NETS:
NETWORK ESTIMATION FOR TIME SERIES

Matteo Barigozzi[‡] Christian Brownlees[†]

This Draft: October, 2013

Abstract

This work proposes novel network analysis techniques for multivariate time series. We define the network of a multivariate time series as a graph where vertices denote the components of the process and edges denote non–zero long run partial correlations. We then introduce a two step LASSO procedure, called NETS, to estimate high–dimensional sparse Long Run Partial Correlation networks. This approach is based on a VAR approximation of the process and allows to decompose the long run linkages into the contribution of the dynamic and contemporaneous dependence relations of the system. The large sample properties of the estimator are analysed and we establish conditions for consistent selection and estimation of the non–zero long run partial correlations. The methodology is illustrated with an application to a panel of U.S. bluechips.

Keywords: Networks, Multivariate Time Series, Long Run Covariance, LASSO

JEL: C01, C32, C52

[‡] Department of Statistics, London School of Economics and Political Science,
e-mail: `m.barigozzi@lse.ac.uk`;
[†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE,
e-mail: `christian.brownlees@upf.edu`;

The proofs of the main results of the paper are contained in an appendix available online from the authors' websites. The procedures presented in this paper are available in the package `nets` for `R`.

1 Introduction

Over the last years, network analysis has become increasingly popular in economics and finance. In a nutshell, network analysis aims at representing the interconnections of a large multivariate system as a graph, and the graph representation is then used to study the properties of the system. In finance, these methods have been applied to study interconnections among financial institutions in order to identify channels of contagion. Applications in this area have gauged considerable interest in the aftermath of the 2007-2009 financial crisis. Example of contributions on network analysis in the literature include, *inter alia*, Billio, Getmansky, Lo, and Pellizzon (2012), Diebold and Yilmaz (2013) and Hautsch, Schaumburg, and Schienle (2012, 2013).

In this work we propose a novel network analysis technique to represent the cross–sectional conditional dependence structure of a high–dimensional multivariate time series. This entails two tasks. We first introduce a measure of cross–sectional conditional dependence for time series that is going to be used to define the network. We then develop an estimation approach that allows to estimate the dependence measure from the data.

The standard network dependence measure used in the statistics and graphical modelling literature is partial correlation (see e.g. Dempster, 1972; Meinshausen and Bühlmann, 2006). The partial correlation network is defined as an undirected weighted graph where vertices denote the components of the process and the presence of an edge between the i-th and j-th vertices indicates that the i-th and j-th components are partially correlated given all other components. The network edges are then associated with the value of the partial correlation between the vertices they connect to express the weight of the link. Partial correlation networks are typically motivated by the analysis of i.i.d. Gaussian data. In this setting, absence of partial correlation implies conditional independence, thus the partial correlation network exhaustively characterizes cross–sectional dependence. This definition however has limitations in the analysis of time series processes exhibiting serial dependence. Partial correlation only captures contemporaneous cross–sectional dependence, while in a time series setting cross–sectional dependence can arise at different

leads or lags. In fact, in the the time series network literature Billio *et al.* (2012) and Diebold and Yilmaz (2013) have proposed network definitions that take into account the dynamic properties of the data.

In this work we propose a network definition based on long run partial correlation, a dependence measure constructed from the long run covariance matrix of the process. The Long Run Partial Correlation network is defined analogously to the (simple) partial correlation network by using the long run partial correlation as a measure of dependence between two time series. Long run partial correlation is a comprehensive and model–free measure of cross–sectional conditional dependence. It captures contemporaneous correlation as well as lead/lag effects. Moreover, it is model–free in the sense that it does not hinge on specific modelling assumptions on the process. These are appealing features for economic and financial applications where dependence across series is not necessarily only contemporaneous and correct model specification is often a concern. Long run partial correlation can also be interpreted as a special case of partial spectral coherence, a cross–sectional conditional dependence measure based on the spectral density matrix used in frequency domain analysis (see Brillinger, 1981; Dahlhaus, 2000; Eichler, 2007).

In order to estimate the Long Run Partial Correlation network from the data, we propose an estimation approach that combines ideas from the high–dimensional sparse graph estimation literature in statistics (see Meinshausen and Bühlmann, 2006; Peng, Wang, Zhou, and Zhu, 2009) and the heteroskedastic and autocorrelation consistent (HAC) covariance estimation literature in econometrics (see White, 1984; Gallant, 1987; Newey and West, 1987; Andrews and Monahan, 1992; Den Haan and Levin, 1994). Following Meinshausen and Bühlmann (2006) and Peng *et al.* (2009), we are concerned with the case in which the network is sparse and the main inferential problem of interest consists of selecting and estimating the non–zero long run partial correlations of the system. Graphically this corresponds to detecting the network linkages and estimating their weight. It can be shown that long run partial correlations are a function of the inverse of the long run covariance matrix (i.e. the long run concentration matrix). Moreover, the long run (i, j)-th partial correlation is zero if and only if the (i, j)-th element of the long run concentration

matrix is zero. This implies that network estimation can be formulated as a sparse long run concentration matrix estimation problem. A sufficiently regular multivariate time series can be represented as a Vector Autoregression (VAR) of infinite order and the long run covariance matrix of such representation is equal to the long run covariance of the original process. Furthermore, the long run covariance is available in closed form and its inverse is a "sandwich" form made up of two components: (*i*) a matrix that depends on the VAR parameters that summarises the long run Granger causality structure of the process (which we name Granger component) and (*ii*) the concentration matrix of the VAR innovations that synthesises contemporaneous dependence conditionally on the past information (contemporaneous component). These two components can be further associated to two networks representing predictive and contemporaneous association among series. From a graphical perspective such representation is appealing in that the long run partial correlation linkages can be interpreted as a (nontrivial) combination of the Granger and contemporaneous links.

We propose a natural regression procedure based on the VAR representation of the data generating process to estimate the long run correlation network. The algorithm consists of estimating the VAR parameters and the concentration matrix of the VAR residuals using LASSO regressions. The outputs of the procedure are the Long Run Partial Correlation network as well as the Granger and Contemporaneous networks. We name this procedure NETS (Network Estimator for Time Series). The properties of this estimation strategy are established under the assumption that the VAR representation is sparse. The sparsity of the VAR representation determines in turn the sparsity structure of the Granger, Contemporaneous, and Long Run Partial Correlation networks. We establish conditions for consistent selection of the network edges and consistent estimation of edges' weights. An important highlight of the theory developed in this work is that results are derived in a high–dimensional settings that allows for the number of series and order of the VAR to increase with the sample size. This implies that the number of possible linkages is allowed to be larger than the number of observations available. The theory is developed for a zero–mean, purely–nondeterministic, weakly stationary process. Standard results from

the HAC literature (see Den Haan and Levin, 1994) lead us to conjecture that the result should also hold for more general classes of process that allow, among other features, for heteroskedasticity.

We illustrate this technique with an application to financial network analysis based on a panel of financial monthly equity returns for a set of 41 bluechips across different industry groups between 1990 and 2010. Building up on standard models used in empirical finance, we model the returns using a CAPM type one factor model: the returns of each firm are a linear function of the market return and an idiosyncratic shock. We then use NETS to analyse the network structure of the idiosyncratic shocks. We find that the market accounts on average for 25% of the variation of returns while links of the idiosyncratic network account on average for another 15%. Moreover, a number of network analysis indices are used to summarises the properties of the network and we find that the idiosyncratic risk network shares several of the empirical regularities found in social network analysis.

Our work relates to different strands of literature. This paper is inspired and motivated by the econometric literature on the analysis of financial networks that has developed in the aftermath of the 2007–2009 financial crisis. Contributions in this area include Billio *et al.* (2012), Diebold and Yilmaz (2013), Hautsch *et al.* (2012, 2013), Dungey, Luciani, and Veredas (2012). Our contribution builds up on the statistical literature on networks and graphical models (see Lauritzen, 1996). In particular, the relation between partial correlation and the inverse of the covariance matrix was first put forward in Dempster (1972). Recently research in the area was boosted by a number of contributions developing techniques for high–dimensional partial correlation network estimation for independent data using the LASSO, *inter alia* Meinshausen and Bühlmann (2006), Friedman, Hastie, and Tibshirani (2008), Peng *et al.* (2009). See also, for instance, Peterson, Vannucci, Karakas, Choi, Ma, and Maletić-Savatić (2013) for a contribution on graph modelling from a Bayesian perspective. Also, our work relates to the research of De Mol, Giannone, and Reichlin (2008), Song and Bickel (2011), Jarociński and Maćkowiak (2011), Davis, Zang, and Zheng (2012), Kock (2012), Kock and Callot (2012), and Medeiros

and Mendes (2012) on sparse estimation for large VAR models. The estimation approach developed in this paper is inspired by the HAC covariance estimation literature, which includes contributes by White (1984), Gallant (1987), Newey and West (1987), Andrews and Monahan (1992), Den Haan and Levin (1994). The estimation approach developed in this work is high–dimensional, in the sense that we allow for the number of parameters to be estimated to be larger than the number of observations available. Contributions and surveys in the high–dimensional literature include Belloni, Chernozhukov, and Hansen (2011) and Bühlmann and van de Geer (2011) among others. This research also relates to the literature on regularized covariance estimation by Bickel and Levina (2008), Lam and Fan (2009), Ledoit and Wolf (2004), Pourahmadi (2011), Fan, Liao, and Mincheva (2011, 2013), and Bai and Liao (2012). Finally, theoretical research on networks in economics and finance, also in relation to the financial crisis, was put forward by Allen and Gale (2000) and Xavier, Parigi, and Rochet (2000). More recent influential contributions include Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2013).

The paper is structured as follows. Section 2 introduces the definition of network for time series and Section 3 presents the estimation approach. Section 4 presents simulation results. Section 5 describes the empirical application to a panel of equity returns for a panel of U.S. bluechips. Concluding remarks follow in Section 6.

2 Networks for Time Series

We are concerned in using a graphical model to represent the cross–sectional conditional dependence structure of a n–dimensional multivariate time series

$$\mathbf{y}_t = (y_{t\,1}, \ldots, y_{t\,n})'.$$

2.1 Partial Correlation Network

First let us assume that $\mathbf{y}_t$ is a white noise process. The standard network definition commonly used in the statistics literature is the partial correlation network, (see Dempster, 1972; Meinshausen and Bühlmann, 2006). The process $\mathbf{y}_t$ is represented as an undirected weighted 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 $\mathbf{y}_t$. 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 partially correlated conditionally on all other components of the process. Partial correlation is a linear measure of cross–sectional conditional dependence and it is defined as the correlation between $y_{t\,i}$ and $y_{t\,j}$ conditionally on all other variables in the system, that is

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

The set of edges of the partial correlation network can then be expressed as

$$\mathcal{E} = \left\{(i, j) \in \mathcal{V} \times \mathcal{V} \,:\, \rho^{ij} \neq 0\right\},$$

and ρ^{ij} is the weight associated with the edge between i and j. It is typically assumed that the partial correlation structure of the process is sparse (see e.g. Meinshausen and Bühlmann, 2006; Peng *et al.*, 2009). The objective of network estimation is to use a realization of the process to detect which edges are present and to estimate the corresponding partial correlations.

It is useful to recall a number of properties of partial correlation. First, if the white noise is also Gaussian, then absence of partial correlation implies conditional independence. Second, partial correlation is related to linear regression. Consider the linear regression model where variable i is regressed on all other remaining variables in the system, that is

$$y_{t\,i} = c + \beta_{1i} y_{t\,1} + \ldots + \beta_{(i-1)i} y_{t\,i-1} + \beta_{(i+1)i} y_{t\,i+1} + \ldots + \beta_{ni} y_{t\,n} + e_{t\,i}$$

where $\mathsf{E}(e_{t\,i}) = 0$, $\mathsf{Var}(e_{t\,i}) = \sigma_i^2$ and $\mathsf{E}(y_{t\,j}e_{t\,i}) = 0$ for all $j \neq i$. It can be shown that, for any $j \neq i$, the ρ^{ij} partial correlation is different from zero if and only if β_{ij} is different from zero. In other words, partial correlation signals that $y_{t\,j}$ has additional information for predicting $y_{t\,i}$ after conditioning on all the other variables in the system (and vice versa). Partial correlation is also related to (simple) correlation. If in the partial correlation network there exists a path between the two vertices, then the two vertices are correlated (and vice versa). Finally, partial correlation can be expressed as a function of the inverse of the covariance matrix of the process. Let $\mathbf{\Sigma} = \mathsf{Cov}(\mathbf{y}_t, \mathbf{y}_t)$ and let $\mathbf{K} \equiv \mathbf{\Sigma}^{-1}$, which is also commonly referred to as the concentration matrix. Let k_{ij} denote the (i, j)-th entry of $\mathbf{K}$. Then, the (i, j)-th partial correlation can be expressed as

$$\rho^{ij} = \frac{-k_{ij}}{\sqrt{k_{ii}k_{jj}}}. \tag{1}$$

This relation implies that the structure of the partial correlation network is fully characterized by the concentration matrix of the process. This property is important for estimation, in that network estimation is typically reformulated as a sparse concentration matrix estimation problem.

EXAMPLE 1. *Let $\mathbf{y}_t$ be a zero mean Gaussian white noise process with $n = 5$ and with concentration matrix*

$$\mathbf{K} \equiv \mathbf{\Sigma}^{-1} = \begin{bmatrix} 1.0 & -0.2 & 0.0 & 0.0 & 0.0 \\ -0.2 & 1.0 & 0.5 & -0.3 & 0.0 \\ 0.0 & 0.5 & 1.0 & 0.0 & 0.0 \\ 0.0 & -0.3 & 0.0 & 1.0 & -0.7 \\ 0.0 & 0.0 & 0.0 & -0.7 & 1.0 \end{bmatrix}.$$

The partial correlation network associated with this process is plotted in Figure 1.

2.2 Long Run Partial Correlation Network

Partial correlation is an exhaustive measure of conditional dependence for white noise processes. However, this measure has limitations when considering more general time series

processes. The main shortcoming is that partial correlation captures only contemporaneous dependence between the variables in the system. In a time series context, however, cross–sectional dependence is not necessarily just contemporaneous.

In this work we propose a network definition for time series that attempts to overcome the limitations of partial correlations. The network definition we adopt is based on long run partial correlation, a measure of cross–sectional conditional dependence constructed from the long run covariance matrix of the process. The long run covariance matrix of the process $\mathbf{y}_t$ can be defined as

$$\mathbf{\Sigma}_L \equiv \lim_{M\to\infty} \frac{1}{M} \mathsf{Cov}\left(\sum_{t=1}^{M} \mathbf{y}_t, \sum_{t=1}^{M} \mathbf{y}_t \right), \tag{2}$$

assuming the limit exists. The main highlight of the definition in equation (2) is that the rescaled limiting covariance of the aggregated process summarises cross–sectional dependences at any lead/lag. Also, our definition simplifies to the standard definition of partial correlation network in case the time series process is a white noise. Clearly, our measure of dependence in (2) only captures linear dependence, and it does not capture other types of cross–sectional dependence.

We present a stylized example for a 2–dimensional process in order to point out the differences between the (simple) covariance and the long run covariance matrix.

EXAMPLE 2. *Let* $\mathbf{y}_t$ *be a* VMA *process defined as*

$$\begin{bmatrix} y_{t\,1} \\ y_{t\,2} \end{bmatrix} = \begin{bmatrix} \epsilon_{t\,1} \\ \epsilon_{t\,2} \end{bmatrix} + \begin{bmatrix} 0 & \psi \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{t-1\,1} \\ \epsilon_{t-1\,2} \end{bmatrix}$$

with

$$\begin{bmatrix} \epsilon_{t\,1} \\ \epsilon_{t\,2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right),$$

and $\mathsf{Cov}(\epsilon_{t\,j}, \epsilon_{t-h\,i}) = 0$*, for any* i, j *and* $h \neq 0$*. The contemporaneous correlation between*

the two components is zero

$$\mathsf{Cor}\,(y_{t\,1}, y_{t\,2}) = 0.$$

The correlation of the aggregated process over $M = 12$ *periods is*

$$\mathsf{Cor}\left(\sum_{t=1}^{12} y_{t\,1}, \sum_{t=1}^{12} y_{t\,2}\right) = \left(\frac{11}{12}\right) \frac{\psi}{\sqrt{1+\psi^2}},$$

where it is trivial to see that the correlation is increasing in the absolute value of the spillover parameter ψ*. The long run correlation between the two processes is*

$$\lim_{M\to\infty} \mathsf{Cor}\left(\sum_{t=1}^{M} y_{t\,1}, \sum_{t=1}^{M} y_{t\,2}\right) = \lim_{M\to\infty} \frac{M-1}{M} \frac{\psi}{\sqrt{1+\psi^2}} = \frac{\psi}{\sqrt{1+\psi^2}}. \tag{3}$$

This example highlights a number of important properties. The (simple) covariance does not exhaustively characterize cross–sectional dependence in a dynamic setting. On the other hand, the covariance of the aggregated process provides more comprehensive description of the conditional dependence structure. Also, the example shows that if the memory of the process has a sufficiently fast rate of decay, then the convergence of the long run covariance to its limit is fast. The bottom line of the example is that in a serially dependent multivariate system the cross–sectional dependence structure can be richer than what is measured by the covariance.

Our network definition has a general interpretation in terms of spectral density matrix of a covariance stationary vector process, which is defined as

$$\mathbf{s}_{\mathbf{y}}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \mathbf{\Gamma}_{\mathbf{y}\,h}\, e^{-ih\omega}, \quad \omega \in [-\pi, \pi],$$

where $\mathbf{\Gamma}_{\mathbf{y}\,h} = \mathsf{Cov}(\mathbf{y}_t, \mathbf{y}_{t-h})$ and $i = \sqrt{-1}$. As it is well known, the spectral density is related to the long run covariance in that the spectrum evaluated at zero frequency corresponds to the long run covariance:

$$\mathbf{\Sigma}_L = 2\pi\, \mathbf{s}_{\mathbf{y}}(0) = \sum_{h=-\infty}^{+\infty} \mathbf{\Gamma}_{\mathbf{y}\,h},$$

which highlights how $\mathbf{\Sigma}_L$ contains the linear dependences of $\mathbf{y}_t$ at every lead and lag.

EXAMPLE 3. *For the* VMA *defined in the previous example we have:*

$$\mathbf{\Gamma}_{\mathbf{y}\,0} = \begin{bmatrix} \sigma^2(1+\psi^2) & 0 \\ 0 & \sigma^2 \end{bmatrix}, \quad \mathbf{\Gamma}_{\mathbf{y}\,1} = \begin{bmatrix} 0 & \psi\sigma^2 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{\Gamma}_{\mathbf{y}\,-1} = \mathbf{\Gamma}'_{\mathbf{y}\,1},$$

and $\mathbf{\Gamma}_{\mathbf{y}\,h} = \mathbf{0}$ *for* $|h| > 1$. *The spectral density matrix is, for* $\omega \in [-\pi, \pi]$,

$$2\pi\, \mathbf{s}_{\mathbf{y}}(\omega) = \mathbf{\Gamma}_{\mathbf{y}\,0} + \mathbf{\Gamma}_{\mathbf{y}\,1} e^{-i\omega} + \mathbf{\Gamma}_{\mathbf{y}\,-1} e^{i\omega}.$$

which can also be computed as

$$2\pi\, \mathbf{s}_{\mathbf{y}}(\omega) = \left(\mathbf{I} + \begin{bmatrix} 0 & \psi \\ 0 & 0 \end{bmatrix} e^{-i\omega} \right) \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \left(\mathbf{I} + \begin{bmatrix} 0 & 0 \\ \psi & 0 \end{bmatrix} e^{i\omega} \right).$$

In both cases we have

$$\mathbf{\Sigma}_L = 2\pi\, \mathbf{s}_{\mathbf{y}}(0) = \begin{bmatrix} \sigma^2(1+\psi^2) & \sigma^2\psi \\ \sigma^2\psi & \sigma^2 \end{bmatrix},$$

and the long run correlation in (3) *follows.*

As for contemporaneous partial correlations, it is possible to characterize the network on the basis of the inverse of the long run covariance matrix. Let long run concentration matrix be defined as

$$\mathbf{K}_L \equiv \mathbf{\Sigma}_L^{-1},$$

and let $k_{L\,ij}$ denote its (i,j)-th entry. We can then express long run partial correlation coefficient for series i and j as

$$\rho_L^{ij} = \frac{-k_{L\,ij}}{\sqrt{k_{L\,ii} k_{L\,jj}}}. \tag{4}$$

The Long Run Partial Correlation network is defined as a weighted undirected network

$\mathcal{N}_L = (\mathcal{V}, \mathcal{K})$ where the set of edges $\mathcal{K}$ is defined as

$$\mathcal{K} = \left\{(i,j) \in \mathcal{V} \times \mathcal{V} \,:\, \rho_L^{ij} \neq 0\right\}, \tag{5}$$

and ρ_L^{ij} is the weight associated with the edge between i and j.

In this work we develop an estimation approach for $\mathbf{K}_L$ which is inspired by the literature on HAC covariance estimators (Newey and West, 1987; Andrews and Monahan, 1992; Den Haan and Levin, 1994). Under appropriate assumptions on the data generating process, we can approximate $\mathbf{y}_t$ with an infinite order VAR:

$$\mathbf{y}_t = \sum_{k=1}^{\infty} \mathbf{A}_k \mathbf{y}_{t-k} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}), \tag{6}$$

where $\mathbf{A}_k$ are $n \times n$ matrices such that the process is stable. The spectral density matrix of representation (6) is

$$\mathbf{s}_{\mathbf{y}}(\omega) = \frac{1}{2\pi} \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k e^{-ik\omega}\right)^{-1} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}} \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k' e^{ik\omega}\right)^{-1}, \quad \omega \in [-\pi, \pi]. \tag{7}$$

Therefore, the long run covariance and the long run concentration matrix are given by

$$\boldsymbol{\Sigma}_L = \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k\right)^{-1} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}} \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k'\right)^{-1},$$

$$\mathbf{K}_L \equiv \boldsymbol{\Sigma}_L^{-1} = \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k'\right) \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}^{-1} \left(\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k\right). \tag{8}$$

This expression has an appealing interpretation in that it factorizes the long run concentration in a sandwich form determined by the term $\mathbf{I} - \sum_{k=1}^{\infty} \mathbf{A}_k$, which captures long run dynamic relations of the system, and the term $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}^{-1}$, which accounts for the contemporaneous dependence of the system innovations. For ease of notation we express (8) more compactly as

$$\mathbf{K}_L = (\mathbf{I} - \mathbf{G})' \, \mathbf{C} \, (\mathbf{I} - \mathbf{G}), \tag{9}$$

where $\mathbf{G} \equiv \sum_{k=1}^{\infty} \mathbf{A}_k$ and $\mathbf{C} \equiv \mathbf{\Gamma}_{\boldsymbol{\epsilon}}^{-1}$, and we are going to refer to the two matrices respectively as the Granger and Contemporaneous components of the long run concentration matrix.

The formula in (9) has an appealing graphical interpretation. In order to better appreciate this, let us first introduce the Granger and Contemporaneous networks associated with the VAR representation of equation (6). We define g_{ij} as the (i,j)-th entry of $\mathbf{G}$, and c_{ij} as the (i,j)-th entry of $\mathbf{C}$. Notice also that in general $g_{ij} \neq g_{ji}$ while $c_{ij} = c_{ji}$. The Granger network is defined as a weighted directed network $\mathcal{N}_G = (\mathcal{V}, \mathcal{G})$ where the presence of an edge from i to j denotes that i Granger causes j in the long run, that is

$$\mathcal{G} = \{(i,j) \in \mathcal{V} \times \mathcal{V} \,:\, g_{ji} \neq 0\}, \tag{10}$$

and g_{ji} is the weight associated with the edge from i to j. The contemporaneous network is defined as the partial correlation network of the VAR innovations. It is a weighted undirected network $\mathcal{N}_C = (\mathcal{V}, \mathcal{C})$ where an edge between i and j denotes that i is partially correlated to j conditionally on the past, that is

$$\mathcal{C} = \left\{(i,j) \in \mathcal{V} \times \mathcal{V} \,:\, \rho^{ij} \neq 0\right\}, \tag{11}$$

where $\rho^{ij} = -c_{ij}/\sqrt{c_{ii}c_{jj}}$ is the weight associated with the edge between i and j. The Long Run Partial Correlation network can be interpreted as a combination of the Granger and Contemporaneous networks.

Building on Meinshausen and Bühlmann (2006) in this work we are concerned in the case in which the Long Run Partial Correlation network is sparse. However, in this work we do not formulate sparsity assumptions on the $\mathbf{K}_L$ matrix directly. Instead, we formulate sparsity assumptions on the approximating VAR model. Precise details on the sparsity assumptions are given in the theoretical analysis in Section 3 and the Appendix. Under such assumptions, the sparsity of the VAR model determines the sparsity structure of the $\mathbf{G}$ and $\mathbf{C}$ components and of the long run concentration matrix $\mathbf{K}_L$. Due to the

sandwich form of $\mathbf{K}_L$ its sparsity structure is not trivial. The generic (i,j)-th element of $\mathbf{K}_L$ is given by

$$k_{Lij} = \sum_{h=1}^{n}\sum_{k=1}^{n} g_{ih}c_{hk}g_{jk}, \qquad i,j = 1,\dots n.$$

In order to see which relations are mapped into an edge of the Long Run Partial Correlation network, we decompose, in the case that $i \neq j$, the previous sums as follows

$$\begin{aligned}
k_{Lij} &= \sum_{h=1}^{n}\sum_{k=1}^{n} g_{ih}c_{hk}g_{jk} = \\
&= g_{ii}\sum_{\ell=1}^{n} c_{i\ell}g_{j\ell} + g_{jj}\sum_{h=1}^{n} g_{ih}c_{hj}(1-\delta_{hi}) + \sum_{h=1}^{n}\sum_{k=1}^{n} g_{ih}c_{hk}g_{jk}\,(1-\delta_{hi})(1-\delta_{kj}) \\
&= g_{ii}c_{ij}g_{jj} + g_{ii}c_{ii}g_{ji} + g_{ij}c_{jj}g_{jj} \\
&\quad + g_{ii}\sum_{\ell=1}^{n} c_{i\ell}g_{j\ell}\,(1-\delta_{\ell i})(1-\delta_{\ell j}) + g_{jj}\sum_{\ell=1}^{n} g_{i\ell}c_{\ell j}(1-\delta_{\ell i})(1-\delta_{\ell j}) \\
&\quad + \sum_{h=1}^{n}\sum_{k=1}^{n} g_{ih}c_{hk}g_{jk}\,(1-\delta_{hi})(1-\delta_{kj})(1-\delta_{hk}) \\
&\quad + \sum_{\ell=1}^{n} g_{i\ell}c_{\ell\ell}g_{j\ell}(1-\delta_{\ell i})(1-\delta_{\ell j}),
\end{aligned}$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$. Using this formula we see that there is an edge between nodes i and j, i.e. $k_{Lij} \neq 0$ if

1. i and j are contemporaneously partially correlated, i.e. $c_{ij} = c_{ji} \neq 0$;
2. i Granger causes j in the long run, i.e. $g_{ji} \neq 0$;
3. j Granger causes i in the long run, i.e. $g_{ij} \neq 0$;
4. there exists another node k such that $g_{ik} \neq 0$ and $c_{kj} \neq 0$ or $g_{jk} \neq 0$ and $c_{ki} \neq 0$.

Being linked in the Granger and Contemporaneous networks is a sufficient condition to be linked in the long run but it is not necessary. The last condition shows that the set of links of the Long Run Partial Correlation network is larger than the union of the links of the Granger and Contemporaneous networks.

Notice that, by looking at the formula of the spectral density matrix of a VAR process (see equation (7)) it is clear that with our method we can estimate not only the Long Run

Partial Correlation network but also the inverse of the spectrum at different frequencies. The sparsity structure will be the same at all frequencies but the weights attached to the edges will change with the frequency.

We conclude this Section with two examples showing the sparseness of the Long Run Partial Correlation network as a combination of the two components.

EXAMPLE 4. *Consider a case with* $n = 3$ *and* $g_{11} \neq 0$, $g_{23} \neq 0$ *and* $c_{13} = c_{31} \neq 0$, *while* $g_{13} = g_{31} = g_{12} = g_{21} = c_{12} = c_{21} = c_{23} = c_{32} = 0$ *and we can show that still* $k_{L12} \neq 0$. *Indeed:*

$$\begin{aligned} k_{L12} &= g_{11}c_{12}g_{22} + g_{11}c_{11}g_{21} + g_{12}c_{22}g_{22} + g_{11}c_{13}g_{23} + g_{13}c_{32}g_{22} \\ &\quad + g_{12}c_{21}g_{21} + g_{12}c_{23}g_{23} + g_{13}c_{31}g_{21} + g_{13}c_{33}g_{23} = g_{11}c_{13}g_{23} \end{aligned}$$

The intuition is that, since 1 and 3 are contemporaneously partially correlated ($c_{13} = c_{31} \neq 0$), *and 3 Granger causes 2 in the long run* ($g_{23} \neq 0$), *then, because 1 has serial dependence* ($g_{11} \neq 0$), *we have also that 1 indirectly Granger causes series 2 in the long run, therefore 1 and 2 are partially correlated in the long run given 3. The resulting Long Run Partial Correlation netowrk is shown graphically in Figure 2.*

EXAMPLE 5. *Consider a* VAR*(1) with* $n = 6$

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}})$$

where

$$\mathbf{A}_1 = \begin{bmatrix} 0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 0.0 & 0.6 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.4 & 0.2 & -0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.4 \end{bmatrix},$$

and

$$\mathbf{\Gamma}_{\boldsymbol{\epsilon}}^{-1} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & -0.2 & 0.0 & -0.3 \\ 0.0 & 0.0 & -0.2 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & -0.3 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

Then, the long run concentration matrix is

$$\mathbf{K}_L = \begin{bmatrix} 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & -0.1 \\ 0.0 & 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.1 & -0.5 & -0.2 & -0.2 \\ 0.0 & 0.0 & -0.5 & 0.6 & 0.2 & 0.0 \\ 0.0 & 0.0 & -0.2 & 0.2 & 0.6 & 0.0 \\ -0.1 & 0.0 & -0.2 & 0.0 & 0.0 & 0.4 \end{bmatrix}.$$

Figure 3 displays the Long Run Partial Correlation network, as well as the Granger and Contemporaneous networks. Note that vertices 3 and 5 are not connected nor in the Contemporaneous or in the Granger network, however they are connected in the Long Run Partial Correlation network. Notice that the strength of the Long Run Partial Correlation network links differs from the strength of the Granger and Contemporaneous network links.

3 Network Estimation

3.1 LASSO Estimation

We obtain sparse estimators using the LASSO (Least Absolute Shrinkage and Selection Operator), an estimation technique introduced in the statistics literature by Tibshirani (1996). Consider a general linear regression model

$$\mathcal{Y}_i = \boldsymbol{\vartheta}_0' \mathcal{X}_i + e_i, \qquad i = 1, \ldots, T,$$

with $\mathcal{X}_i$ and $\boldsymbol{\vartheta}_0$ being P-dimensional vectors, and $e_i \sim i.i.d.(0, s^2)$. The LASSO estimator of this model is defined as the minimizer of the ℓ_1–penalised least squares objective function

$$\widehat{\boldsymbol{\vartheta}}_{T,\mathsf{L}} = \arg\min_{\boldsymbol{\vartheta}} \frac{1}{T}\sum_{i=1}^{T}\left(\mathcal{Y}_i - \boldsymbol{\vartheta}'\mathcal{X}_i\right)^2 + \frac{\lambda_T}{T}\sum_{j=1}^{P}|\vartheta_j|, \tag{12}$$

where $\lambda_T \geq 0$ is a tuning parameter that determines the degree of the penalization and ϑ_j is the j–th component of the parameters vector $\boldsymbol{\vartheta}$. The LASSO is a shrinkage type estimation procedure that shrinks the OLS estimate towards zero. The amount of shrinkage is determined by the value of the penalty λ_T: the greater the λ_T the greater the shrinkage effect. An important feature of the ℓ_1–penalty is that the solution of equation (12) delivers estimates of the parameter vector that contain exact zeros, i.e. a "sparse" solution. Under appropriate conditions on the model and the degree of penalization, it can be shown that if the true parameter vector $\boldsymbol{\vartheta}_0$ is sparse, then the LASSO consistently estimates the non–zero parameters while the others are shrunk to exact zero. Another advantage of the LASSO (which is also shared by other shrinkage estimators) is the fact that the estimator can be well defined even when the number of parameters P is larger than the number of observations T (see Fan and Peng, 2004). This property is crucial in high–dimensional modelling problems where this is indeed the case.

Several variants of the LASSO have been proposed in the literature in order to consider the case of dependent data. In particular, in this work we make use of the Adaptive LASSO estimator proposed by Zou (2006), and defined as

$$\widehat{\boldsymbol{\vartheta}}_{T,\mathsf{AL}} = \arg\min_{\boldsymbol{\vartheta}} \frac{1}{T}\sum_{i=1}^{T}\left(\mathcal{Y}_i - \boldsymbol{\vartheta}'\mathcal{X}_i\right)^2 + \frac{\lambda_T}{T}\sum_{j=1}^{P} w_j|\vartheta_j|,$$

where the w_j is a weight on the penalty for ϑ_j and is typically chosen as the reciprocal of the absolute value of a pre-estimator of ϑ_j. Among the possible choices of the pre-estimator the most popular are the OLS estimator or the ridge estimator when P is larger than T.

3.2 NETS

The VAR approximation has two appealing features in this setting. The expression of the long run concentration matrix in (8) lends itself to relatively straightforward estimation. Moreover, estimators based on this representation deliver consistent estimates of the long run concentration matrix for quite general classes of time series processes, including, for instance, heteroskedastic processes.

In this Section we propose an algorithm called NETS (Network Estimator for Time Series) to estimate sparse Long Run Partial Correlation networks. The NETS procedure consists of reformulating the estimation of the long run concentration matrix of the VAR representation of the process in (9) in two LASSO regression problems. Our strategy consists of estimating the two components $\mathbf{G}$ and $\mathbf{C}$ of the network and to then combine the two to obtain an estimator of $\mathbf{K}_L$. In the rest of this Section we describe the estimation procedure and provide its large sample properties. In particular, the method we propose delivers not only an estimate of the Long Run Partial Correlation network but also consistent estimates of its two components: (*i*) the Granger network that summarises the long run Granger causality structure of the process and (*ii*) the Contemporaneous network of the VAR innovations that synthesises contemporaneous dependence conditionally on the past information. Depending on the data at hand both components may contain interesting information and it might be worth to consider them separately. Detailed assumptions and proofs of the results of this Section are in Appendix.

In what follows we assume that a sample of T observations is available for the $\mathbf{y}_t$ process. Assumptions 1, 2, 3 and 4 part (a) characterize the process. More in detail, assumption 1 states that $\mathbf{y}_t$ is a zero mean, purely stochastic process belonging to the class of weakly dependent processes. This is a general class of processes introduced by Doukhan and Louhichi (1999) which nests several types of mixing processes as well as covariance stationary processes. Indeed, in assumption 2 we assume to have covariance stationary process whose spectral density and covariance matrices are positive definite. This assumption is in turn used to establish the existence of a VAR(∞) representation in

the same way as in Den Haan and Levin (1994). In particular, assumption 3 specifies the rate of decay of the autocovariances which allows us to compute the asymptotic bias due to the truncation error when estimating a VAR(p).

Finally, assumption 4 part (a) contains assumptions on the rates of the growth of the dimension of the model with respect to the sample size. More specifically, the rate of growth of the number of series is $n = O(T^{\zeta_1})$ where $\zeta_1 > 0$. The data is then approximated by a finite p-th order VAR

$$\mathbf{y}_t = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{y}_{t-k} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}}), \quad t = 1, \ldots, T. \tag{13}$$

where the order of the autoregression is allowed to grow with the sample size at the rate $p = O(T^{\zeta_2})$ where $\zeta_2 > 0$. Notice that due the truncation at lag p the VAR is biased and the bias depends on the number of series n, the lag p and rate of decay of the autocovariance. Assumption 4 part (a.iii) gives the relation between the relative rates of n and p in order to have a truncation error which is asymptotically negligible. In particular, for a given rate of decay of the autocovariances, the larger is n the larger must be p.

Hereafter, we adopt the following conventions. Although both n and p depend on T we omit this dependence for notational convenience while we keep the dependence on T for the estimated quantities, the penalization constant, and the number of non–zero parameters to be introduced below. We use the index $_0$ to denote the true values of the parameters and we use $\widehat{\ }$ to denote estimated quantities.

Granger Network Estimator. The Granger component matrix $\mathbf{G}_0$ with generic elements $g_{0\,ij}$ is estimated by estimating a p-th of order VAR model via the Adaptive LASSO. As in least squares, estimation of VAR can be decomposed in n separate regressions each with T observations and np variables, which is typically more convenient for the numerical implementation. For each of the n equations $\boldsymbol{\alpha}_i = (a_{1i1} \ldots a_{1in} \ldots a_{pi1} \ldots a_{pin})'$, where $a_{k\,ij}$

is the (i,j)-th element of $\mathbf{A}_k$. The Adaptive LASSO estimator of $\boldsymbol{\alpha}_i$ is then defined as

$$\widehat{\boldsymbol{\alpha}}_{Ti} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{T}\sum_{t=1}^{T}(y_{t\,i} - \boldsymbol{\alpha}_i'\mathbf{z}_t)^2 + \frac{\lambda_T^{\mathbf{G}}}{T}\sum_{j=1}^{np}\left|\frac{\alpha_{ij}}{\widetilde{\alpha}_{Ti\,j}}\right|, \quad i = 1,\ldots,n, \tag{14}$$

where $\widetilde{\boldsymbol{\alpha}}_{Ti}$ is a pre-estimator. We denote the set of non–zero coefficients of one VAR equation as $\mathcal{A}_i$ which has $q_{Ti}^{\mathbf{G}}$ elements. In assumption 4 part (b) we require for any i

$$q_{Ti}^{\mathbf{G}} = o\left(\sqrt{\frac{T}{\log T}}\right), \quad \frac{\lambda_T^{\mathbf{G}}}{T}\sqrt{q_{Ti}^{\mathbf{G}}} = o(1), \text{ and } \frac{\lambda_T^{\mathbf{G}}}{T}\sqrt{\frac{T}{\log T}} \to \infty.$$

The following theorem establishes the large sample properties of the Adaptive LASSO estimator (14).

THEOREM 1. *Under assumptions 1, 2, 3, 4 parts (a) and (b), 5, and condition 1 in Appendix, the following propositions are true for any $i = 1,\ldots,n$.*

(a) (Consistent Estimation of the Restricted Problem) Consider the constrained minimizer $\widehat{\boldsymbol{\alpha}}_{Ti}^{\mathcal{A}_i}$ of (14) with $\widehat{\alpha}_{Ti\,j}^{\mathcal{A}_i} = 0$ for $j \in \mathcal{A}_i^c$. Then for T large enough and for any $\eta > 0$ the estimator exists and there exists a constant $\kappa_{\mathbf{G}}$ such that

$$\left|\left|\widehat{\boldsymbol{\alpha}}_{Ti}^{\mathcal{A}_i} - \boldsymbol{\alpha}_{0i}\right|\right|_2 \leq \kappa_{\mathbf{G}}\,\frac{\lambda_T^{\mathbf{G}}}{T}\sqrt{q_{Ti}^{\mathbf{G}}},$$

and, for any $j \in \mathcal{A}_i$, $\operatorname{sign}(\widehat{\alpha}_{Ti\,j}^{\mathcal{A}_i}) = \operatorname{sign}(\alpha_{0i\,j})$ with at least probability $1 - O(T^{-\eta})$.

(b) (Consistent Selection of the Unrestricted Problem) Consider the minimizer $\widehat{\boldsymbol{\alpha}}_{Ti}$ of (14). Then for T large enough and for any $\eta > 0$

$$\widehat{\alpha}_{Ti\,j} = 0 \;\; \textit{for} \;\; j \in \mathcal{A}_i^c,$$

with at least probability $1 - O(T^{-\eta})$.

(c) (Oracle) Consider the minimizer $\widehat{\boldsymbol{\alpha}}_{Ti}$ of (14). Then for T large enough and for any $\eta > 0$

$$||\widehat{\boldsymbol{\alpha}}_{Ti} - \boldsymbol{\alpha}_{0i}||_2 \leq \kappa_{\mathbf{G}}\,\frac{\lambda_T^{\mathbf{G}}}{T}\sqrt{q_{Ti}^{\mathbf{G}}},$$

and, for any j, $\text{sign}(\widehat{\alpha}_{Tij}) = \text{sign}(\alpha_{0ij})$, *with at least probability* $1 - O(T^{-\eta})$.

The first statement of the theorem says that the Adaptive LASSO estimator, when restricted to the non–zero coefficients, is consistent (as $\frac{\lambda_T^{\mathbf{G}}}{T}\sqrt{q_{Ti}^{\mathbf{G}}} = o(1)$) and their sign is also estimated consistently. The second statement says that the unrestricted LASSO estimator correctly selects the non–zero coefficients asymptotically. Finally, the last statement says that the unrestricted estimator is also consistent.

The strategy adopted to prove the previous theorem follows the literature of LASSO estimation of large dimensional linear time series models (see e.g. Fan and Peng, 2004; Meinshausen and Bühlmann, 2006; Peng *et al.*, 2009). We are contributing to the existing results in two ways. First we consider a large dimensional panel of time series which requires Adaptive LASSO estimation. Second we address the VAR(∞) case and the bias due to truncation at lag p. Such bias is of order $O(n^2/p^\beta)$, where β is the rate of decay of the autocovariances and by assumption 4 part (a.iii) it can be controlled for. While the first contribution is similar to the ones by Kock (2012); Kock and Callot (2012); Medeiros and Mendes (2012), the second contribution, to our knowledge, is new to the literature.

The estimator of the VAR matrices constructed from $\widehat{\boldsymbol{\alpha}}_{Ti}$ is denoted by $\widehat{\mathbf{A}}_{Tk}$ and we have

$$\widehat{\mathbf{G}}_T = \sum_{k=1}^{p} \widehat{\mathbf{A}}_{Tk},$$

with generic entries $\widehat{g}_{Tij}$. The estimator of the set of the Granger network edges, $\mathcal{G}$ as defined in (10), is

$$\widehat{\mathcal{G}}_T = \{(i,j) \in \mathcal{V} \times \mathcal{V} : \widehat{g}_{Tij} \neq 0\}.$$

Under the additional assumptions on the structure of the coefficients a_{0kij} of the VAR approximation, we are able to establish consistency of the estimated Granger network. In particular, assumption 7 imposes that (i) if for each lag $k < p$ we have $a_{0kij} = 0$, then for each $k > p$ we require $a_{0kij} = 0$; and that (ii) if for any lag k, $a_{0kij} \neq 0$, then, for any h, we have that $\sum_{k=1}^{h} a_{0kij} \neq 0$. These are reasonable assumptions that respectively rule out the cases of (i) zero coefficients followed by non–zero ones and (ii) exact cancellations of non–zero coefficients. We then have the result on the Granger network estimation.

COROLLARY 1. *Under assumptions 1, 2, 3, 4 parts (a), (b), 5, 7 and condition 1 in Appendix, for any* $\eta > 0$ *and for* T *large enough*

$$\widehat{\mathcal{G}}_T = \mathcal{G},$$

with probability at least $1 - O(T^{-\eta})$. *Moreover, for each* $i, j = 1, \ldots, n$

$$p\text{–}\lim_{T\to\infty} \widehat{g}_{T\,ij} = g_{0\,ij}.$$

Contemporaneous Network Estimator. The contemporaneous component matrix $\mathbf{C}_0$ which has generic elements $c_{0\,ij}$ is estimated using a strategy put forward in Peng *et al.* (2009). It consists of estimating the concentration matrix of the innovation term in the VAR representation (6). Define for any VAR equation $\widehat{\epsilon}_{t\,i} = (y_{t\,i} - \widehat{\boldsymbol{\alpha}}'_{i\,T}\mathbf{z}_t)$. Then, the entries of the inverse correlation matrix are $\rho_0^{ij} = -c_{0\,ij}/\sqrt{c_{0\,ii}c_{0\,jj}}$ are the coefficients of the regression

$$\widehat{\epsilon}_{t\,i} = \sum_{j\neq i}^{n} \rho_0^{ij} \sqrt{\frac{\widehat{c}_{T\,ii}}{\widehat{c}_{T\,jj}}} \widehat{\epsilon}_{t\,j} + u_{t\,i}, \quad i = 1, \ldots, n. \tag{15}$$

where $\widehat{c}_{ii\,T}$ are pre-estimators of $c_{0\,ii}$. This estimation problem is a regression of n variables on $\frac{n(n-1)}{2}$ explanatory variables. Given sparsity of $\mathbf{C}_0$, we estimate the n equations as (15) contemporaneously by means of the LASSO estimator

$$\widehat{\boldsymbol{\rho}}_T = \arg\min_{\boldsymbol{\rho}} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \left(\widehat{\epsilon}_{t\,i} - \sum_{j\neq i}^{n} \rho^{ij} \sqrt{\frac{\widehat{c}_{T\,ii}}{\widehat{c}_{T\,jj}}} \widehat{\epsilon}_{t\,j} \right)^2 + \frac{\lambda_T^C}{T} \sum_{i=2}^{n} \sum_{j=1}^{i-1} |\rho^{ij}|. \tag{16}$$

The estimator of the set of the Contemporaneous network edges, $\mathcal{C}$ as defined in (11), is

$$\widehat{\mathcal{C}}_T = \{(i,j) \in \mathcal{V} \times \mathcal{V} : \widehat{\rho}_T^{ij} \neq 0\},$$

while the estimator of the matrix $\mathbf{C}_0$ is denoted as $\widehat{\mathbf{C}}_T$ and has generic entries $\widehat{c}_{T\,ij} = -\widehat{\rho}_T^{ij} \sqrt{\widehat{c}_{T\,ii}\widehat{c}_{T\,jj}}$. Finally, the pre–estimators $\widehat{c}_{T\,ii}$ must satisfy condition 2 in Appendix

and are given by

$$\widehat{c}_{T\,ii} = \left[\frac{1}{T-1} \sum_{t=1}^{T} \widehat{u}_{t\,i}^{2} \right]^{-1} \tag{17}$$

where $\widehat{u}_{t\,i}$ are the residuals of regression (15). Estimation is achieved by iterating between (16) and (17).

If we denote the total number of non–zero parameters, i.e. the number of elements of $\mathcal{C}$, by $q_T^{\mathbf{C}}$, then in assumption 4 part (c) we require

$$q_T^{\mathbf{C}} = o\left(\sqrt{\frac{T}{\log T}} \right), \quad \frac{\lambda_T^{\mathbf{C}}}{T} \sqrt{q_T^{\mathbf{C}}} = o(1), \text{ and } \frac{\lambda_T^{\mathbf{C}}}{T} \sqrt{\frac{T}{\log T}} \to \infty.$$

The following theorem establishes the large sample properties of the LASSO estimator (16).

THEOREM 2. *Under assumptions 1, 2, 3, 4 parts (a), (b), (c), and (d), 5, 6, and conditions 1, 2 in Appendix, the following propositions are true*

(a) (Consistent Estimation of the Restricted Problem) Consider the constrained minimizer $\widehat{\boldsymbol{\rho}}_T^{\mathcal{C}}$ *of* (A-5) *with* $\widehat{\rho}_T^{ij\,\mathcal{C}} = 0$ *for* $(i,j) \in \mathcal{C}^c$. *Then for* T *large enough and for any* $\eta > 0$ *the estimator exists and there exists a constant* $\kappa_{\mathbf{C}}$ *such that*

$$\left|\left| \widehat{\boldsymbol{\rho}}_{T\,\mathcal{C}}^{\mathcal{C}} - \boldsymbol{\rho}_{0\,\mathcal{C}} \right|\right|_2 \leq \kappa_{\mathbf{C}} \frac{\lambda_T^{\mathbf{C}}}{T} \sqrt{q_T^{\mathbf{C}}},$$

and, for any $(i,j) \in \mathcal{C}$, $\operatorname{sign}(\widehat{\rho}_{T\,ij}^{ij\,\mathcal{C}}) = \operatorname{sign}(\rho_0^{ij})$, *with at least probability* $1 - O(T^{-\eta})$.

(b) (Consistent Selection of the Unrestricted Problem) Consider the minimizer $\widehat{\boldsymbol{\rho}}_T$ *of* (A-5). *Then for* T *large enough and for any* $\eta > 0$

$$\widehat{\rho}_T^{ij} = 0 \;\; \textit{for} \;\; (i,j) \in \mathcal{C}^c,$$

with at least probability $1 - O(T^{-\eta})$.

(c) (Oracle) Consider the minimizer $\widehat{\boldsymbol{\rho}}_T$ *of* (A-5). *Then for* T *large enough and for any*

$\eta > 0$

$$||\widehat{\boldsymbol{\rho}}_T - \boldsymbol{\rho}_0||_2 \leq \kappa_{\mathbf{C}} \frac{\lambda_T^{\mathbf{C}}}{T} \sqrt{q_T^{\mathbf{C}}},$$

and, for any (i,j), $\text{sign}(\widehat{\rho}_T^{ij}) = \text{sign}(\rho_0^{ij})$, *with at least probability* $1 - O(T^{-\eta})$.

The first statement of the theorem says that the LASSO estimator, when restricted to the non–zero coefficients, is consistent (as $\frac{\lambda_T^{\mathbf{C}}}{T}\sqrt{q_{Ti}^{\mathbf{C}}} = o(1)$) and their sign is also estimated consistently. The second statement says that the unrestricted LASSO estimator correctly selects the non–zero coefficients asymptotically. Finally, the last statement says that the unrestricted estimator is also consistent.

In the proof, analogous to the one of theorem 1, we have also to take into account both the truncation bias term due to the VAR(p) approximation, which as explained before is controlled by assumption 4 part (a), and the bias due to the first step estimation which is of order $O(n\sqrt{q_{Ti}^{\mathbf{G}}}\lambda_T^{\mathbf{G}}/T)$ and can be controlled by assumption 4 part (d). In particular, as also explained in remark 1 in Appendix, this last assumption implies that if we want a consistent estimator of both network components we must have a tradeoff relation between the number of non–zero parameters in the VAR and the total number of series n. The higher is the first the smaller must be the second (and viceversa). To our knowledge, the analysis of the consistency of the high–dimensional sparse partial correlation network estimator when the data is subject to this type of measurement error is novel in the literature.

Finally, the following corollary is a straightforward consequence of theorem 2.

COROLLARY 2. *Under assumptions 1, 2, 3, 4 parts (a), (b), (c), and (d), 5, 6, 7, and conditions 1, 2 in Appendix, for any* $\eta > 0$ *and for* T *large enough*

$$\widehat{\mathcal{C}}_T = \mathcal{C},$$

with probability at least $1 - O(T^{-\eta})$. *Moreover, for each* $i, j = 1, \ldots, n$

$$p\text{–}\lim_{T\to\infty} \widehat{\rho}_T^{ij} = \rho_0^{ij}.$$

Long Run Partial Correlation Network Estimator. Finally, we can estimate the long run concentration matrix $\mathbf{K}_{L0}$ with generic entries $k_{L0\,ij}$, by combining the sparse estimators of the Granger and Contemporaneous networks

$$\widehat{\mathbf{K}}_{L\,T} = (\mathbf{I} - \widehat{\mathbf{G}}_T)'\,\widehat{\mathbf{C}}_T\,(\mathbf{I} - \widehat{\mathbf{G}}_T),$$

which has generic entries $\widehat{k}_{LT\,ij}$. Then the generic entries of the Long Run Partial Correlation network and their estimators are respectively

$$\rho^{ij}_{L0} = -\frac{k_{L0\,ij}}{\sqrt{k_{L0\,ii}k_{L0\,jj}}} \text{ and } \widehat{\rho}^{ij}_{LT} = -\frac{\widehat{k}_{LT\,ij}}{\sqrt{\widehat{k}_{LT\,ii}\widehat{k}_{LT\,jj}}}.$$

Finally, the estimator of the set of the Long Run Partial Correlation network edges, $\mathcal{K}$ as defined in (5), is

$$\widehat{\mathcal{K}}_T = \{(i,j) \in \mathcal{V} \times \mathcal{V} : \widehat{\rho}^{ij}_{LT} \neq 0\},$$

and we have the final result on network estimation.

COROLLARY 3. *Under assumptions 1, 2, 3, 4 parts (a), (b), (c), and (d), 5, 6, 7, and conditions 1, 2 in Appendix, for any $\eta > 0$ and for T large enough*

$$\widehat{\mathcal{K}}_T = \mathcal{K},$$

with probability at least $1 - O(T^{-\eta})$. Moreover, for each $i, j = 1, \ldots, n_T$

$$p\text{–}\lim_{T\to\infty} \widehat{\rho}^{ij}_{LT} = \rho^{ij}_{L0}.$$

4 Simulation

In this Section we provide a simple illustration of our estimation algorithm using simulated data. The exercise consists of simulating a multivariate time series process and using the NETS algorithm to detect the linkages of the Long Run Partial Correlation network. We

replicate the experiment for 1000 times and report the Monte Carlo estimates of the ROC curves of the procedure to assess the performance of the edge detection algorithm.

In line with our empirical application, we simulate a 40 dimensional system. Recall that the number of possible links in a 40 dimensional system is 780. We assume that the system is generated by a VAR(1). The sparsity structure the Granger and Contemporaneous networks is determined by the Erdős-Rényi random graph model. The parameters of the VAR(1) are set as

$$\mathbf{A}_1 = 0.3\,\boldsymbol{\iota}_G \text{ and } \boldsymbol{\Gamma}_\epsilon^{-1} = \mathbf{I} - 0.2\,\boldsymbol{\iota}_C,$$

where $\boldsymbol{\iota}_G$ and $\boldsymbol{\iota}_C$ are adjacency matrices obtained by the Erdős-Rényi random graph model. The networks are kept fixed across simulations and in the simulation results we report below the number of Granger links is 78 and the number of contemporaneous links is 44. For each replication we run the NETS algorithm over a range of pairs of $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ values. The series of pairs is such that both $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ simultaneously increases. We perform the experiment for different sample sizes $T = 250, 500, 750, 2500$.

We replicate the simulation 1000 times for each sample size and in each replication we compute the proportion of type 1 and type 2 errors. The simulation results are used to compute the ROC curve, that is the plot of the true positive rate (TPR) versus the false positive rate (FPR). Note that the penalization coefficient determines the true positive rate and the false positive rate. When $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ are small (large), the proportion of type 1 errors is high (low) while the proportion of type 2 errors is low (high). Unfortunately, besides special cases, the mapping between the penalization constants and the true positive rate and false positive rate is unknown.

Figure 4 reports the ROC curves for the different sample sizes. The curves show that as the sample size T increases the performance of the classifier, which is measure by the area underneath the ROC curves, increases rapidly. Overall, the simulation results convey that the estimation algorithm performs satisfactorily in a sparse high–dimensional setting.

5 Empirical Application

We illustrate the methodology proposed in this paper with an application to financial risk networks. The application is motivated by the 2007/2009 financial crisis and is connected to the strand of literature concerned with the development of quantitative techniques for the measurement of systemic risk in the economy (see also Adrian and Brunnermeier, 2009; Brownlees and Engle, 2012; Bisias, Flood, Lo, and Valavanis, 2012; Brunnermeier and Oehmke, 2012). One of the lessons learnt from the crisis is that high levels of interconnectedness between companies can be damaging for the entire economy. When the degree of codependence is high, a large individual shock to a small set of firms can have potentially vast rippling effects to all other interconnected institutions.

We use stock returns to construct a market based estimate of the network of interconnections for a panel of companies. The application is inspired by the recent contributions of Billio *et al.* (2012) and Diebold and Yilmaz (2013). Billio *et al.* (2012) use a network definition based on bivariate Granger causality and, similarly to this application, analyse the network of interconnections among firms using monthly returns. On the other hand, Diebold and Yilmaz (2013) use a network definition based on variance decompositions and analyse the network of interconnections among volatilities using volatility measures.

We consider a panel of U.S. bluechips across different industry sectors. The list of company names and industry groups is provided in Table 6. We work with compound monthly returns in between January 1990 to December 2010, which correspond to a sample size of 251 observations. Classical asset pricing theory models like the CAPM or APT imply that the unexpected return of these risky assets can be expressed as a linear function of a few common factors and an idiosyncratic component. The presence of common factors is however in contrast with the assumption of sparsity. Common factors represent systematic components through which all asset returns are correlated. Thus in order to estimate the Long Run Partial Correlation network from the data we first have to control for common sources of variation. We consider here the case when factors are

observable and in particular we rely on a simple one factor model, that is

$$r_{t\,i} = \beta_i r_{t\,m} + z_{t\,i}, \tag{18}$$

where $r_{t\,i}$ is the rate of return for firm i, $r_{t\,m}$ is the return of the market factor and $z_{t\,i}$ is the idiosyncratic shock of firm i. The objective is then to estimate the network structure of the idiosyncratic shocks $z_{t\,i}$. As it is commonly assumed in empirical finance, the market factor is treated as observed and here we use the monthly rate of return on the S&P 500 index. In this setting, our asymptotic results will also depend on the estimation error of β_i which can be easily accounted for. In some cases, the common factors are unobservable and need to be estimated (see e.g. Forni, Hallin, Lippi, and Reichlin, 2000; Stock and Watson, 2002; Bai, 2003). However, in this case estimation of the network will depend also on the estimation error of the unobservable factors. We plan to address this issue in future research.

The idiosyncratic risk network is estimated as follows. For each asset in the panel, we first estimate the β_i coefficient of equation (18) by least squares and construct the series of residuals. We then run NETS on the panel of residuals. The order of the VAR approximation p is set to one. The LASSO penalties $\lambda_T^{\mathbf{G}}$ and $\lambda_T^{\mathbf{C}}$ are determined on the basis of the BIC information criterion by searching over a grid of possible values.

The NETS algorithm estimate of the idiosyncratic risk network contains 57 edges out of 820 possible ones. The estimated long run covariance matrix associated with the network is also positive definite. Approximately 88% of the edges come from the Contemporaneous network component. The vast majority of the partial correlations are positive and the ones that are negative have small magnitudes from an economic perspective. Interestingly, the linkages detected in the idiosyncratic risk network account for a significant portion of the overall variation in stock returns. While the variability explained by the market factor is 25% on average, the detected linkages account for an extra 15% on average.

The estimated idiosyncratic risk network is reported in Figure 5. In the figure, the vertex diameter is proportional to the size of the firm the vertex color denotes the industry

group. We can group vertices in two categories: vertices that belong to a connected component and vertices that are not connected to any other company. The vertices that are not interconnected can be associated with industry groups that can be considered as more peripheral from an idiosyncratic risk perspective. In fact, all Consumer Discretionary like Disney (DIS), McDonalds (MCD) and Home Depot (HD) are disconnected. Figure 6 shows the subnetwork of the connected component. In this subgraph many of the links are within the same industry group, a phenomenon that in the social network analysis is called similarity: similar firms tend to be connected. Besides the intra–industry linkages, the network gives insights on the inter–industry linkages and shows how Financial and Technology company are the industries that are more interconnected with the other groups.

Network plots can be visually rich and it is sometimes challenging to grasp all the information encoded in these graphs. To this end, it is useful to use techniques introduced in the network literature to decode and synthesise the information in the graph. In particular, in this analysis we focus on centrality and clustering.

The objective of centrality analysis is to determine which vertices are more central in the network. Notions of centrality have been associated with systemic risk: a shock to a node that is very central can potentially have vast rippling effects. This is put forward in Dungey *et al.* (2012). Several centrality indices have been proposed in the literature. Among other proposals, a simple measure of centrality is the number of edges of a node: the higher the number of edges, the more interconnected and central a node is. Another popular measure of centrality is the measure of eigenvector centrality, which is at the basis of the PageRank algorithm used by Google to rank search results (Dungey *et al.*, 2012). In Figure 7, we report another plot of connected component of the idiosyncratic risk network. In this plot however, we set the diameter of the vertices proportional to the number of edges. Interestingly, the plot shows how Financial companies and some Technology companies are central in the networks: AIG (AIG), Bank of America (BAC), Citigroup (C) and Apple (AAPL). In order to get further insights on the drivers of centrality, we regress firm network characteristics on firm characteristics, that is we regress the number

of vertices of each firm on industry dummies and (log) size. The results are reported in Table 6. Results show that the only significant effects are the Financial, Material and Technology Fixed Effects. Interestingly, results show that size does not matter. This might be driven by the fact that size is not comparable across different industries. Also, we acknowledge that regression results should be interpreted with caution given the small sample size of 41 firms. Lastly, Figure 8 reports the Lorentz curve associated with the number of edges in the network. The plot shows that there is a moderate degree of "network inequality", in the sense the top 20% vertices in terms of number of edges account for more than 50% of the interconnections in the network.

Another interesting network characteristic that is often encountered in social networks is clustering. Clustering refers to the fact that neighbouring vertices tend to be connected to a similar set of neighbours: A is connected to C and B is connected to C, then it is likely that A and B are also connected. Again, different types of indices can be used to measure clustering. Here we use a local measure of clustering: for each pair of vertices, we compute the proportion of common neighbour of the two vertices. We compute this index for the idiosyncratic risk network. Out of 802 pairs, 172 have a clustering index greater than zero (approximately 20%). We report the histogram of the index in Figure 9 (conditioning on the pairs with positive clustering). The histograms shows that the density of ties in the network can be quite high. For those pairs whose clustering index is greater than zero, the average proportion of common neighbours is 0.246. High degree of clustering implies that the network has "small–world" characteristics: if clustering is large then the number of edges connecting any two vertices is going to be small. From a risk perspective, *ceteris paribus*, higher clustering implies that a shock to an individual node is going to have larger effects in the network.

Overall, the results show that the idiosyncratic risk network estimated by the NETS algorithm provides interesting insights on the conditional dependence structure of idiosyncratic shocks in the panel. The network exhibits several of the empirical regularities typically encountered in social networks.

6 Conclusions

In this work we have introduced novel network techniques for the analysis of high–dimensional multivariate time series data. We provide a network definition that is specifically tailored for multivariate time series based on long run partial correlations. We propose a simple estimation procedure based on the LASSO that we name NETS to estimate Long Run Partial Correlation networks. The large sample properties of the procedure are analysed and we provide conditions for consistent estimation of the sparse high–dimensional networks. Finally, we illustrate the methodology with a financial network analysis application.

Acknowledgements

We would like to thank Tony Hall, Nikolaus Hautsch, Julien Idier, Clifford Lam, Marco Lippi, Gabor Lugosi, Tommaso Proietti, Kimmo Soramaki for their helpful comments. Special thanks go also to Giorgio Calzolari. We also thank the participants to the workshop "Recent Developments in Econometrics" at Universitat Rovira i Virgili (October 26, 2012), the UPF Finance Lunch Seminar (November 9, 2012), the conference "New Tools for Financial Regulation" at Banque de France (November 15-16, 2012), the CORE–ECORE/LSM Seminar at UCLouvain (December 17, 2012), the conference ICEEE 2013 at Università di Genova (January 16–18, 2013), the ULB–ECARES Statistics seminar (February 8, 2013), the Bundesbank seminar (February 12, 2013), the LSE Joint Econometrics and Statistics workshop (March 1, 2013), the London School of Hygiene and Tropical Medicine Statistics seminar (March 5, 2013), the 3rd Humboldt–Copenhagen Conference on Financial Econometrics (March 16–18, 2013), the Economics seminar at IHS in Vienna (March 26, 2013), the Economics seminar at Università di Pisa (April 15, 2013), the Economics seminar at IAE/UAB (April 17, 2013), the workshop on "High Dimensional Time Series in Macroeconomics and Finance" at IHS in Vienna 2013 (May 2–4, 2013), the Financial Econometrics conference at the Toulouse School of Economics (May 17–18, 2013), the Economics seminar at Universidad de Alicante (May 26, 2013), the Barcelona GSE Summer Forum on Time Series Analysis in Macro and Finance workshop (June 10–11, 2013), the Italian Statistical Society conference (June 19–21, 2013), the EUI Measuring and Modeling Financial Risk with High Frequency Data workshop (June 27–29, 2013), the 2013 NBER-NSF Time Series Conference at the Federal Reserve Board (September 26-27, 2013), the Economics seminar at Koç University (October, 2, 2013). All mistakes are ours.

References

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

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2013). Systemic Risk and Stability in Financial Networks. Technical report.

Adrian, T. and Brunnermeier, M. K. (2009). CoVaR. Technical report, FRB of New York. Staff Report No. 348.

Allen, F. and Gale, D. (2000). Financial Contagion. *Journal of Political Economy*, **108**, 1–33.

Andrews, D. W. K. and Monahan, J. C. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, **60**, 953–966.

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

Bai, J. and Liao, Y. (2012). Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood. arXiv 1209.5911v2.

Belloni, A., Chernozhukov, V., and Hansen, C. (2011). Inference for High–Dimensional Sparse Econometric Models. *Advances in Economics and Econometrics*. 10th World Congress of Econometric Society.

Bickel, P. J. and Levina, E. (2008). High Dimensional Inference and Random Matrices. *The Annals of Statistics*, **36**, 2577–2604.

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.

Bisias, D., Flood, M., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. Working paper #0001, Office of Financial Research.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

Brownlees, C. and Engle, R. (2012). Volatility, Correlations and Tails for Systemic Risk Measurement. Technical report, New York University.

Brunnermeier, M. K. and Oehmke, M. (2012). Bubbles, financial crises, and systemic risk. Technical report, Princeton.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High–Dimensional Data: Methods, Theory and Applications*. Springer, New York.

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

Davis, R. A., Zang, P., and Zheng, T. (2012). Sparse Vector Autoregressive Modeling. arXiv 1207.0520v1.

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.

Dedecker, J., Doukhan, P., Lang, G., León, J., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. Springer, New York.

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

Den Haan, W. J. and Levin, A. (1994). Inferences from Parametrics and Non–Parametric Covariance Matrix Estimation Procedures. Technical report, UCSD and Federal Reserve Board.

Diebold, F. and Yilmaz, K. (2013). On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics*, page forthcoming.

Doukhan, P. and Louhichi, S. (1999). A New Weak Dependence Condition and Applications to Moment Inequalities. *Stochastic Processes and their Applications*, **84**, 313–342.

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.

Doukhan, P. and Neumann, M. H. (2008). The Notion of ψ–weak Dependence and its Applications to Bootstrapping Time Series. *Probability Surveys*, **5**, 146–168.

Dungey, M., Luciani, M., and Veredas, D. (2012). Ranking Systemically Important Financial Institutions. Technical report, ECARES.

Eichler, M. (2007). Granger Causality and Path Diagrams for Multivariate Time Series. *Journal of Econometrics*, **137**, 334–353.

Fan, J. and Peng, H. (2004). Nonconcave Penalized Likelihood with a Diverging Number of Parameters. *The Annals of Statistics*, **32**, 928–961.

Fan, J., Liao, Y., and Mincheva, M. (2011). High Dimensional Covariance Matrix Estimation in Approximate Factor Models. *The Annals of Statistics*, **39**, 3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society, Series B*. forthcoming.

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.

Gallant, R. A. (1987). Nonlinear Methods in Econometrics. In J. Eatwell, M. Milgate, and P. Newman, editors, *The New Palgrave*, volume 3 (K–P). Stockton Press New York.

Hautsch, N., Schaumburg, J., and Schienle, M. (2012). Financial Network Systemic Risk Contributions. Technical report, Discussion Paper 2012-053, CRC 649, Humboldt-Universität zu Berlin.

Hautsch, N., Schaumburg, J., and Schienle, M. (2013). Forecasting Systemic Impact in Financial Networks. Technical report, Discussion Paper 2013-008, CRC 649, Humboldt-Universität zu Berlin.

Jarociński, M. and Maćkowiak, B. (2011). Choice of Variables in Vector Autoregressions. Technical report.

Kock, A. B. (2012). On the Oracle Property of the Adaptive Lasso in Stationary and Nonstationary Autoregressions. CREATES Research Papers 2012-05, School of Economics and Management, University of Aarhus.

Kock, A. B. and Callot, L. A. (2012). Oracle Inequalities for High Dimensional Vector Autoregressions. CREATES Research Papers 2012-16, School of Economics and Management, University of Aarhus.

Lam, C. and Fan, J. (2009). Sparsistency and Rates of Convergence in Large Covariance Matrix Estimation. *The Annals of Statistics*, **37**, 4254–4278.

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.

Medeiros, C. M. and Mendes, E. (2012). Estimating High–Dimensional Time Series Models. Technical report, Pontifical Catholic University of Rio de Janeiro.

Meinshausen, N. and Bühlmann, P. (2006). High Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, **34**, 1436–1462.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

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.

Peterson, C., Vannucci, M., Karakas, C., Choi, W., Ma, L., and Maletić-Savatić, M. (2013). Inferring Metabolic Networks Using Baysian Adaptive Graphical Lasso with Informative Priors. *Statistics and Its Interface*, page forthcoming.

Pourahmadi, M. (2011). Covariance Estimation: The GLM and the Regularization Perspectives. *Statistical Science*, **26**, 369–387.

Song, S. and Bickel, P. J. (2011). Large Vector Auto Regressions. arXiv 1106.3915v1.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, **20**, 147–162.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

White, H. (1984). *Asymptotic Theory for Econometricians*. New York: Academic Press.

Xavier, F., Parigi, B. M., and Rochet, J.-C. (2000). Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank. *Journal of Money, Credit and Banking*, **32**, 611–638.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418–1429.

Table 1: U.S. Bluechips

ID	Tick	Name	Sector
1	BRK	Berkshire Hathaway B	Financial
2	C	Citigroup Inc.	Financial
3	BAC	Bank of America Corp.	Financial
4	AXP	American Express Co.	Financial
5	AIG	American International Group Inc.	Financial
6	MCD	McDonalds Corp.	Consumer Discretionary
7	DIS	Walt Disney Co.	Consumer Discretionary
8	HD	Home Depot Inc.	Consumer Discretionary
9	PG	Procter & Gamble Co.	Consumer Staples
10	KO	Coca-Cola Co.	Consumer Staples
11	WMT	Wal-Mart Stores Inc.	Consumer Staples
12	PEP	PepsiCo Inc.	Consumer Staples
13	SLB	Schlumberger Ltd.	Energy
14	OXY	Occidental Petroleum Corp.	Energy
15	PFE	Pfizer Inc.	Healthcare
16	JNJ	Johnson & Johnson	Healthcare
17	MRK	Merck & Co. Inc.	Healthcare
18	ABT	Abbott Laboratories	Healthcare
19	AMGN	Amgen Inc.	Healthcare
20	GE	General Electric Co.	Industrials
21	UTX	United Technologies Corp.	Industrials
22	UNP	Union Pacific Corp.	Industrials
23	MMM	3M Co.	Industrials
24	CAT	Caterpillar Inc.	Industrials
25	BA	Boeing Co.	Industrials
26	DD	E.I. DuPont de Nemours & Co.	Materials
27	FCX	Freeport-McMoRan Copper & Gold Inc.	Materials
28	DOW	Dow Chemical Co.	Materials
29	NEM	Newmont Mining Corp.	Materials
30	PPG	PPG Industries Inc.	Materials
31	AAPL	Apple Inc.	Technology
32	IBM	International Business Machines Corp.	Technology
33	MSFT	Microsoft Corp.	Technology
34	T	AT&T Inc.	Technology
35	ORCL	Oracle Corp.	Technology
36	CSCO	Cisco Systems Inc.	Technology
37	INTC	Intel Corp.	Technology
38	DUK	Duke Energy Corp.	Utilities
39	SO	Southern Co.	Utilities
40	D	Dominion Resources Inc. (Virginia)	Utilities
41	AEP	American Electric Power Co. Inc.	Utilities

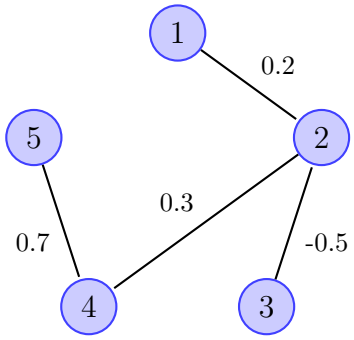
List of company names and sectors.

Table 2: Network Regression

Variable	Estimate
Consumer Staples FE	0.62
Energy FE	1.15
Financial FE	5.66**
Healthcare FE	2.05
Industrials FE	1.33
Materials FE	4.91*
Technology FE	2.76*
Utilities FE	1.23
(Log) Size	0.06
R^2	0.40

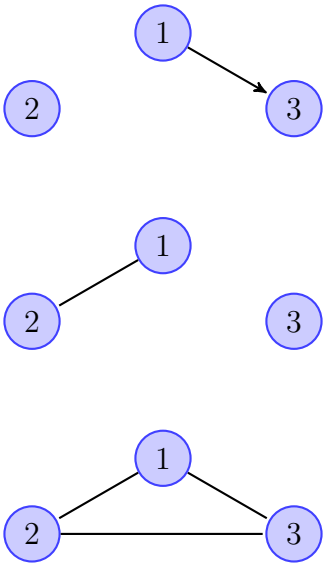
Number of Node Degree and Firm Characteristics.

Figure 1: Partial Correlation Network Example.



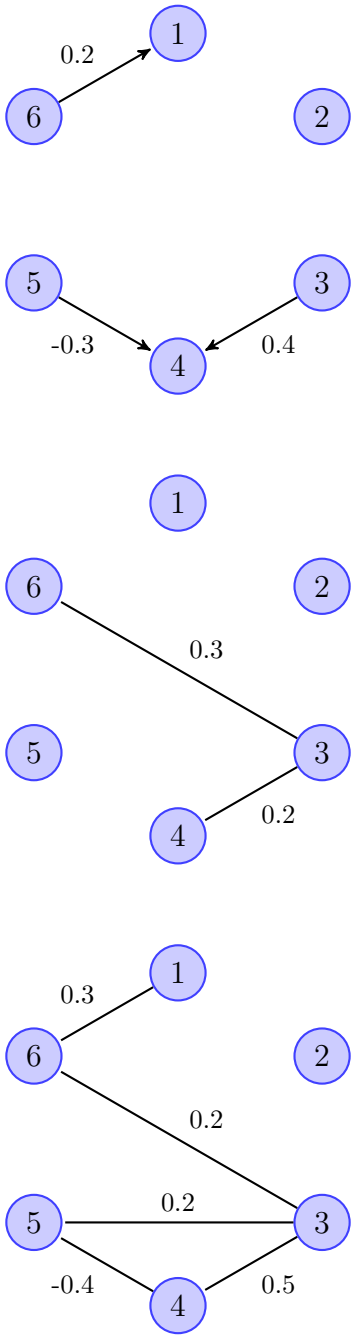
The figure displays the partial correlation network associated with Example 1.

Figure 2: Long Run Partial Correlation Network Example.



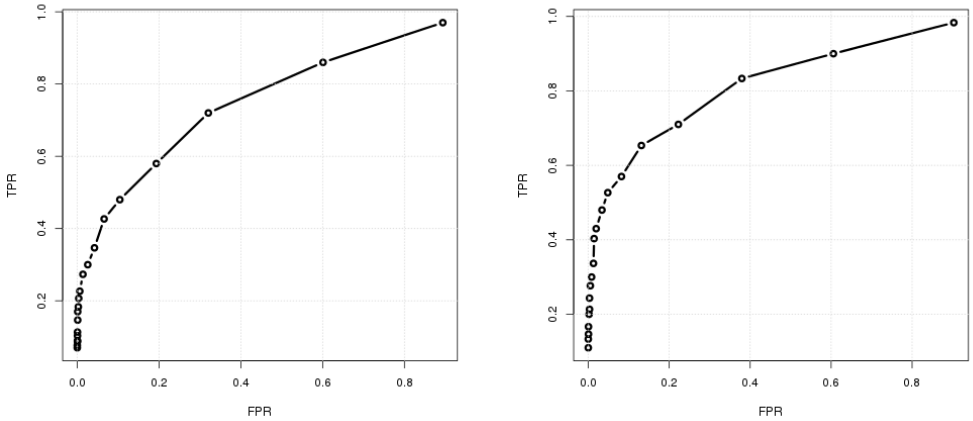
The figure displays the Granger (top), Contemporaneous (middle), and Long Run Partial Correlation (bottom) networks associated with Example 4.

Figure 3: Long Run Partial Correlation Network Example.

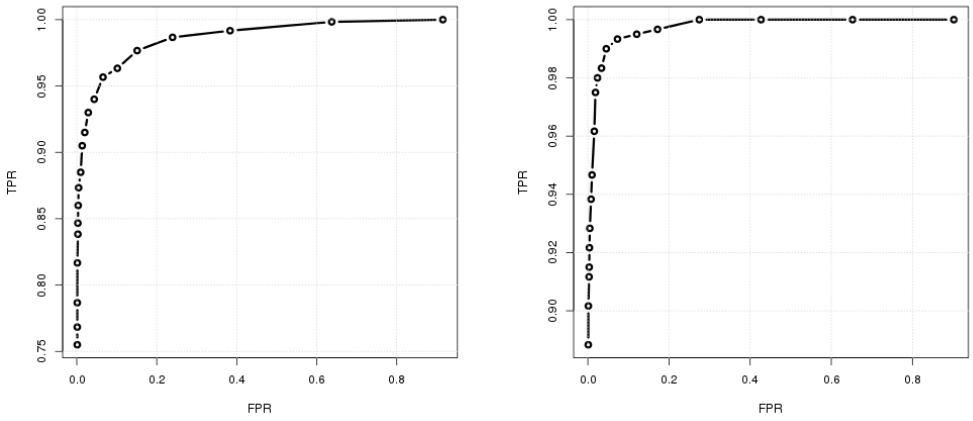


The figure displays the Granger (top), Contemporaneous (middle), and Long Run Partial Correlation (bottom) networks associated with Example 5.

Figure 4: Monte Carlo Evidence



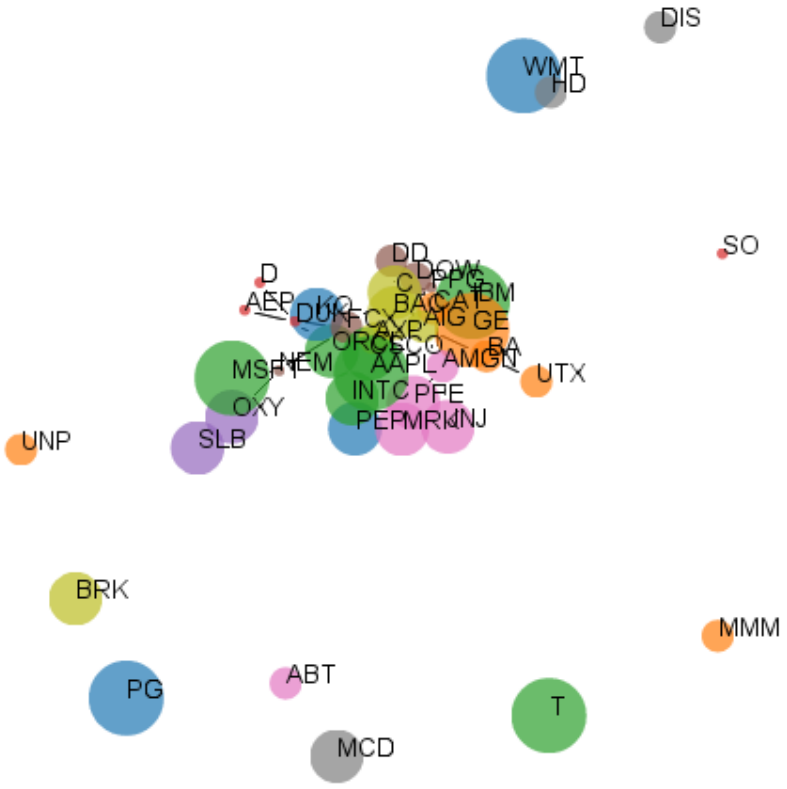
T=250 T=500



T=750 T=2500

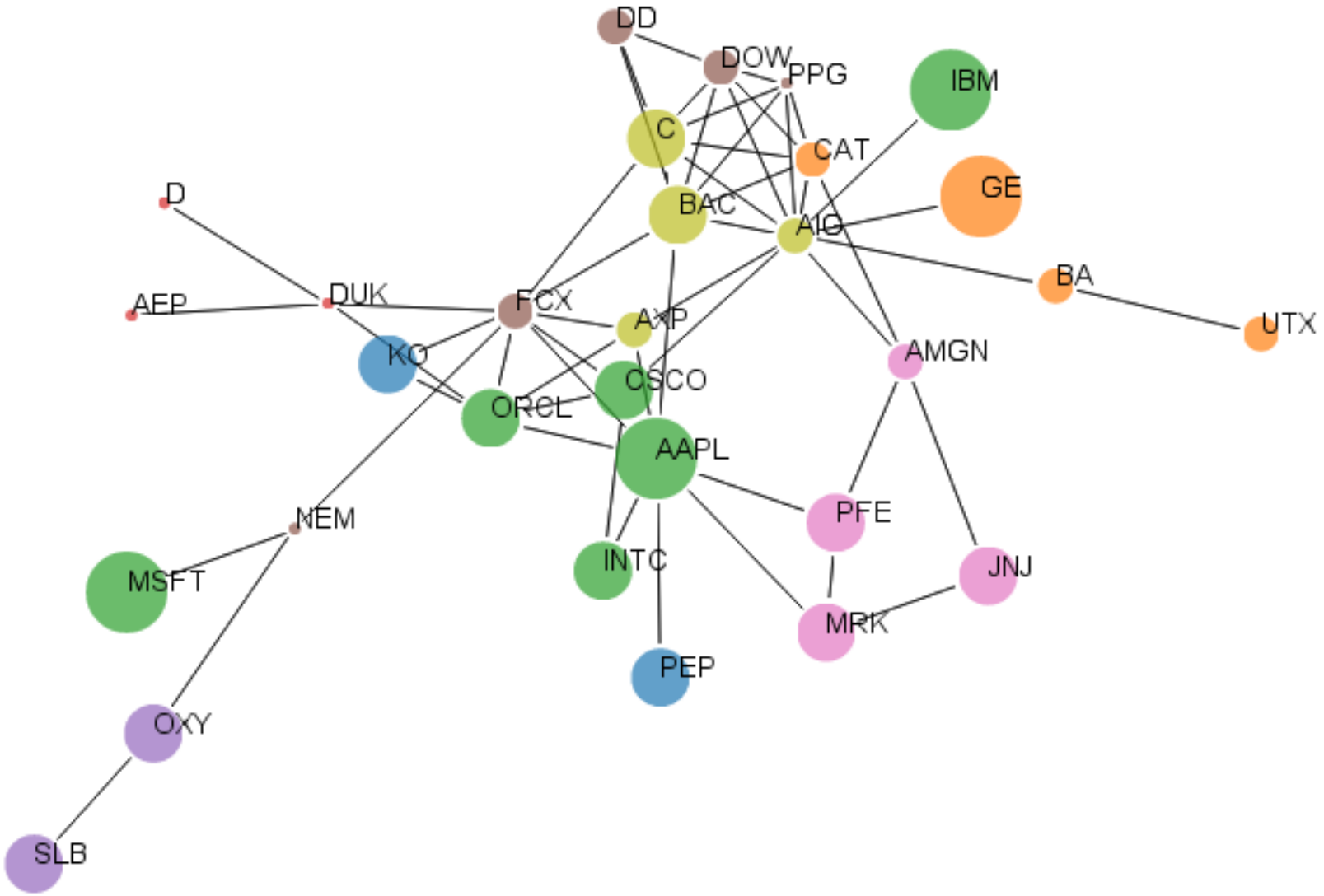
The figure displays the ROC curves associated with the Monte Carlo experiment of Section 4. The set of penalization constants is $\lambda = \{l\, 0.1\, T : l = 1, ..., 20\}$ for both LASSO regressions.

Figure 5: Idiosyncratic Risk Network.



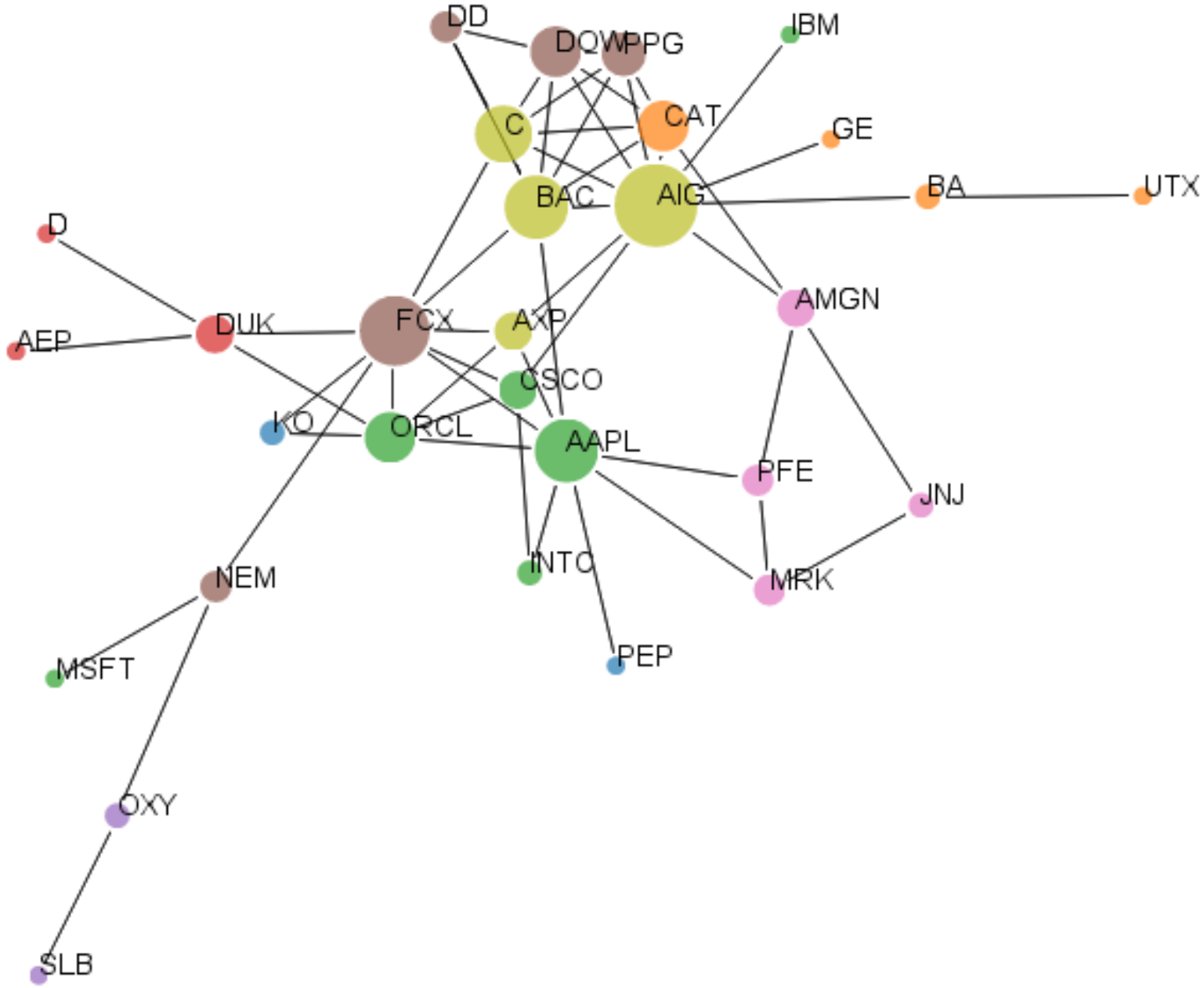
Plot of the full idiosyncratic risk network.

Figure 6: Idiosyncratic Risk Network.



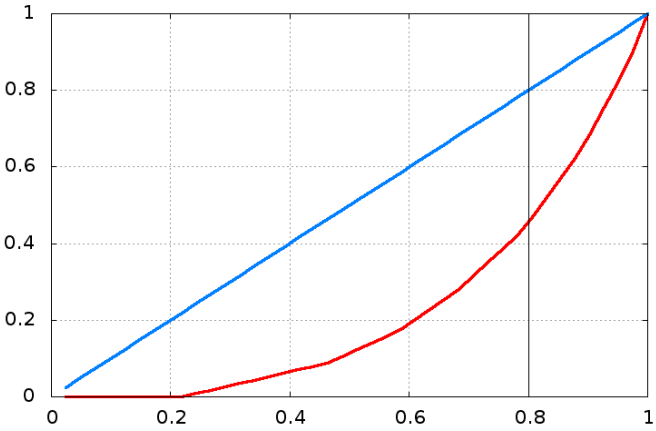
Plot of the connected component of the idiosyncratic risk network.

Figure 7: Idiosyncratic Risk Network.



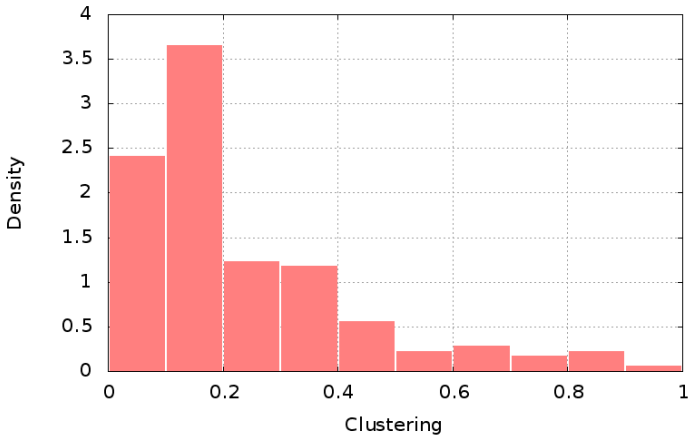
Plot of the connected component of the idiosyncratic risk network. The node size denotes centrality in the network.

Figure 8: Edge Concentration.



The figure shows the Lorentz curve of the edge concentration.

Figure 9: Clustering



The figure shows the histogram of the number of common neighbours for all the nodes pairs with at least one neighbour in common.