimator when the FDR is controlled at 1%, 5% and 10%. For comparison purposes, the table also reports the MSE of the pre-estimator. The MSE of the `nets` estimator decreases steadily as the sample size get larger. When the sample size is 250 the efficiency gains with respect to the pre-estimator are substantial. As the sample size increases the pre-estimator becomes progressively more efficient relative to the LASSO estimator, however the efficiency gain of `nets` are still large. As far as the TPR is concerned, the table shows that when the TPR is controlled at 1%, 5% and 10% levels, the procedure has a fair amount of power even when the sample size T is 250, and that as the sample size increases power raises steadily. In particular, the power is roughly around 80% when the sample size is 750 and the FDR is controlled at the 1% level. Overall, the simulation results convey that the `nets` algorithm performs satisfactorily, and that the gains with respect to the traditional estimator can be large for sparse VAR systems.

4 Empirical Application

We use the methodology introduced in this work to study interconnectedness in a panel of volatility measures. The application is close in spirit to, among others, the research of Diebold and Yılmaz (2009, 2014, 2015) and Engle et al. (2012). We consider a panel of ninety U.S. bluechips across different industry sectors. The list of company names and industry groups is in Table C-1 in the online appendix. Our sample spans January 2nd 2004 to December 31st 2015, which corresponds to 3021 trading days. During this sample period most of the stocks in the list have been part of the S&P 100 index. Following Diebold and Yılmaz (2015) we measure volatility using the high-low range (Parkinson, 1980)

$$\tilde{\sigma}_{it}^2 = 0.361 \left(p_{it}^{\mathsf{high}} - p_{it}^{\mathsf{low}} \right)^2 ,$$

where p_{it}^{high} and p_{it}^{low} denote respectively the max and the min log price of stock i on day t.[4]

[4]Several advanced estimators of volatility based on high frequency data have been proposed over the last years (Andersen et al., 2003). However, a number of contributions have pointed out that simple estimators,

We focus on analyzing volatility interconnectedness conditional on a market wide and sector specific volatility factors. There is a large literature documenting evidence of a factor structure in volatility (see, *inter alia*, Barigozzi et al., 2014; Barigozzi and Hallin, 2016). As previously pointed out, it is straightforward to check that when common factors are present the dependence structure of the data is not sparse. To this extent, we study the interconnectedness of the residuals of the regression

$$\log \tilde{\sigma}^2_{i\,t} = \beta_0 + \beta_1 \log \tilde{\sigma}^2_{m\,t} + \beta_2 \log \tilde{\sigma}^2_{s\,t} + z_{i\,t}, \tag{13}$$

where $\tilde{\sigma}^2_{m\,t}$ and $\tilde{\sigma}^2_{s\,t}$ denote respectively a market wide and a sector specific volatility factors. Thus with the notation used in the previous sections we have $z_{i\,t} = y_{i\,t}$. The market and sectoral volatilities are measured using the high-low range estimator applied to the S&P 500 index and the SPDR sectoral indices of S&P 500.[5] The residuals are obtained after estimating the model by least squares. In what follows we refer to the volatility residual panel as the volatility panel for short.

TABLE 2 AND FIGURE 3 ABOUT HERE

Table 2 reports summary statistics on the variance, kurtosis, autocorrelation, average cross correlation and average cross autocorrelation of order one for the volatility residuals. Moreover, in Figure 3 we show the heatmaps of the sample autocorrelation matrix of order one and the sample correlation matrix. We note that after netting out the factors, the volatility residuals still exhibit autocorrelation. It is important to emphasize that while the raw volatility measure exhibit long range dependence, the volatility residuals exhibit a considerably weaker autocorrelation structure. In a way, the volatility residuals can be thought of the short run idiosyncratic volatility component of a volatility component model (Barigozzi et al., 2014; Wang and Ghysels, 2015). Inspection of the average correlations and

like the high-low range, perform satisfactorily (see also Alizadeh et al., 2002; Brownlees and Gallo, 2010).

[5]The SPDR sectoral indices of the S&P 500 we use are Consumer Discretionary (XLY), Consumer Staples (XLP), Energy (XLE), Financials (XLF), Health Care (XLV), Industrials (XLI), Materials (XLB), Technology (XLK) and Utilities (XLU).

the heatmaps shows that contemporaneous and lagged cross correlation is still present in the volatility residuals. Interestingly, tickers in the same industry still exhibit a moderate degree of correlation even after conditioning on the sectoral factors.

4.1 In-Sample Estimation Results

We analyse the panel of volatility measures using the `nets` algorithm over the entire sample. The order of the VAR model p is set to one. The pre-estimator of the $\boldsymbol{\alpha}$ parameters is the least squares estimator of the VAR(1) autoregressive matrix, while the pre-estimator of the $\boldsymbol{\rho}$ parameters is the partial correlation estimator obtained from the sample covariance of the VAR(1) residuals. Last, we initialize $\mathbf{c}$ using the reciprocal of the sample variances of each series. The penalties $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ are determined by a cross-validation procedure. We split the entire sample in an estimation and a validation samples. The estimation sample corresponds to the first 75% of the entire sample and the validation sample to the last 25%. For given values of the tuning parameters, we first estimate the model in the estimation sample and then compute the residual sum of squares (RSS) in the validation sample. We perform these steps over a grid of $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ values and then choose the optimal tuning parameters as the ones that minimize the validation RSS. We then estimate the model over the entire sample using the optimal value of the tuning parameters.

TABLE 3 AND FIGURE 4 ABOUT HERE

We report the estimated Granger and contemporaneous networks in Figure 4. In the Granger network plot the diameter of each vertex is proportional to the out-degree (the number of non-zero spillovers effects toward others) while in the contemporaneous network the diameter is proportional to the degree. In both plots we use the vertex color to denote the different industry groups. We exclude from the graphs the vertices that do not have any connections, which is one ticker in the Granger network and seven tickers in the contemporaneous network.

Table 3 reports the sum of the degrees of the vertices in the Granger and contemporaneous networks over the entire panel and individual sectors. The estimated Granger volatility network has approximately 3% of the total edges, while the contemporaneous volatility network approximately 4% of the total edges. The estimated networks share some common features. For instance, the number of industry linkages of the two networks are highly correlated and the financial sector is in particular the sector that accounts for most linkages.

We compute an in-sample R^2 type goodness-of-fit criteria for each series in the panel to summarise the amount of variability explained by the sparse VAR, which is defined as the proportion of variance explained by the regression equation (8). Table 3 reports the average of the R^2 index over the entire panel as well as the individual sectors. The index has a strong positive correlation with the number of linkages in each sector and is on average around 22%. For comparison purposes, Table 3 also reports in-sample factor and sectoral R^2. The factor R^2 is defined as the R^2 obtained by regressing the volatility measure on the market factor and the sector R^2 is defined as the R^2 obtained by regressing the volatility measure on the market wide and sector factor minus the factor R^2. The market and sector factors account for most of the variability in the series, which is roughly 56%. A back of the envelope computation shows that the networks explains around an additional 11% of the overall variability, which roughly matches the amount of variability explained by sectoral factors.

TABLE 4 AND FIGURE 5 ABOUT HERE

In order to get better insights on the industry linkages in Table 4 we report the total number of links between industry groups. It is interesting to note that after conditioning on the sectoral factors there are still a moderate number of interconnections between firms within the same industry. The table also shows that firms in the financial sector in particular have a high degree of interconnectedness across industries. In Figure 5 we report the degree distribution of the estimated networks and the distribution of the non-zero $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ coefficients. As far as the degree distribution is concerned, the number of connections has a high degree of heterogeneity in the cross section. In particular, in the contemporaneous network

the most interconnected tickers account for a large number of connections relative to the total. The histogram of the non-zero coefficients shows that the majority of the coefficients are positive and that positive coefficients are on average larger than the negative ones.

TABLE 5 ABOUT HERE

We rank the firms in the panel on the basis of their influence in the Granger and contemporaneous networks. We measure the influence of series j in the Granger and contemporaneous networks using, respectively, the indices $\sum_{i\neq j}^{N} |\widehat{\alpha}_{ij1}|$ and $\sum_{i\neq j}^{N} |\widehat{\rho}_{ij}|$. We report the top ten most influential tickers of the Granger and contemporaneous networks according to this criteria in Table 5. The table shows clearly that large financials firms are highly influential. In particular, the results shows that the financial firms that have been heavily involved in the great financial crisis like Bank of America (BAC), AIG and Citigroup (C) are the stocks associated with the largest spillover effects in the Granger network.

Overall, the in-sample estimation results show that, after conditioning on market wide and sectoral factors, the sparse VAR captures an important proportion of overall variability, and that the financial industry in particular has the highest degree of interconnectedness.

FIGURE 6 ABOUT HERE

Last, we investigate the stability of the sparsity patterns of the estimated network through time. In Figure 6 we report the heatmap of the adjacency matrices of the Granger and contemporaneous networks estimated from the beginning of the sample until the end of December 2015 (the end of the in-sample period) and until the end of December 2013. We find that the sparsity patterns of the network exhibit a moderate degree of time-variation. In particular, we have that for the Granger (contemporaneous) network 80.6% (85.3%) of the edges present over the entire sample are also present at the end December 2013, whereas 98.7% (98.1%) of the edges not present over the entire sample are also not present at the end December 2013.

4.2 Out-of-Sample Forecasting

We carry out a forecasting exercise to evaluate the out-of-sample performance of the methodology. The exercise is designed as follows. We split the sample in an in-sample period spanning January 2nd 2004 to December 31st 2013 and an out-of-sample period spanning January 2nd 2014 to December 31st 2015. We first estimate the sparse VAR in-sample using the same steps outlined in the previous section and we then evaluate the model in the out-of-sample period.

TABLE 6 ABOUT HERE

The prediction evaluation is divided into two parts. The first part focuses on the evaluation of the `nets` estimator of the autoregressive component by predicting one-step-ahead volatility residuals. The benchmark forecast for this exercise is the constant zero forecast. Notice that the constant zero forecast represents the optimal forecast in case the dependence in the panel is fully captured by the factor part of model (13) without exploiting the information in the residuals. The competing forecasts are the ones obtained from a VAR(1) model estimated via `nets`, univariate AR(1) models estimated by least squares, a VAR(1) model estimated by ridge regression and VAR(1) estimated via Bayesian methods. The ridge VAR(1) is fitted by estimating the VAR equation by equation using standard ridge regression. The degree of shrinkage of each equation is chosen via Generalized Cross Validation. The Bayesian VAR(1) relies on traditional conjugate priors (Koop and Korobilis, 2009; Sims and Zha, 1998). Note that the volatility residuals are obtained from the estimation results of model (13) estimated over the entire sample.

We report the forecasting results in the top panel of Table 6. The first row of the table reports the MSE of the benchmark while the remaining rows report the out-of-sample R^2 of the competitors. The out-of-sample R^2 index is defined as one minus the ratio of the MSE of the competing models over the MSE of the benchmark. The performance indices are averaged over the entire panel and the industry sectors. The results show that the VAR forecasts obtained by the `nets` estimator systematically improve forecasting ability over the

benchmark by roughly 8% on average and it is the best performing forecast method overall. The ridge and Bayesian VAR do not perform particularly well in this setting but this may be due to the forecasting design. Conditional on the factors the data exhibit a substantially lower degree of collinearity and sparsity appears to be reasonable modelling assumption. In fact, NETS as well as the simple AR(1) end up performing quite well.

The second part focuses on the evaluation of the `nets` estimator of the contemporaneous component by predicting the contemporaneous volatility residuals conditional on the estimated autoregressive component. We construct the series of VAR residuals $\widehat{\epsilon}_{it}$ of the autoregressive component estimated by `nets`, and the focus is on predicting each residuals series conditional on the remaining ones on the basis of the regression

$$\widehat{\epsilon}_{it} = \sum_{\substack{h=1 \\ h \neq i}}^{n} \gamma_{ij} \widehat{\epsilon}_{ht} + u_{it}, \quad i = 1, \ldots, n.$$

The benchmark forecast for this exercise is again the constant zero forecast, which is the optimal forecast in case the residuals do not have any cross-correlation. The competing forecasts are ones the ones obtained from the contemporaneous component of the VAR estimated by `nets`, the ones obtained from a linear regression estimated by least squares and a linear regression estimated by ridge regression (with tuning parameter chosen by Generalized Cross Validation). The linear regression and the ridge regression are estimated in the in-sample period using the in-sample one-step ahead forecast errors.

We report the forecasting results in the bottom panel of Table 6. The first row of the table reports the average MSE of the benchmark model while the remaining rows report the out-of-sample R^2 of the competitors. Results show that the `nets` forecasts systematically improve out-of-sample predictive ability across sectors and on average improve forecasting over the benchmark by 13%.

5 Conclusions

In this work we introduce network techniques for the analysis of lagre panels of time series. We model a panel as a VAR where the autoregressive matrices and the inverse covariance matrix of the system innovations are assumed to be sparse. The system has a natural network representation in terms of a directed graph representing predictive Granger relations and an undirected graph representing contemporaneous partial correlations. A LASSO estimation algorithm called `nets` is introduced to estimate simultaneously the autoregressive matrices and the inverse covariance matrix of the model. The large sample properties of the estimator are established in a high-dimensional setting. The methodology is used to analyse a panel of volatility measures of US bluechips between January 2004 and December 2015 conditional on market wide and sector specific volatility factors. The analysis shows that the series exhibit a hight degree of interconnectedness and that financial firms have the highest degree of interdependence. A forecasting exercise shows that the methodology introduced in this work allows to improve forecasting over a number of benchmarks.

References

Acemoglu, D., Carvalho, V., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80:1977–2016.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57:1047–1091.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Ando, T. and Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*. available online.

Anufriev, M. and Panchenko, V. (2015). Connecting the dots: Econometric methods for uncovering networks with an application to the australian financial institutions. *Journal of Banking & Finance*, 61:S241–S255.

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

Barigozzi, M., Brownlees, C., Gallo, G. M., and Veredas, D. (2014). Disentangling systematic and idiosyncratic dynamics in panels of volatility measures. *Journal of Econometrics*, 182:364–384.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, 19:C33–C60.

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

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

Billio, M., Getmansky, M., Lo, A., and Pellizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104:535–559.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Springer, New York.

Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, 8:29–56.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19:1212–1242.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, 51:157–172.

Davis, R. A., Zang, P., and Zheng, T. (2015). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*. available online.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics*, 146:318–328.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28:157–175.

Den Haan, W. J. and Levin, A. (1996). Inferences from parametrics and non–parametric covariance matrix estimation procedures. NBER Technical Working Paper 195.

Diebold, F. X. and Yılmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119:158–171.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182:119–134.

Diebold, F. X. and Yılmaz, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.

Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117:878–903.

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137:334–353.

Engle, R. F., Gallo, G. M., and Velucchi, M. (2012). Volatility spillovers in East Asian financial markets: A MEM-based approach. *The Review of Economics and Statistics*, 94:222–223.

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series*. Springer, New York.

Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1).

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.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432–441.

Fu, W. J. (1998). Penalized regression: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7:397–416.

Giannone, D., Lenza, M., and Primicieri, G. E. (2018). Economic predictions with big data: The illusion of sparsity. Technical report, Federal Reserve Bank of New York Staff Reports 847.

Hagströmer, B. and Menkveld, A. J. (2016). A network map of information percolation. Technical report, Stockholm Business School & VU University Amsterdam.

Härdle, W. K., Wang, W., and Yu, L. (2016). Tenet: Tail-event driven network risk. *Journal of Econometrics*, forthcoming.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Financial network systemic risk contributions. *Review of Finance*. available online.

Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18:1603–1618.

Kock, A. B. (2016). Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions. *Econometric Theory*, 32:243–259.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186:325–344.

Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3:267–358.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large–dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.

Mantegna, R. (1999). Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 11:193–197.

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 Econometrics*, 191:255–271.

Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34:1436–1462.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151:435–474.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53:61–65.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104:735–746.

Shao, J. and Deng, X. (2012). Estimation in high-dimensional linear models with deterministic design matrices. *The Annals of Statistics*, 40:812–831.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39:949–968.

Song, S. and Bickel, P. (2011). Large vector auto regressions. Technical report, SFB 649 discussion paper 2011-048.

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.

Wang, F. and Ghysels, E. (2015). Econometric analysis of volatility component models. *Econometric Theory*, 31(2):362393.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429.

Table 1: Simulation study

	FDR=1%		FDR=5%		FDR=10%		MSE
T	TPR	MSE	TPR	MSE	TPR	MSE	pre-estimator
250	0.49	0.32	0.55	0.31	0.60	0.34	5.33
500	0.75	0.10	0.84	0.17	0.88	0.25	1.45
750	0.79	0.06	0.89	0.11	0.92	0.15	0.81
1000	0.82	0.05	0.93	0.09	0.96	0.15	0.55

The table reports the results of the simulation exercise for different values of the sample size T. The table reports the True Positive Rate (TPR) and the MSE of the `nets` estimator when the False Discovery Rate (FDR) is controlled at the 1%, 5% and 10% levels. The last column of the table shows the MSE of the pre-estimator.

Table 2: Descriptive Stats

	Disc	Stap	Energy	Fin	Heal	Ind	Tech	Mat	Util	All
variance	0.064	0.049	0.042	0.064	0.059	0.050	0.059	0.064	0.041	0.056
kurtosis	4.401	5.434	4.981	4.929	4.993	4.676	4.467	4.413	5.740	4.807
ρ_1	0.260	0.231	0.206	0.309	0.254	0.222	0.257	0.320	0.193	0.253
ρ_5	0.170	0.137	0.144	0.241	0.156	0.133	0.163	0.237	0.120	0.168
ρ_{22}	0.142	0.104	0.123	0.183	0.118	0.109	0.120	0.202	0.075	0.132
$\rho_{0,\text{others}}$	0.091	0.089	0.063	0.077	0.087	0.098	0.080	0.073	0.069	0.083
$\rho_{1,\text{others}}$	0.045	0.048	0.024	0.028	0.046	0.045	0.042	0.034	0.034	0.039

The table reports average descriptive statistics over the industry sectors and the entire panel. The set of descriptive statistics considered contains the sample variance, kurtosis, autocorrelation of order 1, 5 and 22, the average contemporaneous correlation with all other tickers, and the average order 1 autocorrelation with all other tickers.

Table 3: Network Estimation Summary

	Disc	Stap	Ener	Fin	Heal	Ind	Tech	Mat	Util	All
Granger Links	47	15	19	41	34	39	35	16	5	251
Contemporaneous Links	33	20	29	71	30	46	51	11	3	294
`nets` R^2_{is}	22.5	18.6	18.9	32.2	18.9	24.4	20.0	19.5	11.0	22.3
factor R^2_{is}	45.3	37.2	42.7	56.6	33.6	51.3	39.1	45.9	36.8	44.4
sector R^2_{is}	7.3	9.6	25.3	16.1	9.0	5.8	8.4	10.6	25.8	11.6

The table reports summary estimation results over the industry sectors and the entire panel. The first row of the table reports the outer degree of the Granger network, the second row reports the degree of the contemporaneous network, the third row reports the (in–sample) average R^2_{is} of the `nets` regression. For comparison, the table reports in the fourth and fifth rows the average (in–sample) factor R^2_{is} and (in–sample) sector R^2_{is}, respectively.

Table 4: Sectoral Linkages

	Granger Component								
	Disc	Stap	Ener	Fin	Heal	Ind	Tech	Mat	Util
Disc	17.2	31.6	4.8	17.1	27.3	17.5	21.4	13.3	40.0
Stap	13.8	0.0	0.0	2.4	9.1	5.0	3.6	6.7	0.0
Ener	6.9	10.5	19.0	9.8	3.0	7.5	7.1	6.7	0.0
Fin	17.2	0.0	19.0	14.6	9.1	25.0	14.3	40.0	40.0
Heal	17.2	15.8	14.3	14.6	15.2	15.0	28.6	6.7	20.0
Ind	6.9	21.1	9.5	19.5	9.1	12.5	14.3	6.7	0.0
Tech	13.8	15.8	19.0	12.2	15.2	15.0	3.6	20.0	0.0
Mat	6.9	5.3	9.5	9.8	3.0	2.5	7.1	0.0	0.0
Util	0.0	0.0	4.8	0.0	9.1	0.0	0.0	0.0	0.0
	Contemporaneous Component								
	Disc	Stap	Ener	Fin	Heal	Ind	Tech	Mat	Util
Disc	5.3	12.5	10.3	9.9	9.1	23.5	9.3	16.7	0.0
Stap	7.9	8.3	6.9	6.2	6.1	11.8	7.4	0.0	0.0
Ener	7.9	8.3	10.3	12.3	12.1	0.0	7.4	16.7	33.3
Fin	21.1	20.8	34.5	7.4	27.3	33.3	37.0	33.3	66.7
Heal	7.9	8.3	13.8	11.1	6.1	9.8	11.1	16.7	0.0
Ind	31.6	25.0	0.0	21.0	15.2	3.9	16.7	0.0	0.0
Tech	13.2	16.7	13.8	24.7	18.2	17.6	11.1	0.0	0.0
Mat	5.3	0.0	6.9	4.9	6.1	0.0	0.0	16.7	0.0
Util	0.0	0.0	3.4	2.5	0.0	0.0	0.0	0.0	0.0

The table reports the fraction of Granger and contemporaneous linkages between the industrial sectors. The (i, j) entry of the Granger linkages table is defined as the total number of linkages from sector j to sector i standardized by the total number of linkages from sector j. The (i, j) entry of the contemporaneous linkages table is defined as the total number of linkages between sector i and j standardized by the total number of linkages of sector j.

Table 5: Rankings

Rank	Granger		Contemporaneous	
	Company	Sector	Company	Sector
1	BAC	Financials	UNP	Industrials
2	AIG	Financials	T	Technology
3	C	Financials	USB	Financials
4	ALL	Financials	GE	Industrials
5	MCD	Discretionary	TGT	Discretionary
6	HPQ	Technology	WFC	Financials
7	DOW	Material	MS	Financials
8	SPG	Financials	NSC	Industrials
9	GE	Industrials	F	Discretionary
10	CVS	Staples	NOV	Energy

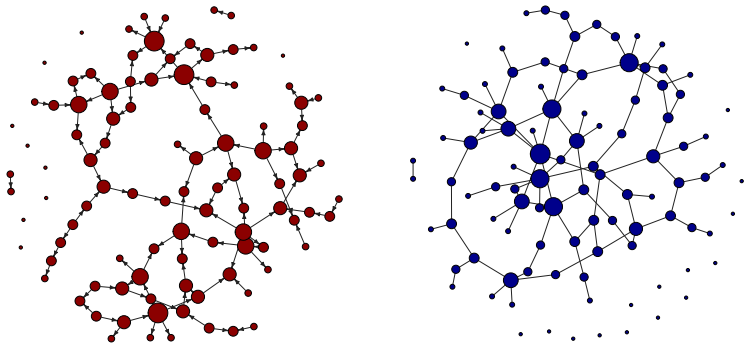
The table reports the top ten of the most interconnected series in the Granger and contemporaneous networks.

Table 6: Forecasting

	Granger Component									
	Disc	Stap	Energy	Fin	Heal	Ind	Tech	Mat	Util	All
benchmark MSE	5.67	3.96	4.52	3.90	5.15	4.32	5.86	8.16	3.14	4.91
`nets` R^2_{oos}	10.21	6.25	5.30	5.86	6.98	5.46	9.49	17.90	0.50	8.08
AR(1) R^2_{oos}	5.65	3.58	5.04	1.65	2.95	4.07	6.30	15.14	-3.49	5.06
ridge R^2_{oos}	5.47	1.79	2.74	-8.46	1.09	3.39	4.89	12.84	-8.20	2.57
BVAR R^2_{oos}	5.89	2.19	3.53	-8.20	1.36	3.77	3.47	8.70	-8.78	1.63
	Contemporaneous Component									
	Disc	Stap	Energy	Fin	Heal	Ind	Tech	Mat	Util	All
benchmark MSE	5.12	3.74	4.30	3.66	4.82	4.11	5.35	6.78	3.11	4.54
`nets` R^2_{oos}	15.76	12.56	12.98	17.21	6.28	19.08	14.03	3.00	-1.17	13.20
reg R^2_{oos}	12.89	9.12	9.02	15.23	2.97	17.44	10.98	-1.94	-6.44	10.19
ridge R^2_{oos}	12.97	9.20	9.05	15.32	3.06	17.50	11.04	-1.87	-6.37	10.26

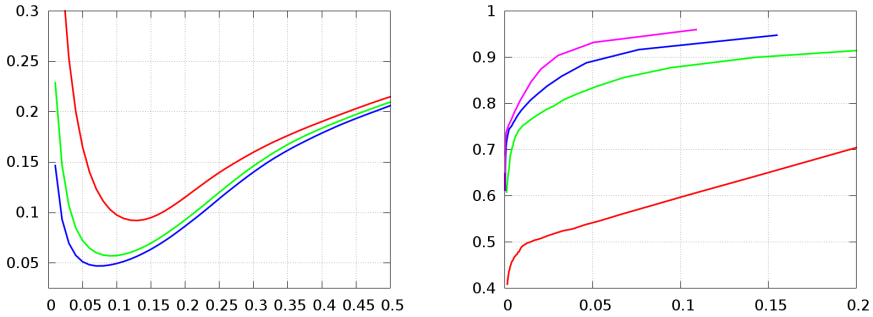
The table reports summary forecasting results over the industry sectors and the entire panel. The first panel reports forecasting exercise for the Granger component. The first row reports the MSE (times 100) of the benchmark. The second to fifth rows report the out-of-sample R^2_{oos} of, respectively, `nets`, an AR(1) estimated by least squares, a VAR estimated using ridge regression and a Bayesian VAR. The second panel reports the forecasting exercise of the contemporaneous component. The first row reports the MSE (times 100) of the benchmark. The second to fourth rows report the out-of-sample R^2_{oos} of, respectively, `nets`, the linear regression estimator and the ridge estimator.

Figure 1: Simulated Granger and Contemporaneous Networks



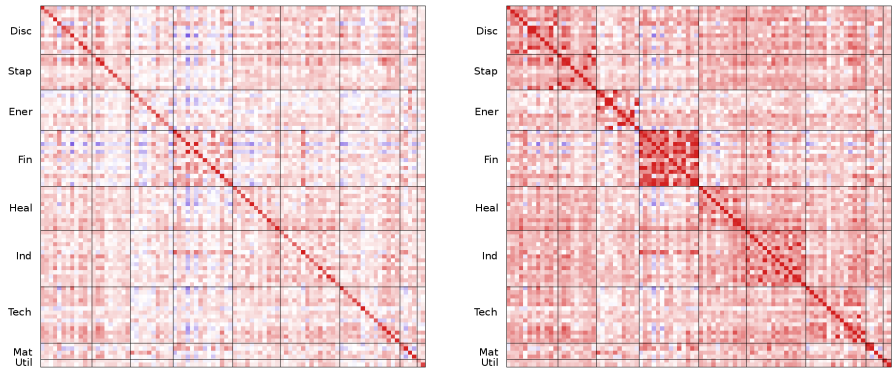
The figure displays realizations of the Erdös-Renyi random graph models used in the simulation exercise. The left picture displays a directed Erdös-Renyi graph used to generate the autoregressive matrix **A** while the right picture displays an undirected Erdös-Renyi random graph used to generate the contemporaneous concentration matrix **C**.

Figure 2: Simulation Study



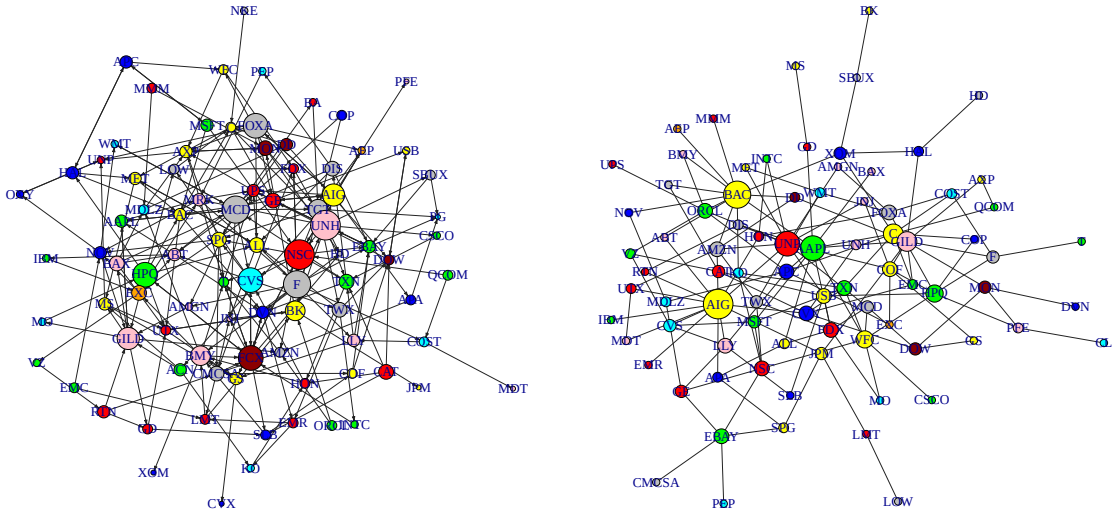
The left picture displays the MSE (multiplied by 100) of the `nets` estimator as a function of the tuning parameter λ_T for (from top to bottom) $T = 500, 750, 1000$. The right picture displays ROC curve of the `nets` estimator for (from bottom to top) $T = 250, 500, 750, 1000$.

Figure 3: Autocorrelation and Correlation Heatmaps



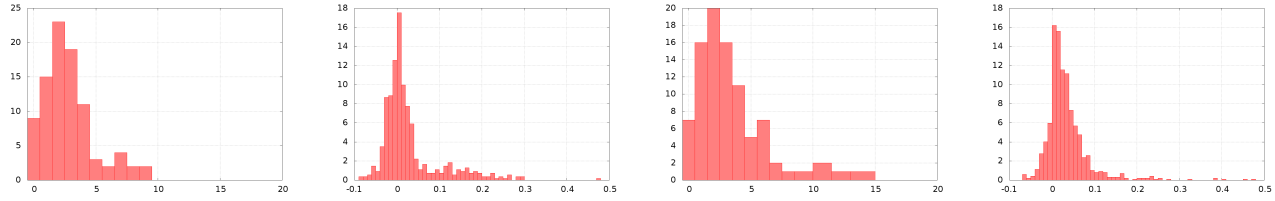
The figure displays the heatmap of the sample autocorrelation (left) and sample correlation matrices of the residuals of regression (13).

Figure 4: S&P 100 Volatility Granger and Contemporaneous Networks



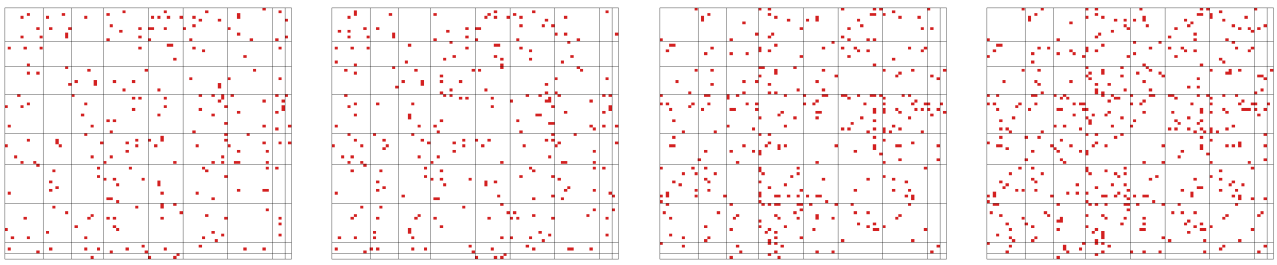
The figure displays the estimated Granger and contemporaneous networks. The size of the vertices is proportional to their degree and the colour of the vertices depends on their industry sector.

Figure 5: Degree and Coefficient Distributions



The first two pictures from the left display the histogram of the degree distribution of the Granger network and the histogram of the estimated non-zero $\boldsymbol{\alpha}$ coefficients. The last two pictures from the left display the degree distribution of the contemporaneous network and the histogram of the estimated non-zero $\boldsymbol{\rho}$ coefficients.

Figure 6: Stability of the Sparsity Patterns



The first two pictures from the left display the adjacency matrix of the Granger network estimated from January 2004 to December 2015 and from January 2004 to December 2013. The last two pictures from the left display the adjacency matrix of the contemporaneous network estimated from January 2004 to December 2015 and from January 2004 to December 2013.