La collana della Scuola di Alta Formazione Dottorale accoglie le migliori tesi di dottorato dell'Università degli Studi di Bergamo, insignite della dignità di stampa e sottoposte a procedura di *blind peer review*.

Network theory is a powerful tool for the analysis of complex systems, and in recent years a growing body of literature highlights its financial applications. This work explores two fields of application of network theory in finance. The first is the modelization of systemic risk and financial contagion in a banking network, the second is portfolio optimization. The work contributes to the literature in several directions: first, the use of tlasso is introduced in the financial literature. Then a novel network version of ΔCoVaR (a common systemic risk measure), and an estimation procedure based on the SCAD penalization framework is proposed. Concerning the study of systemic risk, this work is among the firsts to focus on the presence of a community structure in a banking system, that is particularly relevant in the European setting where national borders are still relevant divisions.

Gabriele Torri is a researcher at the University of Bergamo in the Department of Economics. His research focuses on the analysis of systemic risk in banking systems using network approaches, as well as on the study of quantitative portfolio optimization and risk management. He received his PhD at the university of Bergamo in 2019, and prior to start his position as a resercher, worked as PhD trainee at the European Central Bank, in the Systemic Risk and Financial Institutions division.

ISBN: 978-88-97413-44-8
DOI: 10.6092/978-88-97413-44-8

Gabriele Torri

NETWORK THEORY IN FINANCE

27

Collana della Scuola di Alta Formazione Dottorale
- 27 -

Gabriele Torri

NETWORK THEORY IN FINANCE
Applications to Financial Contagion Analysis and Portfolio Optimization

UNIVERSITÀ DEGLI STUDI DI BERGAMO

2021

Collana della Scuola di Alta Formazione Dottorale

Diretta da Paolo Cesaretti

Ogni volume è sottoposto a *blind peer review*.

ISSN: 2611-9927

Sito web: https://aisberg.unibg.it/handle/10446/130100

Gabriele Torri

NETWORK THEORY IN FINANCE
Applications to Financial Contagion Analysis and Portfolio Optimization

Università degli Studi di Bergamo

2021

Network Theory in Finance: Applications to Financial Contagion Analysis and Portfolio Optimization /
Gabriele Torri. – Bergamo :
Università degli Studi di Bergamo, 2021.
(Collana della Scuola di Alta Formazione Dottorale; 27)

ISBN: 978-88-97413-44-8
DOI: 10.6092/978-88-97413-44-8

Progetto grafico: Servizi Editoriali – Università degli Studi di Bergamo

via Salvecchio, 19
24129 Bergamo
Cod. Fiscale 80004350163
P. IVA 01612800167

https://aisberg.unibg.it/handle/10446/175652

Acknowledgements

First I would like to express my gratitude to my supervisors Rosella Giacometti and Tomáš Tichý for the great support, help and inspiration, and to Sandra Paterlini, that played a fundamental role in my PhD, and was as important as the two supervisors.
I am also grateful to all my professors in Bergamo and Brescia, especially the coordinator of the PhD program Adriana Gnudi, and to the late Marida Bertocchi. I thank my coauthors Vincenzo Russo, Maria Elena De Giuli and Gianluca Farina, as well as Zdeněk Zmeška, Zari Rachev, my former colleagues in Etica SGR, and all the other people that spent their precious time to teach me valuable things. They all influenced positively my academic growth and my career decisions, and I couldn't arrive here without them.
Thanks to all the friends met along the way, too many to greet individually: the fellow PhD students in Bergamo and Brescia, that shared a lot of hard work and legitimate fun in the notorious *Ufficio 206*, the colleagues from EBS University in Wiesbaden, and the ones in Ostrava VŠB-TU, plus many others met in conferences, specialization courses, and in the university department.
Finally, my biggest thanks goes to my family and to Marina, that supported me, even in the hardest moments.

The research project has been supported by the Czech Science Foundation (GACR) under projects 15-23699S, 17-19981S, and SGS research projects of VSB-TU Ostrava SP2017/32, SP2018/34, and within RRC/10/2017 support scheme of the Moravia-Silesia Region.

Contents

Introduction

Network theory in finance

The study of networks has a long history in the mathematical literature, and can be reconducted to the field of *graph theory*, a branch of combinatorics started in the XVIII century by the Swiss mathematician Leonhard Euler. The field witnessed a growth of interest in the 1970s, with the development of applications in the social sciences (see e.g. Granovetter, 1977; Freeman, 1978), but it is in the last two decades that become a fundamental tool for the study of complex systems, mostly thanks to the increasing availability of large and granular datasets and improvements in computational power.

A network can be defined as a collection of nodes and edges that connect pairs of them. It is possible to characterize the network by assigning to the edges certain attributes, for instance, they can have a direction (directed networks) or a weight (weighted networks).

In symbols, we can represent a network as:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \tag{1}$$

where $\mathcal{V}$ is the vertex set, and $\mathcal{E}$ is the set of the edges.

A growing body of literature highlights the usefulness of network theory in finance, presenting several applications to solve traditional problems such as the determination of asset prices in the market, or study the risk exposures of financial institutions, overcoming the limits of other techniques. Moreover, network theory may provide insights on classical results or mathematical techniques, giving new interpretations to traditional methodologies. Network applications span a large set of techniques, from agent-based models, used to study the formation of prices in the market, to models that describe the diffusion of financial contagion and to portfolio optimization. In each of the applications, the nodes and edges in a network are modelled differently. For instance, nodes can represent market traders, banks or assets in a portfolio and the edges

can describe for instance flows of information, the exposures in interbank markets, or statistical measures such as partial correlations.

Banking networks – literature review

In this section we outline the main applications of network theory related to the modelization of systemic risk and financial contagion in banking systems, showing the main channel of transmission and some relevant models. We also briefly review the effects of network topology in the diffusion of contagion.

Financial contagion, systemic risk and network theory

One of the most striking features of financial crises is that they often spread quickly across countries and institutions, and that small shocks affecting a particular region or a small group of institutions can cause contagion in the rest of the system and infect other economic sectors.
The focus on contagion become particularly relevant after the 2008 crisis and become associated with the concept of *systemic risk*, defined by Schwarcz, 2008 as: *The risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either the failure of a chain of markets or institutions or a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility*.
Further insights on the (loose) definition of contagion are proposed by Rigobon, 2016, that underlines how the three expressions "contagion", "spillovers" and "interconnectedness" have been defined in several ways in the academic literature, with the first indicating mostly the unexpected component of transmission of shocks in the system or change in behaviour during crises, while the other two terms refer typically to the expected component of common movements among different institutions and different markets. The distinction between the concepts is often tenuous and model dependent and, beyond the semantic, the two major points of interest in the study of systemic risk are the estimation of transmission during normal times and the the estimation of change in the transmission mechanisms after certain macroeconomic events.
Overall, the interaction structure between financial institutions plays a key role in the analysis of systemic risk, and network approaches can help shedding some light over the analysis of

contagion. A better understanding of the topology of the network of such relationships should then help to evaluate risk and predict the impact of economic shocks. Crucial steps are to identify the channels for the propagation of systemic risk among banks, and to model the mechanisms of diffusion of distress. Given the complexity of financial systems, there are multiple channels for the propagation of systemic risk. Hurd, 2016 suggests the following classification of the main contagion channels:

- **Default contagion**: In case of insolvency of a financial institution, the credit relationships between banks (constituted by traditional interbank lending, but also swaps, derivatives and securitized assets), can represent a contagion channel, and may chain together in a domino effect creating *default cascades*, in particular when the interbank exposure is high compared to the lender's equity.

- **Liquidity contagion**: In a situation of *funding illiquidity*, a bank may try to respond by curtailing its interbank lending, possibly resulting in funding shocks for other banks, triggering a chain that amplifies the distress.

- **Market illiquidity and asset fire sales**: The tendency of distressed financial institutions to liquidate assets in depressed markets can create an *asset fire sales cascade*. This mechanism works in two steps: first the asset sales by a distressed bank decreases the prices, then the marked-to-market losses on others' banks balance sheets lead to losses, and possibly further asset sales.

- **Asset correlation**: Banks often hold common assets in their portfolios. Downward shocks on these assets can increase the banks' leverage and increase the vulnerability of assets' portfolios, making the institutions more vulnerable to other types of contagion.

It is important to underline that the diffusion channels can influence each other and, on top of this, the spread of financial distress can be influenced by several other factors, including for instance the regulatory framework, the behaviour of investors or the possible bail-out of defaulted institutions by the public sector.

Together with the "source-specific approaches" outlined above, Benoit et al., 2017 identify a second family of indicators defined "global measures", not focused on any specific transmission mechanisms, more statistical in nature, and based on the analysis of time series of market instruments. These methods rely on the idea that, if markets are efficient, current prices of securities

can incorporate a vast amont of information. The most common approaches in this class are the ΔCoVaR (Adrian and Brunnermeier, 2016), the systemic risk measure SRISK (Brownlees and Engle, 2016), or network based indicators such as the ones proposed by Billio et al., 2012 or Diebold and Yılmaz, 2014.

The analysis of some network-based global measures will be the focus of Chapters 3 and 4, and we remind to these sections the discussion on "global measures".

Here, we now briefly discuss the main characteristics of "source-specific contagion models", and we present the main findings in the literature. liquidity contagion in a network with a community structure will be the focus of Chapter 5.

Financial contagion models

A common mathematical framework for the analysis of financial contagion is based on the study of *static cascade models* (Hurd, 2016). The general idea is to model banks starting from their balance sheets, and to consider the damaging shocks that can be transmitted over interbank links. The contagion mechanisms can then be modeled, either analytically or by simulations, and the development of the contagion will depend on the nature of the shock, by the balance sheet and bilateral exposure data, and by the behavioural rules that the actors are assumed to follow.

One of the first contribution in the financial literature is the seminal model proposed by Allen and Gale, 2000, that consider a minimal setting of only four banks, where an interbank market arises due to the hetereogeneity of consumers' preferences, and the consequential differences in liquidity needs. The authors then study the effect of liquidity shortage at an aggregated level on the system in different network configurations, showing that the diffusion of distress depends crucially on the network structure, and that the relationship between the density of the bonds in the market and the distress introduced in the system is non-monotonic. Another seminal work is the one proposed by Eisenberg and Noe, 2001, that model default contagion in a banking system by assuming that all banks are part of a single clearing system and introduce an efficient algorithm to obtain iteratively the final state of the system. They also prove the existance and uniqueness of the clearing vector under mild conditions. The Eisenberg and Noe, 2001 model has then be extended, by Gai and Kapadia, 2010, that introduce more realistic balance sheet modelization and contagion mechanism, finding the presence of a *robust-yet-fragile* tendency, that is, while the probability of contagion may be low, the effects can be extremely widespread when problems

occur. In another work, they apply a similar model to liquidity contagion (opposed to default contagion), finding that in period of distress the *liquidity hoarding* phenomenon can decrease the stability of the system. Other authors expanded the literature on static cascade models. We can cite the *generalized liquidity cascade model* introduced by Hurd, 2016, that generalize the Gai, Haldane, and Kapadia, 2011 and the Lee, 2013 models; the Cifuentes, Ferrucci, and Shin, 2005 that includes asset fire sales and the model due to Caccioli et al., 2014 that investigates the case in which contagion is caused by the *fire sales* of assets in presence of overlapping portfolios. An extensive review of the most relevant models, complete with the formal specification is available in Hurd, 2016.

Robust-yet-fragile feature and the topology of network structure

A common theme in the study of financial contagion over networks, is the presence of a *robust-yet-fragile* structure. That is, interconnected banking systems tend to be overall stable and resiliant but, in presence of extreme situations, they tend to be vulnerable to contagion dynamics that amplify the effect of shocks, potentially leading to catastrophic outcomes. Such trade-off comes from the fact that high interconnection can increase the efficiency of the system, allowing banks to obtain liquidity more easily but, on the other hand, it also increases the potential channels of transmission of distress. Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015 study this issue in a model similar to Eisenberg and Noe, 2001, finding a sort of *phase transition* in financial contagion: in presence of shocks of small entity, they find that high level of interconnectivity in the market increases the overall stability and the *complete network* (i.e. a network where all the edges are active) is the most stable configuration. However, when the system faces shocks of high intensity, the complete network becomes one of the least stable configurations. This result is consistent with other works in the literature, including among the others Gai, Haldane, and Kapadia, 2011, Hurd, 2016 and Battiston et al., 2012a. Fragility related to high interconnection is now a major concern also for regulators, that after the crisis acknowledged the role of interconnectivity as a major theme for the stability of the system, see for instance Haldane et al., 2009 and the assessment methodology for the Global Systemically Important Banks (G-SIB) defined by the Basel Committee FSB, 2013 that include the extent of interbank market exposure as one of the criteria to identify systemically important financial institutions.

The resiliance of the system to financial contagion is influenced not only by the density of the

interconnections, but also by their architecture. A relevant stream of literature studies the effects of network configuration on the transmission of contagion. A common stylized fact in banking system is the presence of *hubs*, that are nodes that are much more connected compared to others. The presence of such structure is one of the causes of the *robust-yet-fragile* configuration, since compared to a more homogeneous network, such structure is more resiliant to random failures, but more fragile to failures happening at the nodes (see for instance Barabási and Albert, 1999). Heterogeneity among the banks has been studied also in relation to size (see Iori, Jafarey, and Padilla, 2006), level of connectivity (Amini, Cont, and Minca, 2016) and balance sheet composition Caccioli, Catanach, and Farmer, 2012. The diffusion of contagion can be also influenced by the presence of meso-structure features such as a core-periphery configuration (see Anand, Craig, and Von Peter, 2015) and a community structure, or by behavioural factors such as the presence of imperfect information or bank runs (Battiston et al., 2012b) and moral hazard (Brusco and Castiglionesi, 2007).

Overall, the effect of heterogeneity is complex and non linear, and the identification of contagion patterns has to take into account multiple dimensions. Another source of complexity is given by the fact that the financial system are typically connected through multiple channels, that differ in terms of type of instrument (debt, equity exposures, derivatives, etc) and maturity, while the majority of the studies focus only on the short term interbank lending market (typically overnight). Aldasoro and Alves, 2018 study the multiplex structure of the European banking market, finding a high degree of similarity among these layers, but with some peculiarities that may help to better assess the systemic importance of the banks in the system.

Although some stylized facts have emerged, the literature on financial contagion is still limited, and the effect of several network features on the diffusion of contagion have to be analysed yet. moreover, due to the confidentiality and limited availability of the real-world data, a relevant challenge for regulators and the scientific community is to match the results obtained in theorical and simulation studies to real world settings, in order to develop effective regulatory tools.

Characteristics of the interbank market: evidence from the e-MID platform

The study of the structure of the network requires large and highly disaggregated datasets, in order to build each day the set of all the bilateral exposures among banks. The gathering of such dataset is challenging due to confidentiality, and to the fact that the transactions are often done

over-the-counter. A relevant exception is the e-MID platform, that is the only regulated market for interbank deposits in the Euro Area. The market has been funded in Italy in 1990, and is now used for transactions by banks operating in the European market. A rich dataset, that reports all the transactions settled on the market, is available for research purposes. It includes the univocal code for the two couterparts of each transaction, their nationality, the amount of the loan, its duration, the interest rate and the aggressor. The main limitation of the dataset is the anonimity of the banks, that makes impossible to cross the data with other sources, such as balance sheet data or the market prices of issued securities.

The e-MID dataset has been analysed in several studies, including, among the others, Finger, Fricke, and Lux, 2013, Fricke and Lux, 2015a, Temizsoy, Iori, and Montes-Rojas, 2015, and Anufriev et al., 2016. Next, we present the main findings in the literature, indicating the relevant references.

Concerning the network structure of the market, Fricke and Lux, 2015a show evidence of a rather stable core–periphery configuration. Moreover, they show the presence of a high level of asymmetry between in-coreness and out-coreness and identify the provision of liquidity to the system as the main role of core banks. Fricke and Lux, 2015b analyse the degree distribution of the networks, finding good fit for negative binomial distribution at daily frequency, and Weibull, Gamma, and Exponential distribution for quarterly aggregates. In contrast with some previous studies, they do not find evidence supporting the fit of power-law distributions, casting doubts on the generative mechanism of *preferential attachment* (see Barabási and Albert, 1999) that is often associated to networks in complex systems. Finger, Fricke, and Lux, 2013 show that the network structure is influenced by the time aggregation used for the analysis: at daily level, the network appears to be substantially random, showing little persistence over time, while considering longer time aggregations the links appear more stable, hinting the presence of non-random determinants of link formation based on some preferential relationships. They also find a significant structural break for several network measure in presence of the global financial crisis in 2008. Temizsoy, Iori, and Montes-Rojas, 2015 further analysed the presence of preferential lending channels, showing how they can increase the stability of the system and allow borrowers to obtain more favourable rates. The empirical evidence suggests that such preferential channels have been are particularly relevant during the global financial crisis. Finally, Pecora, Kaltwasser, and Spelta, 2016 found evidence of the presence of a structure characterized by overlapping communities within the core banks in the network.

The literature also shows that the structure of the e-MID market has been subject to significant changes in correspondence to the non conventional monetary policies implemented by the ECB. This is argued by Barucca and Lillo, 2018, that using a stochastic block model to analyse the topological properties of the network, found that after the LTRO (*long term refinancing operation*) in 2012, the stucture becomes more randomized due to the changes in trading behaviour of some banks as a consequence of the flood of cheap liquidity in the market.
Finally, some studies tried to develop agent based models to describe the formation and evolution of the network structure. Anufriev et al., 2016 develop a model in which the behaviour of banks in the market is driven by the trade-off between an uncertain gain on the loaned funds in the interbank market and the need to meet reserve requirements. They found that such framework predicts the emergence of a network structure consistent with the characteristic of the e-MID market, when analysed at daily frequency. This model however does not allow to describe the presence of preferential relationships, that affects the configuration of the market when analysed at longer frequencies. Instead, Iori et al., 2015 propose an alternative agent based model to account for preferential relationships by introducing memory in the trading patterns as a proxy for trust. In such model, the connections active in the past are more likely to be activated, and are reinforced. Such mechanism is regulated by a single memory parameter that can be tuned to market data. The findings are consistent with Finger and Lux, 2017 that, using a stochastic actor oriented model, conclude that in the e-MID market the existence and extent of past credit relationships is a major determinant of credit provision. They also find that after the financial crisis the relevance of core banks increased and the banks tend to limit indirect exposures, probably fearing indirect counterparty risks.

Regulatory framework

The monitoring and controlling of systemic risk and contagion in increasingly relevant for the regulators, that aim to develop effective and coordinate policies. The Basel III framework brings innovations in three main issues: first it aims to improve systemic risk taking by imposing higher capital ratios and with more targeted macroprudential policies such as the use of specific sectorial capital requirements, introducing at the same time countercyclical capital buffers (BCBS, 2011). A second area of intervention is related to liquidity regulation, and aims to reduce maturity mismatches and preventing the need for banks to liquidate their assets generating large

downturns (Basel III, 2013). Finally, a special attention is also dedicated to the so-called *too-big-to-fail* institutions, that is, banks that are so relevant that the entire economy would face severe consequences in case of a default and are potentially characterized by moral hazard risk. In particular, Europe introduced specific requirements in terms of capitalization and transparency for Systemically Important Financial Instititions (SIFIs), with the G-SIB framework in Europe (FSB, 2013), while in the United States the DoddâĂŞFrank Act created the Financial Stability Oversight Council (FSOC), that is in charge of identifying and monitoring systemic risks in the USA, as well as eliminating expectations that American financial companies are too-big-to-fail and will be bailed out.

Global coordination is also a relevant issue, and the institution of the Financial Stability Board (FSB) in 2009 aims to foster greater coordination among supervisors and to identify global systemic weaknesses.

Regulators have still to face several challeges to deal with systemic risk. In particular, one of the main open topics are to effectively develop capital requirements based on the actual network structure of the financial system. Alter, Craig, and Raupach, 2015 propose for instance to define capital requirements based on the centrality of banks in the interbank market network, showing that this approach would reduce bankruptcy costs in case of systemic events. Another issue is related to the use of standardized capital requirement and stress tests methodologies, that on the one hand allow regulators to identify the vulnerabilities of the system, but on the other hand may reduce heterogeneity in risk taking and lead to hearding behaviours (Wagner, 2010).

Finally, as pointed out by Benoit et al., 2017, one of the main challenges is related to the identification of relevant measures of systemic risk, that can be used by regulator to calibrate targeted tools, such as the imposition of systemic risk taxes that, by forcing institutions to internalize the risk, for some authors may be more effective compared to source-specific approaches, that prevent banks to take specific positions.

Network theory and portfolio management

Portfolio management is an increasingly popular field of application of network theory. In recent years, several studies used networks to develop innovative strategies, or to provide new interpretations to traditional methodologies. We can identify three main strains of literature: the first uses networks to support the decision making process in the development of active portfolio

strategies, including stock picking in tilting strategies and technical analysis. Greppi et al., 2018 for instance propose to use bayesian networks to process systematically accounting, market, and sentiment data, in order to identify market signals and develop effective trading rules. Patel et al., 2015 instead use neural networks and other machine learning techniques to forecast stock price movements, and Dash and Dash, 2016 use neural networks to analyse the nonlinear relationships between some popular technical indicators to suggest profitable decisions. All these approach have in common the fact that networks are instrumental to the definition of portfolio decision, and do not represent directly the investable assets. A main issue with such models is that they may result in black-box systems, in which the agent has little or no control and understanding of the investment strategy.

A second strand focuses on the development of investment strategies based on networks in which the investable assets are the nodes, and the edges are defined using statistical relationships among them. These papers use network measures, typically centrality indicators, to define investment strategies. An example is the work of Peralta and Zareei, 2016, that create portfolios by over-weighting assets characterized by a low centrality in correlation networks. The rationale is that by using non-central elements, the resulting portfolio would be better diversified than alternative portfolio strategies. They also found an empirical relationship between the weights of centrality based portfolios and the traditional global minimum variance portfolio, showing that peripheral nodes are typically over-represented in the latter. Such empirical relation has been confirmed by Hüttner, Mai, and Mineo, 2016, that demonstrate how centrality portfolios and minimum variance portfolios are not related algebrically, and that such heuristic relationship is due to specific features of the observed correlation matrix, making it a stylized fact of financial time series. Baitinger and Papenbrock, 2017 propose a centrality-based investment strategy based on mutual information networks capable of accounting for non-linear relationships, and they also found that investment strategies that overweight peripheral assets improve portfolio performances. They also found that such strategies can reduce portfolio turnover, decreasing transaction costs.

Finally, a third strand of literature is also focused on portfolios where the nodes are investable assets. Differently from the second strand, in this case, the network interpretation of the system is functional to the estimation of portfolio optimization parameters, and do not enter the portfolio strategy directly. An example is Goto and Xu, 2015, that consider a partial correlation network interpretation of the covariance matrix, and provide a more stable and reliable estimate of the covariance matrix, used then in the estimation of minimum variance portfolio weights. Another

example is the work of Brownlees and Mesters, 2017, that propose a methodology inspired by the network literature to detect the set of granular time series in large panels, that is, the set of assets that are able to influence the entire cross-section. In Chapter 2 we contribute to this last strand of literature by proposing an extension of the work by Goto and Xu, 2015, in which we consider *tlasso*, a technique for the sparse estimation of partial correlation networks based on the t-Student distribution.

Overall, the literature is still young, and the topic presents several interesting research directions, related in particular to the possibility of merging multiple data sources, including for instance social media data, sentiment analyses, transaction data and financial time series.

The goal of this thesis is to use network theory to address problems in the field of systemic risk in banking systems and portfolio optimization. In particular, in Chapter 2 we use the *tlasso* model to estimate the precision matrix for portfolio optimization, providing an interpretation in terms of partial correlation networks. In Chapter 3 we study the European banking system using partial correlation networks build using Credit Default Swaps (CDS). Also in this case we use the *tlasso* model. In Chapter 4 we consider network-ΔCoVaR, to study the conditional tail risk network between European banks, extending to the multivariate setting the ΔCoVaR model, commonly used for the study of systemic risk. Finally, in Chapter 5 we analyse the diffusion of liquidity contagion in networks with communities, using a static cascade model in a simulation framework.

Chapter 1

Tools

In this chapter we review some of the tools used in the rest of the thesis. The presentation is not supposed to be exhaustive, but instead it is a brief overview, functional to the following chapters, and a collection of references to the relevant literature for the topics.
We first discuss regularization methods, then we present the framework of graphical models and partial correlation networks, discussing the traditional Gaussian case, the t-Student graphical model and quantile graphical models, recently introduced in the statistical literature. We then provide an introduction to quantile regression and penalized quantile regression, and finally we describe some of the main indicators, used to denote the properties of a network.

1.1 Regularization methods

In this thesis, sparsity-inducing regularization methods based on penalized optimization have a relevant role. The idea at the base of this approach is to estimate models with some of the parameters exactly equal to zero, and it can be applied in many different fields and declined in multiple ways. By relying on the assumption that some of the parameters are equal to zero, it is indeed possible to increase the efficiency of the estimators, especially in case of high dimensionality, improving the out-of-sample performances. Moreover they often allow a better intepretability of the model. Typically these benefits come at the cost of increased bias (Tibshirani, 1996).
One of the most common type of regularization is the Least Absolute Shrinkage and Selection Operator (*lasso*), introduced by Tibshirani, 1996 that uses an L_1 penalty in the estimation of the parameters of a linear model.
Let $\boldsymbol{\beta}$ be the $(p \times 1)$ vector of parameters of a linear model, $\boldsymbol{y}$ a $(t \times 1)$ vector of observations of a response variable, $\mathbf{x}$ a $(t \times p)$ matrix of observations of predictor variables, k and λ two

coefficients that act as tuning parameters, the *lasso* estimator is obtained as:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta}) \tag{1.1}$$

$$s.t. \qquad ||\boldsymbol{\beta}||_1 < k, \tag{1.2}$$

or equivalently:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta}) + \lambda ||\boldsymbol{\beta}||_1, \tag{1.3}$$

where $||\boldsymbol{\beta}||_1 = \sum_{i=1}^{p} |\beta_i|$ is the L_1 norm of the vector $\boldsymbol{\beta}$. The estimation can be performed fast and reliably, and lasso estimator for linear models is implemented in many of the most common statistical software packages and programming environments. The sparsity of the model is controlled by the parameter λ that can be calibrated using cross validation, or information criteria such as the Bayesian Information Criterion (BIC).

Lasso is not the only technique that allows to obtain a sparse model. Considering a generic penalty function $p(\boldsymbol{\beta}_i)$, where $\boldsymbol{\beta}_i$ is the i-th element of the vector $\boldsymbol{\beta}$, we can write penalized estimators as:

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta}) + \sum_i p(\boldsymbol{\beta}_i). \tag{1.4}$$

Fan and Li, 2001 describe the class of folded concave penalized estimators, that allow to obtain a sparse estimator, while maintaining better asymptotic properties compared to *lasso* Fan, Xue, and Zou, 2014. One of the most relevant is the Smoothly Clipped Absolute Deviation (SCAD) penalty, defined as follows:

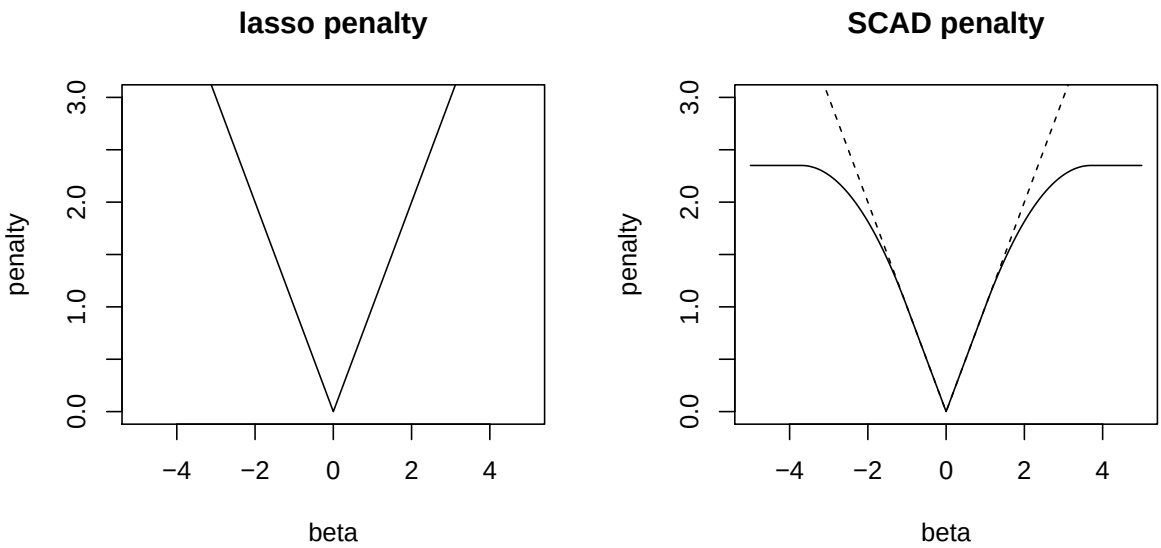
$$p_{\lambda,a}^{\text{SCAD}}(\boldsymbol{\beta}_i) = \begin{cases} \lambda|\boldsymbol{\beta}_i| & \text{if } |\boldsymbol{\beta}_i| \leq \lambda \\ -\dfrac{|\boldsymbol{\beta}_i|^2 - 2a\lambda|\boldsymbol{\beta}_i| + \lambda^2}{2(a-1)} & \text{if } \lambda < |\boldsymbol{\beta}_i| \leq a\lambda \\ \dfrac{(a+1)\lambda^2}{2} & \text{if } |\boldsymbol{\beta}_i| > a\lambda \end{cases}, \tag{1.5}$$

where λ and a are positive constants. The function is linear and equivalent to a lasso penalization near the origin, then quadratic for a trait and finally flat (see figure 1.1).

The non-convexity of the penalization makes the optimization problem much harder to solve, however several specific algorithms have been developed. In the penalized mean regression framework, iterative algorithms based on local linear approximation of the penalty function have

been proposed, allowing for relatively fast estimation (see Zou and Li, 2008; Fan, Xue, and Zou, 2014). Moreover, Fan, Xue, and Zou, 2014 showed that if such algorithm is initialized at a lasso optimum satisfying certain properties, the two-stage procedure produces an oracle solution, and Loh and Wainwright, 2013 show that under suitable conditions, any stationary point of the objective function will lie within the statistical precision of the underlying parameter vector and propose specific algorithms to obtain near-global optimum solutions.

FIGURE 1.1: lasso (left) and SCAD penalty (right)



1.1.1 Penalized quantile regression

Although initially developed for the estimation of mean regression models, *lasso* is quite general, and can be applied to several statistical models, including quantile regression, as described in Koenker, 2005.

Quantile regression is a statistical tool that allows to study the behaviour of the τ conditional quantiles of response variable given the value of predictor variables, in contrast to the more common mean regression, that studies the behaviour of the conditional mean. The basic idea dates back to the work of Boscovich in the 18th Century, but the methodology gained popularity only in the second half of the 20th Century with the diffusion of computers.

It is possible to express quantile regression as an optimization problem by considering the following loss function: $\rho_\tau(u) = (1-\tau)u\mathbb{I}_{u\leq 0} + \tau u\mathbb{I}_{u>0} = (\tau - \mathbb{I}_{u\leq 0})u$, where $\tau \in (0,1)$ and $\mathbb{I}_{\{\}}$

is an indicator function. Given a response variable $\boldsymbol{y}$ and a predictor variable $\mathbf{x}$, the optimization problem is then:

$$\widehat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}_\tau} \mathbb{E}[\rho_\tau(\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta}_\tau)]. \tag{1.6}$$

The problem can be expressed as a linear program, and solved efficiently using the simplex method or interior point approaches (see Koenker, 2005, Chapters 1 and 6).

The introduction of a lasso penalty in the optimization problem (1.6) is quite natural, and allows to maintain the linear structure of the problem (see Koenker, 2005, Chapter 4.9.2 and Belloni, Chernozhukov, et al., 2011). In particular, the optimization problem for lasso penalized quantile regression is:

$$\widehat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}_\tau} \mathbb{E}[\rho_\tau(\boldsymbol{y} - \mathbf{x}\boldsymbol{\beta}_\tau)] + \lambda||\boldsymbol{\beta}_\tau||_1, \tag{1.7}$$

where λ is a coefficient that controls the amount of penalization and $||\cdot||_1$ is the L_1-norm.

An alternative penalization is the Smoothly Clipped Absolute Deviation (SCAD) introduced by Fan and Li, 2001. In the context of penalized linear regression, the non-convexity does not allow a linear programming representation of the problem but, following Wu and Liu, 2009, it is possible to address the problem using a Difference Convex Algorithm (DCA) that uses the representation of SCAD penalty as the difference between a linear and a convex function, and solves a sequence of convex problems to approximate the SCAD problem efficiently.

1.2 Graphical models and partial correlation networks

We focus now on the concept of *partial correlation networks*, in which the system is modelled as a multivariate distribution, the nodes are the individual variables and the edges are the partial correlation between them. In particular, under the Gaussian case we can talk about Gaussian graphical models.

Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(m)})$ be a random vector with a multivariate Gaussian distribution $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix. We can define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in $\mathcal{V}$ correspond to each element of $\mathbf{X}$, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations, $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, and the edge weights correspond to partial correlations ρ_{ij}. The partial correlations are computed

from the inverse of the covariance matrix $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$ (i.e., the precision matrix) as follows:

$$\rho_{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \qquad i, j = 1, \ldots, m,$$

where $\{\omega_{ij}\}$ is an element of the matrix $\boldsymbol{\Omega}$. In matrix form the partial correlation matrix $\boldsymbol{P}$ is:

$$\boldsymbol{P} = -\boldsymbol{D}\boldsymbol{\Omega}\boldsymbol{D}, \tag{1.8}$$

where $\boldsymbol{D} = diag(\frac{1}{\sqrt{\omega_{ii}}})$.

The estimation of the graph structure corresponds then to the estimation of the precision matrix $\boldsymbol{\Omega}$, and its sparse estimation, that is, estimation with most off-diagonal entries of the matrix exactly equal to zero, offers advantages in terms of interpretability and estimation error. A common approach to introduce sparsity is to penalize the maximum likelihood estimation problem by an L_1-norm. In the Gaussian case we can state the graphical lasso (*glasso*) estimation problem as:

$$\widehat{\boldsymbol{\Omega}}_{glasso} = \arg\max_{\boldsymbol{\Omega}} \big(log|\boldsymbol{\Omega}| - tr(\boldsymbol{\Omega}\boldsymbol{S}) - \rho||\boldsymbol{\Omega}^{-}||_1\big), \tag{1.9}$$

where $\boldsymbol{S}$ is the sample covariance matrix, ρ is a tuning parameter that controls the sparsity (i.e. the larger ρ, the larger the number of elements of the precision matrix set equal to zero), $|\cdot|$ is the determinant, $tr(\cdot)$ the trace, $||\cdot||_1$ the L_1-norm of a matrix and $\boldsymbol{\Omega}^{-}$ is a square $m \times m$ matrix with the off-diagonal elements equal to the corresponding elements of $\boldsymbol{\Omega}$ and the diagonal elements equal to zero (Friedman, Hastie, and Tibshirani, 2008).[1]

As pointed out by Yuan and Lin, 2007, the use of an L_1 penalty allows to induce sparsity in the precision matrix, making it possible to perform model selection and parameter estimation simultaneously. The choice of the tuning parameter ρ, that controls the level of sparsity of the estimates assumes therefore a major role in the estimation.

The statistical properties of the *glasso* estimator have been studied, among others, by Banerjee, Ghaoui, and d'Aspremont, 2008, Rothman et al., 2008 and Lam and Fan, 2009. Rothman et al., 2008 show that under some regularity conditions and for a suitable choice of the parameter ρ, the estimator in (1.9) has a rate of convergence to the true parameter $\boldsymbol{\Omega}$ in the Frobenius norm of order $\sqrt{((m + k_{\boldsymbol{\Omega}})\log(m)/t)}$, where $k_{\boldsymbol{\Omega}}$ is the number of non-zero off diagonal entries of the true

[1]The original specification proposed by Friedman, Hastie, and Tibshirani, 2008 applied the penalty to the entire matrix $\boldsymbol{\Omega}$. The version of the model with the penalty applied to $\boldsymbol{\Omega}$ is the one studied by Rothman et al., 2008 and is currently implemented in the R package *'glasso'* (Friedman, Hastie, and Tibshirani, 2014).

matrix $\mathbf{\Omega}$ and t the number of observations. The main implication is that the convergence is faster for matrices that are truly sparse. Lam and Fan, 2009 studied the consistency and *sparsistency* of the estimator (i.e. the property that all parameters that are zero are actually estimated as zero with probability tending to one), showing that sparsistency requires a lower bound on the rate of the regularization parameter ρ, while an upper bound is necessary to control the bias introduced by the L_1 penalty and to obtain a consistent estimator.[2] Under some technical conditions, the compatibility of these bounds requires the number of off-diagonal non-zero entries of the true precision matrix to be no larger than $O(m)$ (Lam and Fan, 2009). In practical terms, the true precision matrix has to be sparse enough, and the *glasso* estimates of dense precision matrices will not be consistent. On the other hand, the modelization of sparse precision matrices has the advantage of reducing the variability of the estimates.

The optimization problem (1.9) is convex, as proved by Banerjee, Ghaoui, and d'Aspremont, 2008. Friedman, Hastie, and Tibshirani, 2008 proposed an efficient algorithm to solve it that has a computational complexity of $O(m^3)$ for dense problems, and considerably less than that for sparse problems. The algorithm is described in Appendix A.

An advantage of the *glasso* model is that the solution $\widehat{\mathbf{\Omega}}_{glasso}$ is always unique and has bounded eigenvalues,[3] also when the number of covariates m is smaller than the number of observations t. Moreover, for the case in which $t > m$, empirical analyses show that the estimate is generally better conditioned compared to the sample covariance matrix (see e.g. Goto and Xu, 2015). Finally, we point out that the sparsity of the precision matrix does not necessariliy correspond to the sparsity of the covariance matrix.

Other than increasing the efficiency of the estimator, *glasso* allows to improve the interpretability of the model, by identifying only the most relevant items. In the estimation of a financial network for instance, we can identify the most relevant edges, similarly to other techniques such as minimum spanning trees (Mantegna, 1999) or planar maximally filtered graphs (Tumminello et al., 2005), but without imposing any topological structure on the resulting network.

1.2.1 The *tlasso* model

The *tlasso* model was recently introduced in the statistical literature to estimate sparse partial correlation matrices under the assumption that the data follow a multivariate t-Student distribution

[2] See Theorem 2 and Technical Condition (B) in Lam and Fan, 2009.

[3] see Banerjee, Ghaoui, and d'Aspremont, 2008, Theorem 1.

(Finegold and Drton, 2011). Compared to *glasso*, the distributional assumptions of this model are less restrictive for financial applications, where time series are typically leptokurtic (see for instance Cont, 2001). Furthermore, *tlasso* is robust to model misspecification and estimation errors due to outliers (Finegold and Drton, 2011).

Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(m)})$ be an m-variate random vector with a multivariate t-Student distribution $t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v)$, with v degrees of freedom, mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Psi}^{-1}$ ($m \times m$ positive semi-definite matrix[4]). The covariance and precision matrices are then $\boldsymbol{\Sigma} = \frac{v}{v-2}\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\Omega} = \frac{v-2}{v}\boldsymbol{\Psi}$, respectively.

Similarly to the Gaussian case, we can associate to the distribution a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, in which $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, where $\{\rho_{ij}\}$ are the elements of the partial correlation matrix $\boldsymbol{P}$ computed from the precision matrix $\boldsymbol{\Omega}$ by standardization. *Tlasso* provides a sparse estimate $\widehat{\boldsymbol{\Psi}}_{tlasso}$ and the corresponding partial correlation matrix $\widehat{\boldsymbol{P}}$ that represents the network structure of the model.

In contrast to the Gaussian set-up, when using the t-Student distribution, or the class of elliptical distributions in general, an absence of correlation does not necessarily correspond to conditional independence (Baba, Shibata, and Sibuya, 2004). However, despite the lack of conditional independence for $\omega_{ij} = 0$, Finegold and Drton, 2011 proved that if two nodes j and k are separated by a set of nodes C in $\mathcal{G}$, then $X^{(j)}$ and $X^{(k)}$ are conditionally uncorrelated given $X^{(C)}$. It is then reasonable to substitute the conditional independence with zero partial correlation or zero conditional correlation. In this case, disconnected vertices in a graphical model can be considered orthogonal to each other after the effects of other variables are removed. The absence of conditional correlations entails that a mean-square error optimal prediction of variable $X^{(j)}$ can be based on the variables $X^{(k)}$, which correspond to neighbours of the node j in the graph.

The *tlasso* model can be estimated using an Expectation-Maximization (EM) algorithm (Finegold and Drton, 2011). The EM estimation procedure exploits the scale-mixture representation of the t-Student distribution, consisting of a multivariate Gaussian and a gamma distribution (Kotz and Nadarajah, 2004). The procedure is computationally efficient as it is based on *glasso*, which is called at every M-step of the algorithm. A description of the algorithm can be found in Appendix A.

[4]We chose the uncommon parametrization based on the inverse of the diffusion matrix to highlight the relevance of Ψ in the graphical model.

1.2.2 Quantile graphical models

Conditional quantiles provide an alternative approach for the construction of a graphical model: Quantile Graphical Models (QGM), recently introduced in the statistical literature to describe the dependency structure of multivariate random variables.[5]

Several specifications have been discussed in the literature, in particular we cite Belloni, Chen, and Chernozhukov, 2016; Ali, Kolter, and Tibshirani, 2016; Chun et al., 2016. In general, we can express the quantiles of the distribution of an item conditional to the others as

$$Q_{\mathbf{X}_i,\tau} = f(\mathbf{X}_{\setminus i}, \tau), \tag{1.10}$$

where the function f is a generic function. The identification and estimation of the function $f(\mathbf{X}_{\setminus i}, \tau)$ is a crucial step, and several approaches can be chosen. We may determine the function in a fully non-parametric form but, due to course of dimensionality, some simplifying assumptions have to be made in order to allow estimation in finite samples. A common choice is to define an additive form such as $f(\mathbf{X}_{\setminus i}, \tau) = \sum_{j \neq i} f_j(\mathbf{X}_j, \tau)$, where $f_j(\mathbf{X}_j, \tau)$ are smooth functions (see e.g. Ali, Kolter, and Tibshirani, 2016). A further simplification is to assume that $f_j(\mathbf{X}_j, \tau)$ are linear functions (Belloni, Chen, and Chernozhukov, 2016); in such case the function f can be expressed in terms of linear quantile regression coefficients: $f(\mathbf{X}_{\setminus i}, \tau) = \alpha_i + \boldsymbol{\beta}_\tau^{i|\cdot} \mathbf{X}_{\setminus i}$, where α_i is a constant, and $\boldsymbol{\beta}_\tau^{i|\cdot}$ is a $1 \times p - 1$ vector. Such coefficients can be conveniently represented by the matrix $\boldsymbol{B}_\tau$ such that:

$$\{\boldsymbol{B}_\tau\}_{ij} = \begin{cases} \beta_\tau^{i|j} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}. \tag{1.11}$$

In this way, under the assumption of a linear specification, we can express the conditional quantile as:

$$f(\mathbf{X}_{\setminus i}, \tau) = \alpha_i + \{\boldsymbol{B}_\tau \mathbf{X}\}_i. \tag{1.12}$$

QGM allow a richer modelization of the interdependence structure among nodes in a network as they consider a larger amount of information. In particular, they allow to focus not only on the

[5]Note that a QGM in the Gaussian case is a simple rescaling of partial correlation networks. The topic is discussed in greater detail in Section 4.3.2.

conditional means, but also on the entire conditional distributions, especially the tails, that are particularly relevant in financial applications.

1.3 Network Indicators

Network theory allows to model complex systems parsimoniously and to explain the behaviour of the systems on the basis of their topological properties. In order to describe this relationship quantitatively, it is necessary to compute suitable indicators to describe the properties of the network.

Such indicators are typically categorized into *local* and *global*, where local indicators refer to characteristics of individual nodes and global indicators refer to properties of the network as a whole.

Focusing on undirected weighted networks (i.e., networks where edges do not have a direction but have a weight), we present here some of the most common network indicators. In the following Sections we use the convenient representation of a network in terms of its adjacency matrix $\boldsymbol{A}$, that is, the square matrix in which columns and rows represent the nodes, and the entries represent the weights of the edges connecting all the couples of nodes. Notice that, in case of undirected networks, the adjacency matrix $\boldsymbol{A}$ is symmetric.

1.3.1 Centrality measures

Centrality measures are indicators used to quantify the importance of individual nodes in the network. We present here three common measures: *strength centrality*, *eigenvector centrality* and *Bonacich power centrality*.

The *strength centrality* of node i is defined as:

$$c_i^S = \sum_{j=1}^{m} w_{ij}, \tag{1.13}$$

where c_i^S is the *strength centrality*, or *strength* of node i, m is the total number of nodes and w_{ij} is the weight of edge ij. This measure can be considered an extension of *degree centrality* to weighted networks (Barrat et al., 2004).

The *eigenvector centrality* is a measure that assigns to each node a score based on the principle that connections to important nodes contribute more to centrality than connections with less central nodes, showing a recursive nature. Formally, the ith node's centrality score is proportional to the sum of the scores of all the nodes which are connected to it:

$$c_i^E = \frac{1}{\delta} \sum_{j=1}^{m} a_{ij} c_j^E, \tag{1.14}$$

where c_j^E is the *eigenvector centrality* of node j, a_{ij} is the ijth element of the adjacency matrix $\boldsymbol{A}$ and δ is a constant. Under suitable conditions, *eigenvector centrality* can be computed as the eigenvector corresponding to the largest eigenvalue of the adjacency matrix $\boldsymbol{A}$. *Eigenvector centrality* can be generalized to weighted networks using the weighted adjacency matrix $\boldsymbol{W}$ instead of $\boldsymbol{A}$.

Bonacich power centrality is another recursive centrality measure, defined as:

$$c_i^B = \sum_{j=1}^{m} (\alpha + \beta c_j^B) a_{ij}, \tag{1.15}$$

where c_i^B is the *Bonacich power centrality* of node i, α and β are two constants, and a_{ij} is ijth element of the adjacency matrix $\boldsymbol{A}$ (Bonacich, 1987). Similarly to *eigenvector centrality*, a node's centrality is influenced by the neighbours' centralities. For convenience, we consider the case in which $\alpha = \beta = 1$.

1.3.2 Freeman centralization

Strength, *eigenvector* and *Bonacich* centralities are local measures that apply to individual nodes. It is also possible to consider global measures that characterize the centralization of the entire network computing the *Freeman centralization* measure as follows (Freeman, 1978):

$$C = \frac{\sum_{i=1}^{m} c_* - c_i}{\max\left(\sum_{i=1}^{m} c_* - c_i\right)}, \tag{1.16}$$

where c_i is a local centrality measure for the node i, c_* is the centrality corresponding to the most central node in the network and $\max \sum_{i=1}^{m} (c_* - c_i)$ is a normalization factor computed on

the network with the highest centralization index for a given number of nodes.[6] In other words, *Freeman centralization* calculates the difference between the centrality of the most central node and all the others, Therefore, a network with a high centralization has a node, or a group of nodes, that has a high centrality and a large number peripheral nodes.

1.3.3 Clustering coefficient

The *clustering coefficient* measures the tendency of the network to create triangles. In a network with communities it is likely that, for a given node, two of its neighbouring nodes are connected to each other, "closing the triangle".
We consider the *clustering coefficient* for weighted networks proposed by Fagiolo, 2007, which is defined as:

$$c_i^w = \frac{(\boldsymbol{W}^{[1/3]})_{ii}^3}{k_i(k_i - 1)}, \tag{1.17}$$

where k_i is the degree of node i and $\boldsymbol{W}^{[1/3]} = \{w_{i,j}^{1/3}\}$, that is, the matrix obtained by the weight matrix W by taking the third root of each entry. The measure takes values in [0,1] and is equal to the non weighted *clustering coefficient* when the weights become binary. Fagiolo's *clustering coefficient* c_i^w is a local measure that applies to each individual node i. Still, it is possible to derive a global clustering measure C^w computed as the average of c_i^w over all the nodes.

1.3.4 Modularity

Modularity is a measure that allows us to measure how well a certain partition describes the network. A high level of *modularity* indicates that the number of edges internal to each partition is higher than expected. Given a partition $G = \{G_1, \ldots, G_p\}$ we can define modularity Q as follows:

$$Q = \frac{1}{2M} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(w_{ij} - \frac{s_i s_j}{2M} \right) \delta(g_i, g_j), \tag{1.18}$$

where w_{ij} is an element of the weighted adjacency matrix, s_i is the strength of node i, $M = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij}$, g_i is the group in the partition in which the element i belongs and $\delta(g_i, g_j)$ is 1 if $g_i = g_j$ and 0 otherwise. *Modularity* can be computed on any partition of the network, and

[6] For several centrality measures the most centralized network is star-shaped, while for the eigenvector centrality it is a disconnected network with a single edge connecting two nodes and a large number of disconnected nodes.

can assume values between -1 and 1, with positive and high values denoting a good division of the network into communities.

It is possible to identify the best partition of a network by an optimization procedure that maximise modularity over the space of all the possible partitions in a network. The procedure proposed in Newman, 2004 solves this problem by using a greedy optimization that, starting with each vertex being the unique member of a community, repeatedly joins together the two communities whose amalgamation produces the largest increase in modularity. This approach can be implemented efficiently on large networks and identifies automatically the optimal number of communities.

1.3.5 Assortativity

A useful indicator to describe the structure of a network is *assortativity*, defined by as the Pearson correlation coefficient for a nodes' characteristic computed over the edges of the network. This indicator measures the extent to which nodes with similar properties are connected to each other and can be computed for virtually any property assigned to the nodes, including categorical ones (Newman, 2010). Assortativity is typically used with the node *degree* or *strength*, but can be computed on any scalar characteristic of the nodes.

In a weighted network, the expression for *assortativity* with respect to a scalar characteristic is the following:

$$r = \frac{\sum_{k=1}^{m} w_k (x_k - \mu_x)(y_k - \mu_y)}{\sqrt{\sum_{k=1}^{m} w_k (x_k - \mu_x)^2 \sum_k w_k (y_k - \mu_y)^2}}, \tag{1.19}$$

where k is the index corresponding to each edge, w_k is the weight of edge k, x_k and y_k are scalar characteristics of the originating and destination nodes (the scalar characteristic in our case is *eigenvector centrality*) and μ_x and μ_y are the respective average values weigthed by the edges' weights. In other words, the *assortativity* is the weighted correlation coefficient over the edges of each network and thus lies in the range [-1,1] with $r = 1$ indicating perfect *assortativity* and $r = -1$ indicating perfect *disassortativity*. Note that *assortativity* is not defined for negative weights, therefore in the computation of the indicator we set the weight for negative edges to zero.

An assortative network can be consistent with a core-periphery structure, where we observe a core of highly central nodes surrounded by a less dense periphery of nodes with lower centrality.

Newman and Park, 2003 show that a high level of *assortativity* can be found also in network characterized by a the presence of communities, typically social networks.

1.3.6 DebtRank

DebtRank is a methodology introduced by Battiston et al., 2012a that allows to evaluate the systemic relevance of banks in a system and to conduct simple stress test analyses. Unlike the network indicators and methodologies presented so far, DebtRank has the peculiarity of being developed explicitly for the study of banking systems. DebtRank allows to estimate the systemic impact of a shock to a node in the network. Here, we focus on a small shock hitting individual banks and we want to measure the final effect, due to the shock reverberations through the network.

We define the impact of node i on node j as $\alpha \cdot b_{i,j}$, where $b_{i,j}$ is the weight of the link in the network and α is a parameter that controls the intensity of the impact. The impact of i on its first neighbours is $\sum_j b_{i,j} v_j$, where v_j is a measure of the economic size of j.

We associate to each node two state variables: $h_i \in [0,1]$ (continuous variable) and $s_i \in \{U, D, I\}$ (discrete variable with 3 possible states: undistressed, distressed, inactive).

Denoting by S_f the set of nodes in distress at time 1, the initial conditions are: $h_i(1) = \Psi \quad \forall i \in S_f; \quad h_i(1) = 0 \quad \forall i \notin S_f$, and $s_i(1) = D, \quad \forall i \in S_f; \quad s_i(1) = U \quad \forall i \notin S_f$. The parameter Ψ measures the initial level of distress: $\Psi \in [0,1]$, with $\Psi = 1$ meaning default.

The dynamics is defined as follows:

$$h_i(t) = \min\left\{1, h_i(t-1) + \sum_j (b_{ji})^+ h_j(t-1)\right\}, \quad \text{where} \quad j|s_j(t-1) = D, \tag{1.20}$$

$$s_i(t) = \begin{cases} D & \text{if } h_i(t) > 0, s_i(t-1) \neq I, s_i(t-1) \neq D \\ I & \text{if } s_i(t-1) = D \\ s_i(t-1) & \text{otherwise} \end{cases}, \qquad \forall i. \tag{1.21}$$

b_{ji}^+ is defined as $\min\{0, b_{ji}\}$ and denotes the positive edges in the network. After a finite number of steps T the dynamics stop and all the nodes in the network are either in state U or I.

The DebtRank of the set S_f is then defined as

$$R = \sum_j h_j(T)v_j - \sum_j h_j(1)v_j, \tag{1.22}$$

and it can be interpreted as the amount of distress induced in the system by the reverberation of the initial shock. Conceptually, DebtRank can be considered a sort of feedback centrality measures such as *eigenvector centrality* or Pagerank, since they can also be computed using iterative algorithm (see Page et al., 1999 and Newman, 2010).

Chapter 2

Sparse precision matrices for minimum variance portfolios

Acknowledgement

This chapter is largely based on the article "Sparse Precision Matrices for Minimum Variance Portfolios" co-authored by Sandra Paterlini and Rosella Giacometti, published in the journal *Computational Management Science* (Torri, Giacometti, and Paterlini, 2019).

2.1 Introduction

Markowitz's mean-variance model (Markowitz, 1952) represents a cornerstone for asset allocation frameworks and financial theory in general (see for example Kolm, Tütüncü, and Fabozzi, 2014 and references therein). Since then, many alternative methods and new strands of research have been established, fostered by the FinTech industry, which relies on the development of data-driven and automatic investment tools. Markowitz's simplistic Gaussian framework, the idea of risk-return optimization and diversification are still the starting points for the largest majority of more sophisticated recent approaches. Possibly, one of the main challenge of Markowitz's mean-variance model is the need to provide reliable estimates of the input parameters: the expected asset returns vector and the expected covariance matrix. Concerning the expected returns, it is acknowledged that they are extremely difficult to estimate and are often the main source for unreliable allocations (Michaud, 1989; Brodie et al., 2009), with suboptimal Sharpe ratios compared to the minimum variance portfolios (Black and Litterman, 1992). Most research has recently focused on building robust and reliable estimators for the covariance matrix to better

control for the estimation error, especially to avoid its impact on asset weights estimates. In the minimum variance setting, the problem with the estimation of the covariance matrix is amplified by the fact that the input required for the analytical solution is its inverse, the precision matrix (Stevens, 1998). Hence, the optimal minimum variance portfolio strongly depends on the largest eigenvalues of the precision matrix, which are the smallest eigenvalues of the covariance matrix and are typically dominated by noise. As widely discussed in the literature, sample estimates often result in ill conditioned covariance matrices when the number of asset is relatively large compared to the depth of the time series (see e.g. Ledoit and Wolf, 2004a; Meucci, 2009; Won et al., 2013). This, in turn leads to optimal portfolios with extreme and unstable positions over time (see for instance Michaud, 1989 and Ledoit and Wolf, 2004b). Furthermore, the presence of positive multicollinearity among asset returns impacts even further the weight estimates by resulting in unrealistic short positions, which have to be offset by corresponding long positions. This is exacerbated when the number of assets is large compared to the number of observations available for the estimates. High volatilities, positive multicollinearity and the presence of extremes are typical during financial crises. Some works have focused on GARCH-based approaches to better capture time-varying volatility (see Engle, 2002). Here, instead, we rely on a rolling-window mechanism to update the input estimates for the minimum variance portfolio. In the aftermath of the 2008 crisis, the academic literature has seen a surge of contributions to improve covariance estimation, such as the Ledoit-Wolf shrinkage estimator (Ledoit and Wolf, 2004b) and random matrix theory (Laloux et al., 1999). Other strands of research instead focus on robust optimization methods (De Miguel and Nogales, 2009) and, most recently, on regularization methods such as lasso (Tibshirani, 1996), which relies on imposing a penalty function on the L_1-norm of the asset weight vector (De Miguel et al., 2009; Brodie et al., 2009; Fan, Zhang, and Yu, 2012). We refer the reader to Kremer, Talmaciu, and Paterlini, 2017 for a comparison of state-of-art techniques within a minimum risk framework.

Here, we introduce two approaches that rely on constraining the L_1-norm of the precision matrix to reduce the estimation error impact on optimal portfolio weights. The proposed methods are based on Markovian graphs to improve the stability of the precision matrix estimates. In particular, we consider two statistical set-ups: the first one assumes that asset returns are normally distributed, hereafter *glasso* (Friedman, Hastie, and Tibshirani, 2008), while the second relies on the assumption of t-Student asset returns, which better fits assets returns that, as it is widely known, are leptokurtic and typically characterized by fat tails, hereafter *tlasso* (Finegold

and Drton, 2011). To our knowledge, within the financial literature the *glasso* approach has been discussed in the context of portfolio optimization by Goto and Xu, 2015 and Brownlees, Nualart, and Sun, 2018, while the *tlasso* has not received any attention yet. Here, we aim to fill some gaps in the literature by showing that *glasso* and *tlasso* are effective tools for the development of data-driven investment strategies. In particular, the main goal of the analysis is to investigate the out-of-sample performances of *glasso* and *tlasso* in an equity portfolio set-up. We focus on the minimum variance framework, since it has an analytical solution that depends on the precision matrix, allowing us to point out the effect of estimation error in the inputs. Moreover, aligned to many studies in the literature (e.g., De Miguel et al., 2009; Fan, Zhang, and Yu, 2012), we focus on the global minimum variance case, given that estimates of expected returns are typically unreliable (Michaud, 1989).

We underline that our focus is on the estimation of the covariance and precision matrix, not on the development of benchmark-beating equity investment strategies. As a consequence, in the empirical part of the work we focus on the comparison of different estimators strategies for the input parameters maintaining the focus on the global minimum variance portfolios for a more meaningful comparison. Still, asset managers relying on quantitative modelling might benefit from our findings when setting up more complex investment strategies thanks to a better estimation of the interconnectivity structure.

The chapter is structured as follows. Section 2 introduces the minimum variance portfolio framework and describes the *glasso* and *tlasso* approaches. Section 3 discusses the simulation set-up and the main results. Section 4 provides empirical results on real-world financial data, when compared with state-of-art methods. Section 5 then concludes.

2.2 Methodology

2.2.1 The minimum variance portfolio

Since the seminal work of Markowitz (Markowitz, 1952), the idea of risk minimization by diversification has become central to modern portfolio theory. Markowitz minimum variance portfolio framework is still considered the reference model for many scholars and practitioners. The simplicity of the risk diversification idea, quantified by linear dependence, resulting in the need to estimate “just” the covariance matrix, and the possibility to rely on an analytical solution, or to

deal with a "simple" quadratic optimization problem, have been possibly among the main factors behind the widespread use of the minimum variance model, as stated below:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ s.t. \quad & \mathbf{1}'\boldsymbol{w} = 1, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\Sigma}$ is the $n \times n$ true covariance matrix, $\boldsymbol{w}$ the $n \times 1$ vector of asset weights and $\mathbf{1}$ a $n \times 1$ unit vector. The optimization problem has then an analytical solution:

$$\mathbf{w}_{MV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{2.2}$$

where $\mathbf{w}_{MV}$ is the vector of weights of the optimal minimum variance portfolio.
As $\boldsymbol{\Sigma}$ is unknown, an estimate $\widehat{\boldsymbol{\Sigma}}$ has to be computed to obtain the weights:

$$\hat{\mathbf{w}}_{MV} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}. \tag{2.3}$$

Markowitz's minimum variance portfolios have been widely critized, mainly due to the sensitivity to estimation error in the input estimates, resulting in unrealistic extreme weights and bad out-of-sample properties.[1] One of the main critiques is related to the use of the sample covariance estimates, and subsequently its inverse, as an input. Indeed such estimator has a slow convergence rate to the true covariance matrix. As a consequence, finite sample estimation is characterized by a higher eigenvalue dispersion compared to the true covariance matrix and by a low accuracy of the eigenvectors corresponding to the smallest eigenvalues, especially for matrices of large dimension estimated on short time series (Meucci, 2009). Another shortcoming is related to the role of the inverse of the covariance matrix $\boldsymbol{\Omega} \equiv \boldsymbol{\Sigma}^{-1}$, the so called *precision matrix*, in the solution of the optimization problem. From (2.2), it is clear that the accuracy of the estimation of the weights is directly related to the accuracy of the precision matrix. Using spectral decomposition, the relationship between the two matrices can be explicitly studied. In fact, the eigenvector decomposition of the covariance matrix can be expressed as $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\boldsymbol{V}$ is the matrix of eigevectors with $\boldsymbol{V}^{-1} = \boldsymbol{V}'$ and $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$

[1] Other critiques of the Markowitz framework focus on the limits of variance as a risk measure, proposing alternative approaches to quantify risk, yielding to different optimal asset allocation strategies (Artzner et al., 1999; Rockafellar and Uryasev, 2013; Giacometti, Torri, and Paterlini, 2020). We do not discuss these issue here, as we focus on the numerical issues in the estimation process of the covariance matrix.

with $\lambda_1 \geq \cdots \geq \lambda_n$ is the diagonal matrix of the eigenvalues sorted in decreasing order on the main diagonal. Analogously, the eigendecomposition of the precision matrix is such that $\boldsymbol{\Omega} = \boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{U}'$,where $\boldsymbol{\Delta} = diag(\delta_1, \ldots, \delta_n)$ with $\delta_1 \leq \cdots \leq \delta_n$ (note that in this case the eigenvalues are sorted in ascending order). By inverting the covariance matrix, we have

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= (\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}')^{-1} \\ &= (\boldsymbol{V}')^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'. \end{aligned} \tag{2.4}$$

It follows that $\boldsymbol{U} = \boldsymbol{V}$ and $\boldsymbol{\Delta} = \boldsymbol{\Lambda}^{-1}$ with the ith element $\delta_i = 1/\lambda_i$, that is, the eigenvectors of the precision matrix are the same as those of the covariance matrix and the eigenvalues are the reciprocal of those of the covariance matrix.
The consequence is that the dominant eigenvectors of the precision matrix (i.e., the ones corresponding to the largest eigenvalues) are the ones most likely influenced by noise and estimation error, especially in ill-conditioned covariance matrices.

2.2.2 Graphical lasso or *glasso*

Graphical models can be useful to describe both the conditional and unconditional dependence structure of a set of variables. Gaussian Graphical Models (GGMs) are probably the most popular tools to capture the network structure of a set of variables. As Markowitz's model relies on the normality assumption of the asset returns, GGMs are the natural choice for capturing and estimating linear dependence among assets (see Dempster, 1972; Murphy, 2012). Such models are also known as *covariance selection* or *concentration graph* models, as they rely on the use of partial correlations as a measure of independence of any two variables, by exploiting the relationship between partial correlations and the inverse of the correlation matrix.

Let the asset return $X \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a random variable with a multivariate normal distribution with $\boldsymbol{\mu}$ the $n \times 1$ vector of the expected returns and $\boldsymbol{\Sigma}$ their $n \times n$ covariance matrix. We define the *precision matrix* as the inverse of the covariance matrix: $\boldsymbol{\Omega} \equiv \boldsymbol{\Sigma}^{-1}$.

We can then associate to the vector X an undirected graph defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in the vertex set $\mathcal{V}$ correspond to each element of X, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations: $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | r_{ij} \neq 0\}$, where r_{ij} denotes the partial correlation between two assets, that is, the correlation of the set of the remaining assets, but i and j. In the graph, the edge weights are the partial correlations. It can be shown that matrix of the partial correlations $\boldsymbol{R} = \{r_{ij}\}$ and the precision matrix $\boldsymbol{\Omega}$ are related as follows (Lauritzen, 1996):

$$\boldsymbol{R} = -\boldsymbol{D}\boldsymbol{\Omega}\boldsymbol{D}, \tag{2.5}$$

where $\boldsymbol{D} = diag(\frac{1}{\sqrt{\omega_{11}}}, \ldots, \frac{1}{\sqrt{\omega_{nn}}})$ and ω_{ii} is an entry of $\boldsymbol{\Omega}$.

The estimation of *sparse* precision matrices, that is, precision matrices with most off-diagonal entries exactly equal to zero, is then an important task. A common approach to introduce sparsity is to penalize the maximum likelihood estimation problem by an L_1-norm. In the Gaussian case we can state the graphical lasso (*glasso*) estimation problem as

$$\widehat{\boldsymbol{\Omega}}_{glasso} = \arg\max_{\boldsymbol{\Omega}} \big(log|\boldsymbol{\Omega}| - tr(\boldsymbol{\Omega}\boldsymbol{S}) - \rho||\boldsymbol{\Omega}^{-}||_1 \big), \tag{2.6}$$

where $\boldsymbol{S}$ is the sample covariance matrix, ρ is a tuning parameter that controls the sparsity (i.e. the larger ρ, the larger the number of elements of the precision matrix set equal to zero), $|\cdot|$ is the determinant, $tr(\cdot)$ the trace, $||\cdot||_1$ the L_1-norm of a matrix and $\boldsymbol{\Omega}^{-}$ is a square $n \times n$ matrix with the off-diagonal elements equal to the corresponding elements of $\boldsymbol{\Omega}$ and the diagonal elements equal to zero (Friedman, Hastie, and Tibshirani, 2008).[2]

As pointed out by Yuan and Lin, 2007, the use of an L_1 penalty allows to induce sparsity in the precision matrix, making it possible to perform model selection and parameter estimation simultaneously. The choice of the tuning parameter ρ, that controls the level of sparsity of the estimates assumes therefore a major role in the estimation and it will be examined in Section 2.3.3.

The statistical properties of the *glasso* estimator have been studied, among others, by Banerjee, Ghaoui, and d'Aspremont, 2008, Rothman et al., 2008 and Lam and Fan, 2009. Rothman et al.,

[2]The original specification proposed by Friedman, Hastie, and Tibshirani, 2008 applied the penalty to the entire matrix $\boldsymbol{\Omega}$. The version of the model with the penalty applied to $\boldsymbol{\Omega}$ is the one studied by Rothman et al., 2008 and is currently implemented in the R package *'glasso'* (Friedman, Hastie, and Tibshirani, 2014).

2008 show that under some regularity conditions and for a suitable choice of the parameter ρ, the estimator in (2.6) has a rate of convergence to the true parameter $\mathbf{\Omega}$ in the Frobenius norm of order $\sqrt{((n + k_{\mathbf{\Omega}}) \log(n)/t)}$, where $k_{\mathbf{\Omega}}$ is the number of non-zero off diagonal entries of the true matrix $\mathbf{\Omega}$ and t the number of observations. The main implication is that the convergence is faster for matrices that are truly sparse. Lam and Fan, 2009 studied *glasso*'s consistency and *sparsistency* (i.e., the property that all parameters that are zero are actually estimated as zero with probability tending to one, also known in the literature as *selection consistency*). They show that sparsistency requires a lower bound on the rate of the regularization parameter ρ, while an upper bound is necessary to control the bias introduced by the L_1 penalty and to obtain a consistent estimator.[3] Under some technical conditions, the compatibility of these bounds requires the number of off-diagonal non-zero entries of the true precision matrix to be no larger than $O(n)$ (Lam and Fan, 2009). In practical terms, the true precision matrix has to be sparse enough, and the *glasso* estimates of dense precision matrices will not be consistent. On the other hand, the modelization of sparse precision matrices has the advantage of reducing the variability of the estimates, as it will be discussed in Section 2.2.3 with reference to the *regression hedge*.

The optimization problem (2.6) is convex, as proved by Banerjee, Ghaoui, and d'Aspremont, 2008. Friedman, Hastie, and Tibshirani, 2008 proposed an efficient algorithm to solve it, that is briefly described in Appendix A. They also show that the resulting matrix remains positive definite and invertible if the procedure is initialized with a positive definite matrix.

Empirical analysis show that *glasso* estimates are better conditioned compared to the sample covariance matrix, even when the number of covariates n is close to the number of observations t (see e.g., Goto and Xu, 2015). Moreover, the solution $\widehat{\mathbf{\Omega}}_{glasso}$ is always unique and has bounded eigenvalues,[4] also when $n \leq t$, allowing the use this method also in high-dimensional setting, in which the sample covariance matrix estimate is singular. Finally, we point out that the sparsity of the precision matrix does not necessariliy correspond to the sparsity of the covariance matrix.

2.2.3 Sparse precision matrix estimation and regression hedge

From a financial point of view, the sparsity of the precision matrix can be considered in the framework of *regression hedge*. In fact, as discussed by Stevens, 1998 and Goto and Xu, 2015, the precision matrix has an interpretation in terms of optimal hedging between assets: specifically,

[3] See Theorem 2 and Technical Condition (B) in Lam and Fan, 2009.

[4] see Banerjee, Ghaoui, and d'Aspremont, 2008, Theorem 1.

the ith row (or column) of $\mathbf{\Omega}$ is proportional to the ith asset's hedge portfolio. Such ith hedge portfolio consists of the combination of a long position in the ith asset and $(n-1)$ positions in all the other assets that minimize the variance of the tracking error of the ith asset w.r.t. the remaining $(n-1)$ assets. The ith tracking portfolio is defined as follows:

$$X_{i,\tau} = \alpha_i + \sum_{k=1,k\neq i}^{n} \beta_k^{(i)} X_{k,\tau} + \varepsilon_{i,\tau} \qquad i = 1, \ldots, n, \tag{2.7}$$

where $X_{i,\tau}$ is the ith asset return at time τ, $\beta_k^{(i)}$ is the coefficient of asset k in the regression for asset i, $\varepsilon_{i,\tau}$ is the unhedgeable component of $X_{i,\tau}$.

The regressions in (2.7) are typically defined in the financial literature as *regression hedge*. As shown in Stevens, 1998, the OLS estimates of the βs can be easily related to the precision matrix. We identify the following partition of the sample covariance matrix $\boldsymbol{S}$

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{\backslash i,\backslash i} & \boldsymbol{s}_{\backslash i,i} \\ \boldsymbol{s}'_{\backslash i,i} & s_{i,i} \end{pmatrix}, \tag{2.8}$$

where $X_{\backslash i}$ denotes all the elements of X except the ith, $\boldsymbol{S}_{\backslash i,\backslash i}$ denotes the firsts $n-1$ rows and columns of $\boldsymbol{S}$, $\boldsymbol{s}_{\backslash i,i}$ the first $n-1$ elements of the last column, and $s_{i,i}$ is the element in the last row and column.[5]

We then have that the following relationship holds:

$$\hat{\beta}^{(i)} = \boldsymbol{S}_{\backslash i,\backslash i}^{-1} \boldsymbol{s}_{\backslash i,i}, \tag{2.9}$$

where $\hat{\beta}^{(i)}$ is the $((n-1)\times 1)$ vector of the coefficients in the ith regression hedge. Moreover, let $v_i = \text{var}(\varepsilon_i)$ be the variance of the residual ε_i, then the elements of $\widehat{\mathbf{\Omega}}$ can then be computed as follows:[6]

$$\hat{\omega}_{ij} = \begin{cases} -\dfrac{\hat{\beta}_j^{(i)}}{v_i} & \text{if } i \neq j \\ \dfrac{1}{v_i} & \text{if } i = j \end{cases}. \tag{2.10}$$

Further details can be found in Stevens, 1998.

[5] Notice that this representation implies a permutation of the rows and columns to have the ith asset as the last one.

[6] v_i can be interpreted as the unhedgeable component of $X_{i,t}$.

In financial applications, the regression hedge framework generally suffers in presence of multicollinearity among the regressors. Regularization techniques that allow to set some βs to zero, can then provide more reliable estimates and better out-of-sample performances, at the cost of introducing some bias. One of the most common techniques is the lasso regression, that introduces an L_1-norm penalty in the estimation problem. As we discuss now, the *glasso* estimator allows to introduce an L_1 regulatization of all the regression hedges, while maintaining the relationship with the precision matrix as presented in (2.10).

The naive application of the lasso penalty on each regression hedge, indeed is not consistent with (2.10), since this approach does not constrain $\widehat{\mathbf{\Omega}}$ to be symmetric and positive definite. Instead, the glasso algorithm estimates all the regression hedges iteratively as n coupled lasso problems. The information is shared between the lasso problems through the common estimate of the matrix $\boldsymbol{G}$, providing a positive definite, symmetric and sparse estimate of $\mathbf{\Omega}$ (Friedman, Hastie, and Tibshirani, 2008). This allows to extend the analysis of Stevens, 1998 to the sparse case (Goto and Xu, 2015). On one hand, the use of the lasso penalization in the regression hedge equations introduces a bias. On the other hand it reduces the estimation variation, leading to a more efficient estimator of the precision matrix.

Overall, the *glasso* method has a shrinkage effect on the βs of the regression hedge, filtering the estimation noise in $\mathbf{\Sigma}$ and its effect when computing $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}$. Empirical evidence suggests that the *glasso* estimates of $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are better conditioned than the sample covariance matrix (Goto and Xu, 2015). The spectrum of *glasso* estimates is therefore typically less disperse than the one of sample covariance.

Despite its appealing properties, so far, we are aware of only two applications of *glasso* within asset allocation frameworks (Goto and Xu, 2015; Brownlees, Nualart, and Sun, 2018). Here, we contribute to the literature by providing further evidence when comparing *glasso* to state-of-art methods. Moreover, as widely known, asset returns normality assumption is too stringent, as stylized facts suggest that asset returns have a leptokurtic distribution, which can be better captured by a t-Student assumption (Cont, 2001). Hence, we move one step further by introducing the so-called *tlasso* model that allows to estimate the precision matrix under the assumption of multivariate t-Student distribution of asset returns.

2.2.4 Robust graphical modeling with *tlasso*

As widely discussed in the statistical literature, deviations of returns from Gaussianity can significantly impact the estimation and the inference on GGMs. Asset return distributions typically deviate from normality by having fatter tails and leptokurtic distributions. Hence, the t-Student assumption with a low number of degrees of freedom is considered a better choice to model asset returns. Moreover, relying on such distribution can provide more robust estimates in presence of outliers or contaminated data (Lange, Little, and Taylor, 1989). Recently, Finegold and Drton, 2011 introduced the so called *tlasso*, replacing the *glasso* Gaussian assumption with a t-Student to provide a tool for robust model selection. The *tlasso* algorithm estimates then a sparse precision matrix under the assumption that the data follow a multivariate t-Student distribution.
Let $X = (X_1, \ldots, X_n)$ be a random vector following a multivariate t-Student distribution $t_n(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, df)$, with df degrees of freedom, mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Psi}^{-1}$ ($n \times n$ positive semi-definite matrix). The covariance matrix is then

$$\boldsymbol{\Sigma} = \frac{df}{df-2}\boldsymbol{\Psi}^{-1}, \tag{2.11}$$

and the precision matrix is

$$\boldsymbol{\Omega} = (\boldsymbol{\Sigma})^{-1} = \frac{df-2}{df}\boldsymbol{\Psi}. \tag{2.12}$$

Similarly to the Gaussian case, we can associate a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ in which $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | r_{ij} \neq 0\}$ and the edge weights are the corresponding partial correlations r_{ij} computed from the precision matrix.
Under the t-Student assumption, in contrast to the Gaussian set-up, the absence of correlation does not necessarily correspond to conditional independence (Baba, Shibata, and Sibuya, 2004). However, despite the lack of conditional independence for $\omega_{ij} = 0$ (where ω_{ij} is an element of $\boldsymbol{\Omega}$), we have that, if two nodes j and k are separated by a set of nodes C in $\mathcal{G}$, then X_j and X_k are conditionally uncorrelated given $X_{\{C\}}$ (see Finegold and Drton, 2011, Proposition 1). Disconnected vertices can be considered orthogonal to each other after the effects of other variables are removed. The absence of conditional correlations entails that a mean-square error optimal prediction of variable X_j can be based on the variables X_k, which correspond to neighbours of the node j in the graph.

We adopt the estimation procedure introduced by Finegold and Drton, 2011, that exploits the scale-mixture representation of the multivariate t-Student distribution consisting of a multivariate Gaussian and a gamma distribution (Kotz and Nadarajah, 2004) and uses an EM-algorithm (Expectation-Maximization) described in Appendix A In particular, the E-step consists in the estimation of the mixing gamma variable and the M-step in the estimation of parameters $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Psi}}$ given the latent variable (the degrees of freedom df are assumed to be known in this version of the algorithm). Since the sparse parameter $\boldsymbol{\Psi}$ is the precision matrix of the conditional Gaussian variable, it can be estimated efficiently in the M-step of the algorithm using *glasso* (Finegold and Drton, 2011). The E- and the M-steps are then iterated until convergence. The estimate of the precision matrix $\widehat{\boldsymbol{\Omega}}_{tlasso}$ of the multivariate t-Student vector is finally obtained by rescaling the estimate $\widehat{\boldsymbol{\Psi}}_{tlasso}$ using (2.12).

The *tlasso* procedure is computationally efficient since it is based on *glasso* algorithm at every M-step. While convergence to a stationary point is guaranteed in the penalized versions of EM (McLachlan and Krishnan, 2007), the algorithm is not guaranteed to converge to the global maximum since the *tlasso* penalized log-likelihood function to be maximized is not concave (Finegold and Drton, 2011).

The scale-mixture representation of the multivariate t-Student also allows the regression hedge interpretation as in the Gaussian case. Indeed, the non penalized version of the EM algorithm can be interpreted as an iteratively reweighted least square estimation of the regression of each variable on all the others, as shown in Lange, Little, and Taylor, 1989. In the penalized case, the estimation is consistent with the iteratively reweighted *glasso* estimation.

Finally, we underline that in the empirical application we use the *tlasso* and *glasso* to estimate the correlation matrix and its inverse (rather than the covariance and precision matrices). We then obtain the corresponding estimates of the covariance and precision matrix by scaling the output using the sample variances. This approach ensures that the penalization is not influenced by the scale of the variables (Højsgaard, Edwards, and Lauritzen, 2012) and, as proved by Rothman et al., 2008, ensures a faster convegence of the estimator in the matrix 2-norm.

2.3 Simulation analysis

We conduct a simulation analysis to test the empirical properties of the precision matrix estimates by *glasso* and *tlasso*. In particular, our two main goals are to measure the quality of the estimates

of the true covariance and precision matrices and to assess their impact on the solution of the minimum variance portfolio. We compare the results obtained with *glasso* and *tlasso* to the traditional sample covariance matrix, as well as to the naive equally weighted portfolio strategy (EW) and to two state-of-art covariance estimation methods: random matrix theory filtering (RMT) (Bouchaud and Potters, 2009) and Ledoit Wolf shrinkage estimation (LW) (Ledoit and Wolf, 2004b). Appendix B provides a brief description of these methods with relevant references.

2.3.1 Statistical performance measures

To test the quality of the covariance and precision matrix estimates we measure the error, bias and inefficiency with respect to the true parameters. For explanatory purposes, we describe the measures referring to the covariance matrix $\mathbf{\Sigma}$, which can then be computed also for the precision matrix $\mathbf{\Omega}$.

First, we introduce the following loss function:

$$\text{Loss}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] \equiv ||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||^2, \tag{2.13}$$

where $||\cdot||^2 = tr\left[(\cdot)^2\right]$ is the square of the Frobenius norm. Then, we can compute three measures to quantify the estimation accuracy. First, the *error*, that is the square root of the expected loss between the estimated and the true parameters:

$$\text{Err}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] = \sqrt{\mathbb{E}\left[||\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}||^2\right]}.$$

Second, the *inefficiency*, which is a measure of dispersion of the estimates and is computed as:

$$\text{Inef}[\widehat{\mathbf{\Sigma}}] = \sqrt{\mathbb{E}\left[||\mathbb{E}\left[\widehat{\mathbf{\Sigma}}\right] - \widehat{\mathbf{\Sigma}}||^2\right]}.$$

Finally, the *bias*, that quantifies the distance between the expected value of the estimated covariance and the true parameter:

$$\text{Bias}[\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}] = \sqrt{||\mathbb{E}\left[\widehat{\mathbf{\Sigma}}\right] - \mathbf{\Sigma}||^2}.$$

As widely known, the following relationship holds:

$$\text{Err}^2[\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}] = \text{Bias}^2[\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}] + \text{Inef}^2[\widehat{\boldsymbol{\Sigma}}]. \tag{2.14}$$

Using regularization and shrinkage techniques, we expect to reduce the estimation error by increasing the efficiency of the estimator compared to the sample covariance, typically at the cost of an increased bias.

Glasso and *tlasso* rely on the direct regularization of the precision matrix. Therefore, we expect them to provide good estimates of the optimal assets' weights in the minimum variance portfolio framework, given that the precision matrix represents the input of the optimization. To evaluate the overall impact of the estimation error in $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Omega}}$, we compute the empirical, actual and oracle risk of optimal portfolios. In particular, considering the standard deviation as a risk measure we have:

$$R_{\text{empirical}} = \sqrt{\hat{\mathbf{w}}_{MV}^T \widehat{\boldsymbol{\Sigma}} \hat{\mathbf{w}}_{MV}}, \tag{2.15}$$

$$R_{\text{actual}} = \sqrt{\hat{\mathbf{w}}_{MV}^T \boldsymbol{\Sigma} \hat{\mathbf{w}}_{MV}}, \tag{2.16}$$

$$R_{\text{oracle}} = \sqrt{\mathbf{w}_{MV}^T \boldsymbol{\Sigma} \mathbf{w}_{MV}}, \tag{2.17}$$

where $\boldsymbol{\Sigma}$ is the true covariance matrix, $\widehat{\boldsymbol{\Sigma}}$ is an estimate, $\mathbf{w}_{MV}$ is the optimal vector of minimum variance weights with $\boldsymbol{\Sigma}$ as input and $\hat{\mathbf{w}}_{MV}$ is the optimal weight vector for $\widehat{\boldsymbol{\Sigma}}$.

These measures give us insights on the impact of the estimation error of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Omega}}$ in the optimization process: the empirical risk represents the perceived risk by the investor, the actual risk is the one which the investor is exposed to, while the oracle risk is the minimum risk possible given the true covariance matrix (Fan, Zhang, and Yu, 2012). Since in the real world the last two are unknown, the estimation process should minimize errors due to estimation and provide an empirical risk as close as possible to the oracle and actual.

2.3.2 Simulation set-up

We consider two different approaches for the simulation set-up. The first one is a three-factors model, similar to the one in Fan, Zhang, and Yu, 2012 (which we denote as the *Factor Model*

data). It assumes that the excess returns of the assets are generated according to:

$$X_i = b_{i1}f_1 + b_{i2}f_2 + b_{i3}f_3 + \varepsilon_i \qquad i = 1, \ldots, n, \tag{2.18}$$

where f_1, f_2 and f_3 are the three factors' returns, b_{ik} are the factor loadings for the kth factor and for the ith asset and ε_i is the idiosyncratic noise. The factors' are generated by a multivariate t-Student distribution with 5 degrees of freedom,[7] while the idiosyncratic terms are generated from a univariate t-Student distribution. The parameters of factor returns, factor loadings and level of idiosyncratic noise are calibrated on real market data (see Fan, Zhang, and Yu, 2012 for more details).

In the second approach (henceforth *Simulated S&P100*), we generate the assets' returns by a multivariate t-Student distribution with 5 degrees of freedom and a covariance matrix estimated on the daily returns of constituents of the S&P100 equity index for the period 01/01/2006 – 31/12/2016, adding to each asset a noise factor distributed as univariate t-Student with variance equal to 0.1 times the variance of each asset.

For each setting, we consider two configurations characterized by different dimensionality: 50 and 85 assets, respectively. In both cases the parameters are estimated over a window of 100 observations and we consider 30 simulation runs. The number of degrees of freedom for the *tlasso* has been set equal to 5.

2.3.3 Optimal choice of ρ

The structure of the precision matrix estimated by *glasso* and *tlasso* depends largely on the choice of the penalization parameter ρ, that controls the level of sparsity in the precision matrix. We select the optimal ρ on a grid of values using the Bayesian Information Criterion (BIC)

$$BIC = -2log(Lik_{\rho_i}) + k_{\widehat{\mathbf{\Omega}}} \times log(t), \tag{2.19}$$

where Lik_{ρ_i} is the value of the likelihood function corresponding to the ith value of ρ in the grid, $k_{\widehat{\mathbf{\Omega}}}$ is the number of non-null elements in the estimate of the precision matrix, and t the

[7] In the original model the factors followed a multivariate normal distribution (Fan, Zhang, and Yu, 2012). We used a t-Student to capture the leptokurtic distribution of financial time series (Cont, 2001).

number of observations.[8] The grid is composed by 20 logarithmically spaced values between 0 and 1. The choice of this interval guarantees that the *glasso* estimates span from a completely dense precision matrix ($\rho = 0$) to a completely sparse one ($\rho = 1$) when estimated using the correlation matrix as input.[9]. For *tlasso* such result is not guaranteed, but we found empirically that such interval is wide enough to include the estimate characterized by the optimal BIC in all the cases (see Figures 2.1 and 2.2).

For sake of brevity, we report exclusively the parameter calibration of *tlasso*; the procedure and the results are analogous for *glasso* and available from the authors upon request.

Figures 2.1 and 2.2 show for *tlasso* the values of error, bias and inefficiency of the covariance matrix estimator, as well as the in- and out-of-sample standard deviation as a function of ρ for the *Factor Model* and *Simulated S&P100*, respectively. In both cases we consider 85 assets estimated on a window of 100 observations. Panel (a) reports the value of BIC, Panel (b) the error, bias and inefficiency of the estimation of covariance matrix and Panel (c) the in-sample and out-of-sample standard deviations. We observe in Panel (a) that the BIC optimal model lies within the interval $\rho \in [0, 1]$, in both cases with values close to 0.4, characterized by a medium sparsity level in $\widehat{\mathbf{\Omega}}$. From Panel (b), the bias increases with the values of ρ, while the inefficiency decreases. This pattern is consistent with the fact that, for the *tlasso* estimates computed with higher values of ρ, the number of parameters to estimate is smaller, given that more elements of the precision matrix are set equal to zero. The overall estimation error reaches a minimum for intermediate values of ρ not distant from the ones chosen by the BIC. Concerning the portfolio performances, we see from Panel (c) that the effect of ρ on the out-of-sample standard deviation is different for the two simulation settings: in the case of *Factor Model*, the portfolios have minimum risk with a rather low value of ρ, while for *Simulated S&P100*, the out-of-sample risk is minimized for a wider range of ρ.

[8] In the case of *glasso* we refer to the likelihood of a multivariate normal distribution, while with *tlasso* we refer to the one of a multivariate t-Student distribution.

[9] The result follows from Corollary 1 in Witten, Friedman, and Simon, 2011, according to which the ith node is fully unconnected to all other nodes if and only if $|\mathbf{\Sigma}_{ij}| \leq \rho \quad \forall i \neq j$. When $\mathbf{\Sigma}$ is the correlation matrix, all its elements are smaller or equal to one and therefore for $\rho = 1$ all the elements are disconnected, that is, the precision matrix is diagonal.

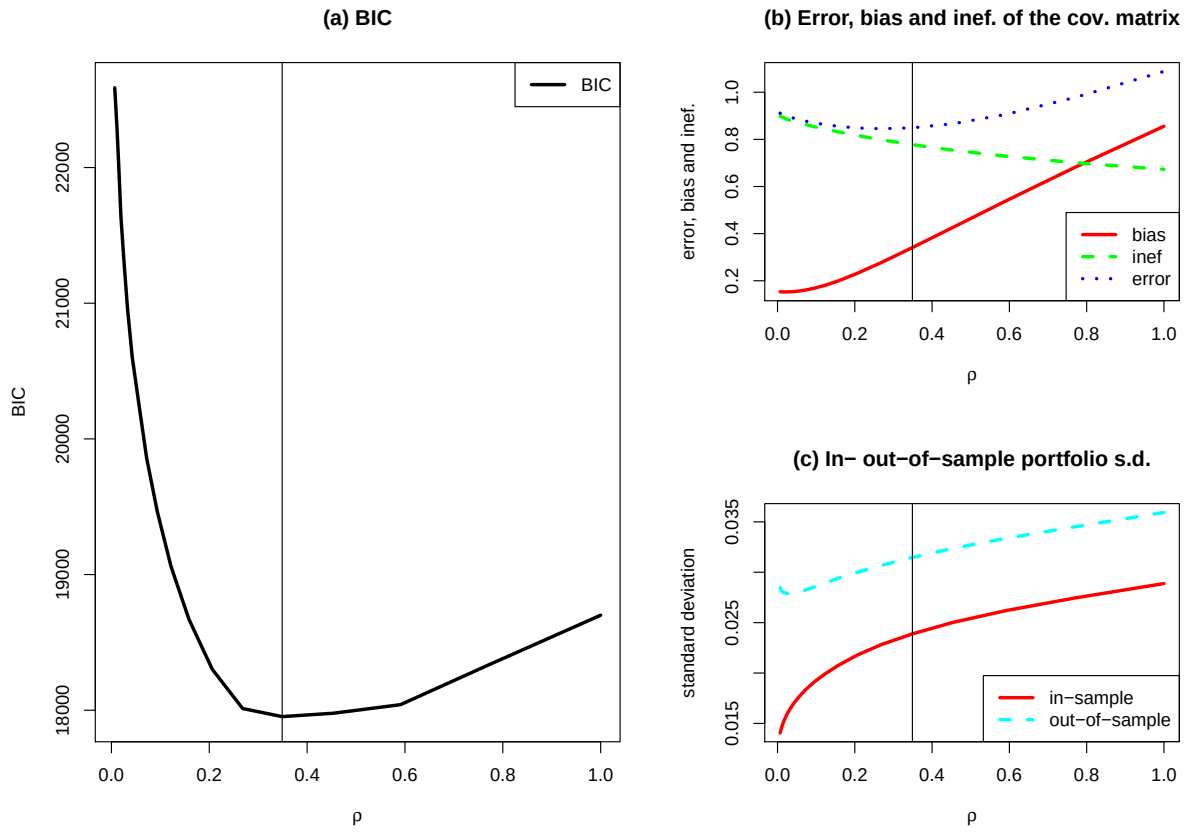


FIGURE 2.1: Optimal Choice of ρ for *tlasso*, *Factor Model*, 85 assets. Panel (a) displays the value of BIC for every value of ρ in the grid. Panel (b) shows the values of error, bias and inefficiency of the estimation of the covariance matrix (see (2.13) and (2.14)). Panel (c) displays the in- and out-of-sample standard deviation of the optimal minimum variance portfolios. The value reported are based on 30 runs and the vertical lines denote the median of the optimal ρ across the runs.

2.3.4 Simulation results

Accuracy of the estimates

Table 2.1 reports bias, inefficiency and error for the covariance (columns 2–4) and the precision matrix (columns 5–7) computed on 30 runs for four test cases: *Factor Model* and *Simulated S&P100*, with 50 and 85 assets. It also reports the average condition numbers of the estimates for 30 runs (column 8) and the ones of the true covariance matrices. Concerning the estimation of the covariance matrix, we observe that *glasso*, *tlasso*, RMT and LW present in all the cases a low inefficiency compared to the sample covariance matrix. This comes at the cost of a higher bias. As a consequence of these two opposite effects, the overall error levels end up being similar. Indeed, *glasso*, *tlasso* and LW provide only minor improvements in terms of overall error with

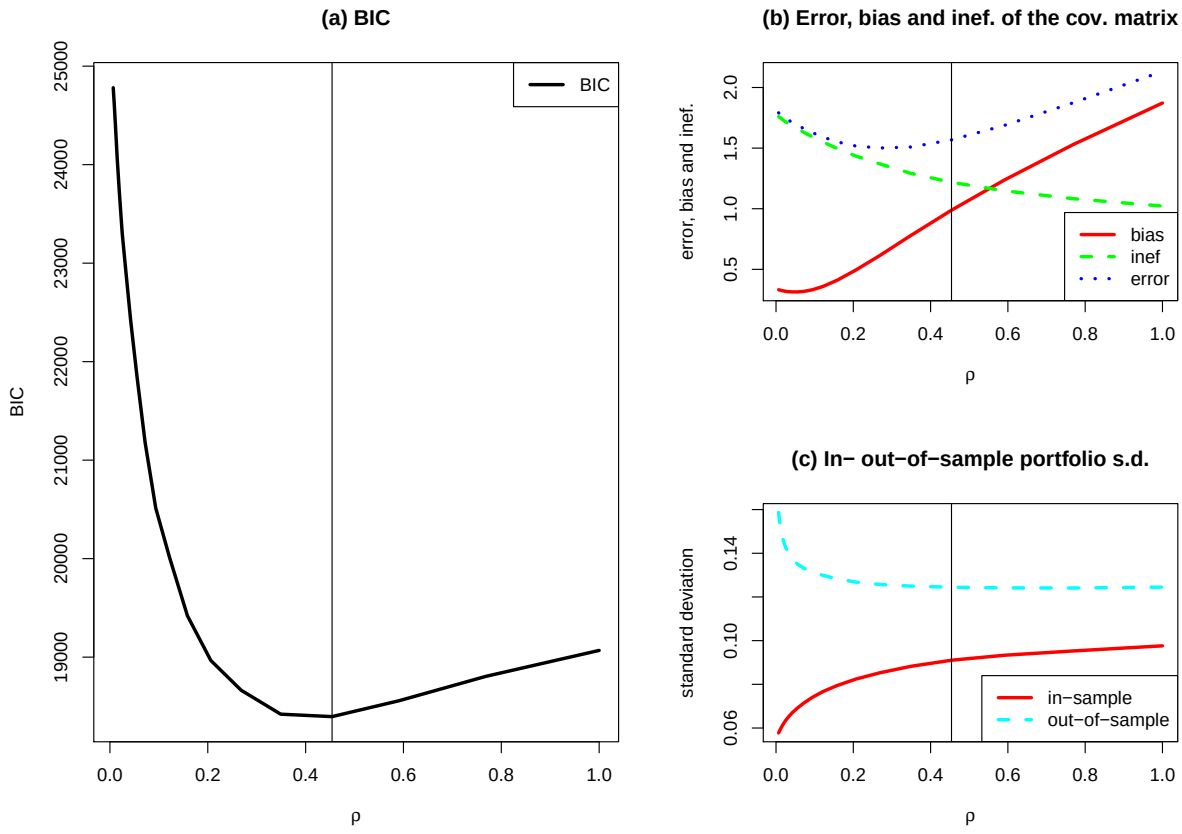


FIGURE 2.2: Optimal Choice of ρ for *tlasso*, *Simulated S&P100*, 85 assets. Panel (a) displays the value of BIC for every value of ρ in the grid. Panel (b) shows the values of error, bias and inefficiency of the estimation of the covariance matrix (see (2.13) and (2.14)). Panel (c) displays the in- and out-of-sample standard deviation of the optimal minimum variance portfolios. The value reported are based on 30 runs and the vertical lines denote the median of the optimal ρ across the runs.

respect to the naive sample covariance approach, while RMT shows in three out of four test cases an estimation error larger than the sample covariance due to a particularly high bias. The results are consistent for all the simulation set-ups. The fact that the *Factor Model* test cases are characterized by larger errors then the *Simulated S&P100* can be explained by higher values of the entries of the covariance matrix, resulting from higher volatility and collinearity in the data. In the context of minimum variance portfolio selection, the focus is on the estimation of the precision matrix, that is the input of the closed form optimal solutions in (2.3). Indeed, the analysis of the estimation error of such matrix displays a rather different picture, more aligned to the well documented pitfalls of minimum variance portfolios estimated using sample covariance (Michaud, 1989). The estimation of the precision matrix obtained by inverting the sample covariance is indeed characterized by an error much higher than the alternative estimates in all

the test cases considered, especially when the number of assets is large. For instance, in Panel 2 (*Factor Model* with $n = 85$), the error for the precision matrix is equal to 102120.68, while the error of the estimates obtained using *glasso* and *tlasso* is equal to less than a tenth of it: 8928.08 and 8985.56, respectively. The large error of the sample covariance estimator is not surprising, given that such matrices are characterized by high condition numbers, and their inverses are therefore highly sensitive to estimation error (the most relevant eigenvectors of the precision matrix are indeed the ones corresponding to the smallest eigenvalues of the covariance matrix, which are typically dominated by noise). *Glasso* and *tlasso* show good performances, with the smallest error in test case 1, 3 and 4. The difference with RMT and LW in terms of error are generally moderate. The results are similar for the *Factor Model* and the *Simulated S&P100* set-ups, although in the latter the absolute value of the estimation errors for the precision matrix are smaller than in the *Factor Model*. This difference is probably due to the worse conditioning of the covariance matrices in the *Factor Model* cases, which amplifies the estimation error of the covariance matrix.

Empirical, actual and oracle risk

Table 2.2 reports empirical, actual and oracle risk. Concerning the *Factor Model* test cases (Panel 1 and 2), we see that, in terms of actual risk, *glasso* and *tlasso* obtain the best results: in the case with 50 assets, they have an actual risk of 0.038 and 0.037 (Panel 1), lower than the LW, RMT and EW portfolios. They are also the portfolios with the lowest actual risk in the setting with 85 assets (Panel 2). The portfolios estimated using the sample covariance are a peculiar case, since they obtain the lowest actual risk for 50 assets, but have the highest actual risk for 85 assets. This is consistent with the high error in the estimation of the precision matrix, as highlighted in Section 2.3.4. Such portfolios are also the ones with the largest difference between actual and empirical risk (especially in Panel 2, 85 assets settings, where they are equal to 0.045 and 0.006, respectively), while the other techniques provide much less divergent values (in Panel 2 for instance the actual and empirical risk for *tlasso* are equal to 0.030 and 0.032). As the empirical risk is the only one known to investors in real-world applications, positive differences between actual and empirical can lead to risk underestimation.

In the *Simulated S&P100* test cases we observe that, again, *glasso* and *tlasso* portfolios have the lowest actual risk. The advantage over RMT and LW in this case is limited, much smaller than the *Factor Model* framework (Panels 3 and 4). Sample covariance portfolios show a particularly

TABLE 2.1: Bias, inefficiency and error for the estimation of covariance and precision matrices; condition number of covariance matrix. Values are computed in 30 runs.

Panel 1 - *Factor Model* - 50 assets, 100 obs. (cond. number of Σ: 12359.42)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.08	0.64	0.64	8390.62	4801.40	9667.26	29770.64
glasso	0.19	0.60	0.63	5476.29	190.52	5479.60	1401.60
tlasso	0.16	0.57	0.59	5331.74	244.77	5337.35	1645.16
RMT	0.73	0.43	0.85	6127.47	79.91	6127.99	425.54
LW	0.21	0.60	0.63	5579.33	327.75	5588.94	1336.68
Panel 2 - *Factor Model* - 85 assets, 100 obs. (cond. number of Σ: 27016.97)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.12	0.96	0.97	74698.65	69632.92	102120.68	296622.82
glasso	0.32	0.89	0.94	8924.83	240.91	8928.08	2119.29
tlasso	0.35	0.82	0.89	8981.38	274.12	8985.56	1938.62
RMT	1.27	0.59	1.41	9706.34	100.23	9706.86	679.27
LW	0.30	0.90	0.95	8732.19	702.21	8760.37	2825.45
Panel 3 - *Simulated S&P100* - 50 assets, 100 obs. (cond. number of Σ: 1064.35)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.20	1.20	1.22	402.24	601.64	723.72	2014.03
glasso	0.47	1.07	1.17	263.69	57.79	269.94	275.73
tlasso	0.87	0.75	1.15	284.75	45.29	288.32	190.13
RMT	1.25	0.73	1.45	312.06	31.60	313.66	137.84
LW	0.53	0.92	1.06	289.43	64.61	296.56	294.64
Panel 4 - *Simulated S&P100* - 85 assets, 100 obs. (cond. number of Σ: 1630.18)							
	Covariance matrix			Precision matrix			Cond. number
	bias	inef	error	bias	inef	error	
sample cov.	0.77	3.78	3.85	4327.64	9170.36	10140.22	37909.67
glasso	1.21	3.42	3.63	309.70	79.65	319.77	446.39
tlasso	1.49	1.58	2.17	330.86	71.30	338.46	315.02
RMT	2.13	2.35	3.17	367.85	50.81	371.34	218.32
LW	0.99	1.82	2.07	332.10	103.70	347.91	428.29

high actual risk (0.182 and 0.270 for the settings with 50 and 85 assets, respectively, against the 0.135 and 0.122 for *tlasso*), highlighting the limits of this simple estimation technique. Equally weighted portfolios, typically considered a difficult benchmark to beat (DeMiguel, Garlappi, and Uppal, 2009), do not seem to show interesting performances in terms of the risk measures that we consider, exhibiting the highest actual risk in three out of four test cases.

Summing up, *glasso* and *tlasso* perform well in all the test cases, exhibiting low portfolio risk

exposures compared to the alternative techniques, especially in markets characterized by ill-conditioned covariance matrices (i.e. the *Factor Model* case). This suggests that these techniques might be particularly suitable in presence of multicollinearity, a characteristic typically associated to financial crises.

TABLE 2.2: Empirical, actual and oracle risk for the optimal portfolios. Average results over 30 runs.

Panel 1 - *Factor Model*						
	50 assets, 100 observations			85 assets, 100 observations		
	empirical	actual	oracle	empirical	actual	oracle
sample cov.	0.015	0.032	0.022	0.006	0.045	0.018
glasso	0.039	0.038	0.022	0.031	0.030	0.018
tlasso	0.038	0.037	0.022	0.032	0.030	0.018
RMT	0.052	0.053	0.022	0.042	0.041	0.018
LW	0.039	0.044	0.022	0.028	0.037	0.018
EW	0.190	0.193	0.022	0.186	0.188	0.018
Panel 2 - *Simulated S&P100*						
	50 assets, 100 observations			85 assets, 100 observations		
	empirical	actual	oracle	empirical	actual	oracle
sample cov.	0.074	0.182	0.122	0.032	0.270	0.103
glasso	0.098	0.136	0.122	0.079	0.123	0.103
tlasso	0.101	0.135	0.122	0.082	0.122	0.103
RMT	0.100	0.136	0.122	0.080	0.125	0.103
LW	0.102	0.138	0.122	0.080	0.129	0.103
EW	0.239	0.244	0.122	0.217	0.228	0.103

2.4 Real-world data analysis

2.4.1 Empirical set-up

We test the performance of the portfolio determined by *glasso* and *tlasso* on four real-world datasets. The first two are provided by Kenneth French and are publicly available on his website.[10] One consists of the monthly returns of 48 US industry portfolios (FF 48) and the other of the returns of 100 portfolios formed on size and book-to-market ratio (FF 100) of US companies. The third and the fourth datasets are made of the stock returns of the constituents of S&P 100, at monthly and daily frequency (S&P 100 monthly and S&P 100 daily). The choice of the datasets

[10] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

spans different combinations of constituents, dimensionality and sampling frequency, providing robustness to the results.

We analyse the out-of-sample performances using a rolling window approach, rebalancing the portfolios every three months by computing the optimal global minimum variance portfolio on a window of fixed size. The out-of-sample period is defined for all the portfolios from January 2006 to December 2016. The estimation windows consist of 1 year (252 observations) in the case of the daily data, while for the monthly data we considered longer time windows in order to have sufficient data points (10 years, 120 observations). The main characteristics of the datasets are summarized in Table 2.3.

We evaluate the resulting portfolios in terms of risk/return profile (computing standard deviation, average return and Sharpe ratio) and in terms of portfolio composition, computing statistics relative to shorting, diversification and turnover. As in the simulation study, we estimated the *tlasso* with 5 degrees of freedom.

TABLE 2.3: Descriptive statistics of the real-world datasets. The first three columns report the number of assets (n), the window size for the calibration (t) and the ratio between these two values (n/t), respectively. Columns 4 and 5 report the period spanned by each dataset and the frequency of the data. Note that, concerning the S&P100, we included in the analysis only the assets whose time series spanned the entire time period.

Portfolio	n	t	n/t	time period	data freq.
FF 48	48	120	0.40	01/1996 - 12/2016	monthly
FF 100	100	120	0.83	01/1996 - 12/2016	monthly
S&P 100 monthly	86	120	0.72	01/1996 - 12/2016	monthly
S&P 100 daily	91	252	0.36	01/2005 - 12/2016	daily

2.4.2 Empirical results

Table 2.4 displays the performance measures for the portfolios estimated on real data. As we construct minimum variance portfolios, we focus in particular on the standard deviation, which is the quantity of interest in the optimization. Still, we also compute the average return and the Sharpe ratio of the portfolios to analyse the risk-adjusted return profiles.

As expected, the out-of-sample standard deviation is larger than the in-sample in all the cases. The difference between the two is especially relevant for the portfolios computed using the sample covariance estimator when the number of assets is large compared to the length of the estimation window. For instance, in the FF 100 case, in which we consider 100 assets and the estimation window is made of 120 observations, the in-sample standard deviation for the sample covariance portfolio is equal to only 3.8%, while its out-of-sample counterpart is equal to 23.0%, more than 6 times larger. This provides further evidence to previous findings related to the large impact of estimation error, when using the sample covariance (e.g., Michaud, 1989). The other estimation techniques do a better job at minimizing the out-of-sample standard deviation and reducing the gap between in- and out-of-sample results. In particular, *tlasso* compares favourably to the alternative approaches, obtaining the lowest out-of-sample standard deviation in the FF 48 and S&P 100 daily datasets and performing well also in the FF 100 and S&P 100 monthly. Table 2.5 shows the differences between the out-of-sample standard deviation of *tlasso* portfolios and the ones optimized using other techniques. The confidence levels are computed using the Ledoit and Wolf bootstrap confidence interval for the ratio of two variances (Ledoit and Wolf, 2011). We observe that the out-of-sample standard deviation of *tlasso* is lower, and statistically significantly different than both *glasso* and LW in the FF 48 and S&P 100 daily datasets, while it is never higher and statistically significantly different from any other model.

We underline that *tlasso* in real-world scenarios shows better performance compared to *glasso*, while in the simulation study they obtain similar results. This can be related to the robustness of *tlasso* to misspecification and outliers in the data (Finegold and Drton, 2011), and therefore the better capability of dealing with the typical fat-tail distribution of asset returns. The comparison between the results on the S&P 100 with daily and monthly returns allows to better characterize the relationship between *glasso* and *tlasso*. Indeed, from Table 2.5, we notice that the difference in the out-of-sample standard deviation of the *tlasso* and the *glasso* portfolios is statistically significant when using data with daily frequency, but not for the monthly ones. This may be due to the stylized property of *aggregational Gaussianity*, as reported by Cont, 2001, which refers to the fact that the distribution of equity returns tends to have ticker tails for shorter time frequencies (e.g., daily), while being better approximated by a Gaussian distribution as the time frequency increases (e.g., monthly). Moreover, when the length of the estimation time interval is large compared to the number of asset weights to be estimated (e.g., 120 observations for FF 48), and consequently the precision matrices are less ill-conditioned, *tlasso* portfolios exhibit better

out-of-sample risk properties than *glasso*. Finally, we notice that the equally weighted portfolios generally show high standard deviations compared to *glasso*, *tlasso*, RMT and LW. In two of the datasets, however, it performs better than the sample covariance matrix portfolios.

Concerning the Sharpe ratio, we obtain different results across the test cases: the sample covariance estimator shows the most inconsistent performance, with the highest out-of-sample Sharpe ratio in the FF 100 case (1.039) and the lowest in the FF 48 and S&P 100 monthly (0.384 and 0.151, respectively). The other estimators generally obtain good performance in all the test cases, with *glasso* and *tlasso* displaying the highest out-of-sample Sharpe ratios in the S&P 100 daily, RMT in the S&P 100 monthly and FF 48 and LW in the FF 100 case. Finally, we notice that the EW portfolio, despite beating the sample covariance portfolio in FF 48 and S&P100 monthly, does not seem to be competitive with the other methods in terms of risk adjusted performance.

Table 2.6 reports summary portfolio statistics. Columns 2 to 4 display the gross exposure (i.e., the sum of absolute values of the portfolio weights), the total negative exposures and the maximum negative exposure of individual assets. We see that the sample covariance portfolios are characterized by extreme exposures, especially for the FF 100 portfolios, where the gross exposures is more than 22 times higher than the initial endowment, due to a short exposure of 10.758. *glasso*, *tlasso*, RMT and LW show considerably lower exposures, both in terms of whole portfolios and of individual securities. The EW portfolio, as it is long only by construction, is trivially the one with the lowest exposures. None of the methods promote sparsity of the weights, therefore in all cases the percentage of active position is 100%. The level of diversification is computed by the modified Herfindahl concentration coefficient H^*. Such measure can deal with short portfolio exposures and takes the lowest value for the most diversified portfolio (i.e., the EW portfolio). The concentration levels are similar for different portfolios, including the sample covariance one. This suggests that the main differences in the portfolio structure result from the allocation of weights and the level of gross exposures and not from the excessive concentration in a limited number of assets. Finally, we compute the turnover rate of the portfolios. The sample covariance portfolios show the worst performance in terms of turnover, due to both the amount of gross exposure and the estimation error. The equally weighted portfolio has zero turnover by construction and all the other techniques show considerably lower turnover levels than the sample covariance case.

TABLE 2.4: In-sample and out-of-sample (oos) standard deviation, mean return and Sharpe ratio for for real-world data analysis.

	Standard deviation		Mean return		Sharpe ratio	
Model	in-sample	oos	in-sample	oos	in-sample	oos
Panel 1 - 48 Industry Portfolios (FF 48)						
sample cov.	0.068	0.124	0.070	0.048	1.042	0.384
glasso	0.083	0.110	0.074	0.077	0.889	0.697
tlasso	0.081	0.106	0.072	0.065	0.887	0.618
RMT	0.085	0.108	0.075	0.081	0.885	0.756
LW	0.074	0.113	0.068	0.070	0.917	0.620
EW	0.166	0.179	0.097	0.105	0.592	0.587
Panel 2 - 100 Size and Book-to-Market Portfolios (FF 100)						
sample cov.	0.038	0.230	0.228	0.239	6.147	1.039
glasso	0.085	0.127	0.102	0.069	1.189	0.546
tlasso	0.085	0.124	0.099	0.075	1.152	0.608
RMT	0.098	0.129	0.088	0.068	0.890	0.531
LW	0.070	0.121	0.105	0.092	1.496	0.758
EW	0.187	0.187	0.091	0.073	0.493	0.392
Panel 3 - S&P 100 2006-2016 - monthly data (S&P 100 mon.)						
sample cov.	0.039	0.206	0.100	0.031	2.652	0.151
glasso	0.084	0.119	0.099	0.079	1.211	0.663
tlasso	0.075	0.116	0.099	0.066	1.330	0.570
RMT	0.074	0.107	0.110	0.084	1.488	0.788
LW	0.069	0.115	0.094	0.052	1.362	0.453
EW	0.160	0.160	0.077	0.082	0.489	0.516
Panel 4 - S&P 100 2006-2016 - daily data (S&P 100 daily)						
sample cov.	0.067	0.138	0.090	0.090	1.531	0.649
glasso	0.086	0.124	0.110	0.108	1.411	0.872
tlasso	0.083	0.121	0.102	0.108	1.377	0.892
RMT	0.085	0.122	0.099	0.097	1.308	0.799
LW	0.074	0.125	0.100	0.094	1.475	0.755
EW	0.181	0.202	0.089	0.084	0.844	0.414

2.5 Conclusion

The estimation of the precision matrix is fundamental to the implementation of several quantitative investment strategies. In this chapter, we consider two innovative methods based on Markovian graphs: *glasso* and *tlasso*. These techniques allow us to regularize the estimation of the precision matrix (i.e., the inverse of the covariance matrix) by imposing a constraint on the L_1-norm, assuming Gaussian and t-Student distributions, respectively. We test the models both on simulated and real world data, measuring the quality of the estimation and the out-of-sample

TABLE 2.5: Difference in out-of-sample standard deviations between the *tlasso* portfolios and the alternative methods. Negative numbers denote a lower standard deviation for *tlasso* compared to the alternative method (i.e., sample covariance, *glasso*, RMT, LW and EW). Statistical significance has been assessed with Ledoit and Wolf procedure (Ledoit and Wolf, 2011). *,**,*** denote 90%, 95% and 99% confidence level, respectively.

	sample cov.	glasso	RMT	LW	EW
FF 48	-0.019**	-0.004***	-0.002	-0.007*	-0.073***
FF 100	-0.106***	-0.002	-0.004	0.003	-0.063***
S&P 100 mon.	-0.091***	-0.003	0.009	0.001	-0.044***
S&P 100 daily	-0.018***	-0.003***	-0.001	-0.004***	-0.081***

performances of the optimized global minimum variance portfolios. We compare them to the naive sample covariance estimator, equally weighted portfolios and two state-of-art techniques: random matrix and Ledoit Wolf shrinkage methods. According to our analysis, *glasso* and *tlasso* show interesting results: in the simulation framework they both improve the estimation of the precision matrix compared to the alternative techniques, reducing the bias and error of the estimates, and the actual risk in simulated portfolios, especially with ill-conditioned matrices. When applied to real data, they obtain good out-of-sample performances. *Tlasso*, which is more robust to misspecification and outliers, stands out for the low out-of-sample standard deviation, providing better results than *glasso* by just paying a small price in computational efficiency compared to *glasso*. The results are consistent across all the dataset considered, and the advantage of *tlasso* over *glasso* is larger when using data with daily frequency compared to monthly. Moreover, *glasso* and *tlasso* limit the portfolio short exposures and reduce considerably the turnover compared to the sample covariance estimator. We underline that the analysis proposed here was focused on a simple investment strategy in order to isolate the effect of covariance matrix estimator, but these regularization methods can be applied in any asset allocation framework that requires the estimation of the covariance and precision matrices, such as risk parity portfolio (Roncalli, 2014) and maximum diversification portfolio (Choueifaty and Coignard, 2008), as well as for index tracking investment strategies. Indeed, the good portfolio performance, together with the simple implementation, make *glasso* and *tlasso* interesting tools for the Fintech industry and for the implementation of data-driven investment models, suitable also for distressed markets, when covariance matrices of the assets tend to be ill-conditioned.

TABLE 2.6: Portfolio statistics on real-world data. The Table reports, from left to right, gross exposure (gross exp.) ($\sum_i |w_i|$), total short exposure (short exp), maximum negative exposure of individual assets (max short), percentage of active positions in the portfolios (active pos.), modified Herfindahl diversification index corrected to account for short portfolio ($H^* = \sum_i w_i'^2$ where $w_i' = w_i/(\sum_i |w_i|)$) and portfolio turnover (turnover). The reported values are the average across all the rebalancing periods.

Model	gross exp.	short exp.	max short	active pos.	H^*	turnover
Panel 1 - 48 Industry Portfolios						
sample cov.	4.426	1.713	-0.213	100%	0.036	1.149
glasso	2.142	0.571	-0.068	100%	0.040	0.282
tlasso	2.253	0.627	-0.070	100%	0.039	0.317
RMT	1.894	0.447	-0.046	100%	0.046	0.201
LW	2.866	0.933	-0.103	100%	0.040	0.455
EW	1.000	0.000	0.000	100%	0.021	0.000
Panel 2 - 100 Size and Book-to-Market Portfolios						
sample cov.	22.516	10.758	-0.807	100%	0.017	11.799
glasso	4.656	1.828	-0.121	100%	0.018	0.703
tlasso	4.573	1.786	-0.114	100%	0.018	0.798
RMT	3.022	1.011	-0.062	100%	0.022	0.272
LW	5.512	2.256	-0.141	100%	0.018	0.973
EW	1.000	0.000	0.000	100%	0.010	0.000
Panel 3 - S&P 100 2006-2016 (monthly data)						
sample cov.	6.954	2.977	-0.297	100%	0.020	3.016
glasso	1.488	0.244	-0.025	100%	0.022	0.309
tlasso	1.707	0.353	-0.030	100%	0.025	0.271
RMT	1.850	0.425	-0.036	100%	0.023	0.258
LW	2.190	0.595	-0.039	100%	0.024	0.338
EW	1.000	0.000	0.000	100%	0.011	0.000
Panel 4 - S&P 100 2006-2016 (daily data)						
sample cov.	4.569	1.785	-0.186	100%	0.022	3.192
glasso	2.376	0.688	-0.061	100%	0.023	1.058
tlasso	2.479	0.739	-0.058	100%	0.025	1.113
RMT	2.121	0.561	-0.043	100%	0.026	0.888
LW	3.043	1.021	-0.078	100%	0.025	1.650
EW	1.000	0.000	0.000	100%	0.011	0.000

Chapter 3

Robust and sparse banking network estimation with *tlasso*

Acknowledgement

This chapter is largely based on the article "Robust and Sparse Banking Network Estimation" co-authored by Sandra Paterlini and Rosella Giacometti, published on the *European Journal of Operational Research* (Torri, Giacometti, and Paterlini, 2018).

3.1 Introduction

The recent financial crises have pointed out the need to capture the interconnectedness between the banks in a system, in order to introduce risk-management tools capable to better control potential spillover effects in case of distress of some institutions. The "too-big-to-fail" slogan that became popular just after the 2008 financial crisis, was soon replaced by "too-interconnected-to-fail". This new expression acknowledges that systemic events can be triggered not only by the distress of large institutions, but also by small entities that cover specific roles in the system, or that are in key positions within the network. The relevance of interconnections is stressed also in the Basel Committee's regulation for assessing and identifying global systemically important banks (G-SIBs). This regulation requires that global systemic importance should be measured in terms of the impact that a bankâĂŹs failure could have on the global financial system and the economy, rather than the probability that such a failure occurs. Hence, Basel introduced an indicator-based measurement approach that considers not only size but also interconnectedness, complexity, substitutability and the level of cross-jurisdictional activities (FSB, 2013).

Applications of network theory and statistical tools that can capture the dependence structure within an entire system have then flourished in finance and economics in the aftermath of the crisis, in conjunction with the rapid development of network modelling in other fields, such as sociology and biology (e.g., Albert and Barabási, 2002; Watts and Strogatz, 1998; Newman and Park, 2003). In a banking system, network nodes are typically financial institutions, and edges capture the relationship among them. The goal of these applications is to model the dynamics of financial co-movements and contagion, the potential impact on the system, and ideally to set up improved risk management tools to avoid spillover effects (see e.g., Battiston et al., 2012a; Paltalidis et al., 2015; Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015; Cont and Minca, 2016).[1]

Still, much research on the estimation of the network structure of a banking system and comparison across different modelling approaches is needed. The first contributions in the field have mostly focused on interbank lending markets, where the network structure consists in the set of bilateral credit exposures between banks (see for instance Mistrulli, 2011 and Lelyveld and Veld, 2014). Most of the initial work was set up by regulators for stress testing and to understand potential consequences of new regulation. A main limitation is still that bilateral exposures are generally non-disclosed and unavailable to researchers outside central banks. In absence of bilateral data, the network structure must be reconstructed from partial data or one must rely on other sources, such as time series of equity prices or CDS spreads.

There exists a strand of literature that reconstructs the bilateral interbank exposures using the total exposures of each bank towards the entire banking system through statistical techniques such as maximum entropy or minimum density algorithms (Elsinger et al., 2013; Anand, Craig, and Von Peter, 2015). Nevertheless, such approaches usually require strong assumptions on the behaviour of the banks in the interbank market and they often fail to represent the true network structure (Mistrulli, 2011).

An alternative is instead to infer the network structure using tools that can capture co-movements and dependence patterns between financial time series to establish the existence of links among banks. An example is the estimation of credit risk networks from credit default swap (CDS) spreads or equity price times series (e.g., Puliga, Caldarelli, and Battiston, 2014; Anufriev and Panchenko, 2015; Billio et al., 2012). One major advantage of these approaches is that they rely

[1] Other common applications of network theory in finance include the assessment of credit risk for SMEs including spatial dependence (Fernandes and Artes, 2016) and the linkages between different markets, such as Bekiros et al., 2017 that analyses the causal linkages between equity and commodity futures markets.

on public data and well known statistical modelling techniques. We focus in particular on graphical models, that are probabilistic models, nowadays widely used in machine learning (Murphy, 2012), in which the graph captures the conditional dependence structure between the nodes. Gaussian graphical models are possibly the most widely known. By relying on the assumption that nodes are random variables with a multivariate normal distribution, edges are estimated as partial correlations in these models. Values of zero imply then independence between nodes, while the conditional probability among nodes factorizes according to the partial correlation graph (Lauritzen, 1996). The advantage of using conditional dependence, compared to direct correlations, is that it allows us to single out the direct co-movements between nodes, controlling for spurious connections due to common exposures. Our work aims to fill some gaps in the literature by proposing a robust method, based on graphical models and regularization techniques, for estimating only the relevant links in the network from CDS spread time series data. The estimated network then provides information on the dependence among banks and possible paths of shock propagation. We underline that we consider synchronous, and not delayed co-movements of time series, and that we estimate sparse undirected networks. As a consequence, we cannot interpret the edges as causal relationships, at least in a Granger (i.e. intertemporal) sense. Other approaches, such as *Granger causality networks* (Billio et al., 2012), variance decomposition (Diebold and Yılmaz, 2014) and transfer entropy (Bekiros et al., 2017) would allow us to consider intertemporal and directed relationships by modeling the data, either in a Vector AutoRegression (VAR) framework or by using an informational theoretical approach. Still, we believe that given the aim of our work, partial correlation networks deliver an appropriate estimate of the connectivity patterns among banks since, assuming that the markets are sufficiently liquid and promptly reacting, the information available to the market should be discounted by the prices fast enough to be reflected by synchronous relationships.[2] Although it is not always possible to match the co-movement networks to the actual network structure of the interbank market, some studies highlight their usefulness to capture salient aspects of the financial system, which allows regulators to design more effective policies and to set-up risk monitoring and mitigating tools (Anufriev and Panchenko, 2015). Other studies found some interdependence between co-movement networks and interbank exposures networks. It is the case of Abbassi et al., 2017, that

[2]Intertemporal causation may indeed indicate the presence of financial turmoil, as reported by Billio et al., 2012, that show how Granger causality networks may allow to identify periods of market dislocation and distress, and Jenkins, Kimbrough, and Wang, 2016 who report that during the financial crisis in 2008, the US CDS market was less efficient (in terms of semi-strong market efficiency).

highlight the capability of detecting similar patterns when comparing partial correlation networks based on CDS data in the German banking system and the actual bilateral exposures provided by the Deutsche Bundesbank credit register, suggesting that market based measures can serve well in the absence of bilateral interbank market data.

One of the main drawbacks of Gaussian graphical models, besides the normality assumption, which is hardly satisfied by financial times series, is the fact that the estimated graph is typically dense and difficult to interpret. Regularization methods can then be exploited to detect sparse graph structures for which not all the edges are active and only the relevant links are detected. The *glasso* model, one of the best known approaches for matrix regularization in the statistical literature, induces sparsity in the network structure by introducing a penalty proportional to the 1-norm of the precision matrix in the estimation process (Friedman, Hastie, and Tibshirani, 2008). While *glasso* has found a range of applications in biology and other fields, its use within the financial literature is still limited (e.g., Goto and Xu, 2015; Anufriev and Panchenko, 2015). This is possibly because the assumption of Gaussianity is often too restrictive when dealing with financial times series.

Here, we introduce the *tlasso*, a model that allows to estimate the partial correlation matrix by relying on the more realistic assumption of multivariate t-Student variables. This allows us to better capture the leptokurtic behaviour of financial times series. The model is also more robust to model misspecification and the presence of outliers compared to *glasso*, providing more stable and reliable estimates of the sparse network structure. However, differently from the Gaussian setting, under the t-Student distributional assumption the absence of an edges does not imply conditional independence, but only conditional zero correlation. To the best of our knowledge, besides Torri, Giacometti, and Paterlini, 2019 on portfolio selection and optimization, *tlasso* has never been applied to the estimation of the networks of financial institutions. After illustrating the robustness properties of *tlasso* compared to *glasso* in a simulation study, we focus on estimating the network structure of a sample of large European banks, providing some financial interpretation of the network in two ways. First, we analyse the structural properties of the system using a relevant set of network measures. Second, we propose a decomposition of the *strength centrality* measure to characterize more accurately the role of each bank in the network and to highlight the most relevant channels for the transmission of financial distress. We thus introduce a tool that can be used to evaluate the risk profile of each bank in the system.

Our empirical results suggest a highly connected network structure, characterized by geographical communities and by the absence of a core-periphery structure. The analysis complements previous studies that focus on domestic banking systems (e.g., Anand, Craig, and Von Peter, 2015; Mistrulli, 2011; Lelyveld and Veld, 2014), which are generally characterized by very sparse, strongly tiered network structures with a small set of large banks that play the role of hubs. Other works conducted on samples of international financial institutions, while relying on different modelling techniques, provide results that are in line with ours, although they put less or no emphasis on the community structure. In particular, Craig and Saldías, 2016 found that, considering a network derived from equity prices, the European banking system is characterized by a high level of interconnectivity, high clustering coefficient, and high assortativity. Our results are also in line with Aldasoro and Alves, 2018, which studied the interbank network of large European financial institutions, and found evidence of a high network density and a core composed by a large number of banks.

The Chapter is organized as follows: Section 2 describes the *tlasso* model, after introducing the Gaussian graphical model and *glasso*. Section 3 illustrates the properties of *tlasso* and compares it to *glasso* in a simulation set-up. Section 4 reports the empirical results obtained by analysing a sample of large European banks using CDS data. Estimation results are presented along with network properties, followed by the proposal of the decomposition of *strength centrality* and a discussion of the financial and economic implications of our analysis. Section 5 draws the main conclusions.

3.2 The *tlasso* model

3.2.1 Preliminaries

We briefly review partial correlation networks in the context of Gaussian graphical models and the *glasso* model that, starting from a dense network, identifies a *sparse* partial correlation matrix under the hypothesis of Gaussianity.

Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(m)})$ be a random vector with a multivariate Gaussian distribution $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix. We can define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in $\mathcal{V}$ correspond to each element of $\mathbf{X}$, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations, $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, and

the edge weights correspond to partial correlations ρ_{ij}. The partial correlations are computed from the inverse of the covariance matrix $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$ (i.e., the precision matrix) as follows:

$$\rho_{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \qquad i, j = 1, \ldots, m,$$

where $\{\omega_{ij}\}$ is an element of the matrix $\boldsymbol{\Omega}$. The estimation of the graph structure corresponds to the estimation of the precision matrix $\boldsymbol{\Omega}$. The partial correlation matrix $\boldsymbol{P}$ is:

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \boldsymbol{\Omega} \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}}, \tag{3.1}$$

where $\boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} = diag(\frac{1}{\sqrt{\omega_{ii}}})$ and $\boldsymbol{I}$ is a conformable identity matrix. The representation of partial correlations in the matrix form $\boldsymbol{P}$ is the *weighted adjacency matrix* of the graph, that is, the $m \times m$ matrix in which the entries are the weights of the edges in the network. Notice that, following a common convention, the diagonal elements of $\boldsymbol{P}$ are equal to zero. The reader is referred to Lauritzen, 1996 for Gaussian graphical models, while Anufriev and Panchenko, 2015 provide a review of partial correlation networks in the Gaussian framework.

Traditional estimation techniques for graphical models, such as maximum likelihood, typically return dense adjacency matrices in which all the elements off the main diagonal are different from zero. It may then be challenging to identify the relevant links among nodes. Moreover, the presence of collinearity in the data would result in ill-conditioned estimates of the covariance matrices and, consequently, large estimation error for the precision matrix (Torri, Giacometti, and Paterlini, 2019).

The *glasso* model, introduced by Friedman, Hastie, and Tibshirani, 2008, was proposed as a tool to solve this issue and to obtain *sparse* estimates of the precision matrix $\boldsymbol{\Omega}$. Under the assumption of Gaussianity, the precision matrix can be estimated by penalized quasi maximum likelihood, where the penalty is proportional to the 1-norm of the precision matrix. Omitting constants and multiplicative factors, the estimates obtained with *glasso* can be computed by solving the following optimization problem:

$$\widehat{\boldsymbol{\Omega}}_{glasso} = \arg\max_{\boldsymbol{\Omega}} (log|\boldsymbol{\Omega}| - tr(\boldsymbol{\Omega}\boldsymbol{\Sigma}_S) - \lambda||\boldsymbol{\Omega}||_1), \tag{3.2}$$

where $\boldsymbol{\Sigma}_S$ is the empirical sample covariance matrix, λ is the regularization parameter (the larger the λ, the sparser the estimated model), $|\cdot|$ is the determinant of a matrix, $tr(\cdot)$ the trace and $||\cdot||_1$

the 1-norm. The optimization problem can be solved efficiently using the algorithm proposed by Friedman, Hastie, and Tibshirani, 2008.

3.2.2 The *tlasso* model

One of the main limitations of *glasso* is the Gaussianity assumption that, as widely discussed in the literature, generally does not hold for financial time series, typically characterized by fat tails and leptokurtic distributions (e.g., Cont, 2001). We then consider *tlasso*, a recently introduced model for the inference of sparse partial correlation networks under the assumption that the data follow a multivariate t-Student distribution (Finegold and Drton, 2011). Compared to *glasso*, the distributional assumptions of *tlasso* are less restrictive for financial applications. Furthermore, the use of t-Student distributional assumption as an alternative to Gaussianity is known to increase the robustness of the estimates in presence of outliers in several applications, such as linear and nonlinear regression, robust estimation of the mean and covariance matrix with missing data, unbalanced multivariate repeated-measures and multivariate nonlinear regression Lange, Little, and Taylor, 1989.

Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(m)})$ be an m-variate random vector with a multivariate t-Student distribution $t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v)$, with v degrees of freedom, mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Psi}^{-1}$ ($m \times m$ positive semi-definite matrix[3]). The covariance and precision matrices are then $\boldsymbol{\Sigma} = \frac{v}{v-2}\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\Omega} = \frac{v-2}{v}\boldsymbol{\Psi}$, respectively. Similarly to the Gaussian case, we can associate to the distribution a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, in which $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$, where $\{\rho_{ij}\}$ are the elements of the partial correlation matrix $\boldsymbol{P}$ computed from the precision matrix $\boldsymbol{\Omega}$ as in (3.1). The *tlasso* model allows us to estimate a sparse $\widehat{\boldsymbol{\Psi}}_{tlasso}$ and the corresponding partial correlation matrix $\widehat{\boldsymbol{P}}$ that represents the network structure.

In particular, we consider a scale-mixture representation of the t-Student distribution. Let $\boldsymbol{W} \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi}^{-1})$ and $\tau \sim \Gamma(v/2, v/2)$ be random variables distributed as a multivariate Gaussian and a gamma distribution, respectively. Then:

$$\mathbf{X} = \boldsymbol{\mu} + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v), \tag{3.3}$$

[3] We choose the uncommon parametrization based on the inverse of the diffusion matrix to highlight the relevance of Ψ in the graphical model.

where $\mathbf{\Psi}^{-1}$ is the dispersion matrix of the t-Student distribution, $\boldsymbol{\mu}$ is a $m \times 1$ vector, and v is the number of degrees of freedom in the multivariate t-Student distribution (Kotz and Nadarajah, 2004).[4]

In contrast to the Gaussian set-up, when using the t-Student distribution, or the class of elliptical distributions in general, an absence of correlation does not necessarily correspond to conditional independence (Baba, Shibata, and Sibuya, 2004). However, despite the lack of conditional independence for $\omega_{ij} = 0$, it can be proved that if two nodes j and k are separated by a set of nodes C in $\mathcal{G}$, then $X^{(j)}$ and $X^{(k)}$ are conditionally uncorrelated given $X^{(C)}$ (Finegold and Drton, 2011). It is then reasonable to substitute the conditional independence with zero partial correlation or zero conditional correlation. In this case, disconnected vertices in a graphical model can be considered orthogonal to each other after the effects of other variables are removed. The absence of conditional correlation entails that a mean-square error optimal prediction of variable $X^{(j)}$ can be based on the variables $X^{(k)}$, which correspond to neighbours of the node j in the graph.

The Expectation-Maximization algorithm

The *tlasso* model can be estimated using an Expectation-Maximization (EM) algorithm (Finegold and Drton, 2011). The EM estimation procedure exploits the scale-mixture representation (3.3), and it is computationally efficient since it is based on the *glasso* algorithm, which is iteratively applied at every M-step of the algorithm.

The EM algorithm treats τ as a hidden variable in the E-step, exploiting the fact that the conditional distribution of $\mathbf{X}$ given τ is $\mathcal{N}_m(\mathbf{0}, \mathbf{\Psi}^{-1}/\tau)$. Then, in the M-step the algorithm maximizes the penalized log-likelihood of the latent Gaussian vector using the *glasso* procedure.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an n-sample drawn from $t_m(\boldsymbol{\mu}, \mathbf{\Psi}^{-1}, v)$. The EM algorithm iterates the following two steps:

- **E-step**
  - Given:

$$\mathbb{E}[\tau|\mathbf{X} = \mathbf{x}] = \frac{v + m}{v + (\delta_{\mathbf{x}}(\boldsymbol{\mu}, \mathbf{\Psi}))}, \tag{3.4}$$

[4]Note that the scale-mixture representation clarifies how the use of t-Student distribution leads to more robust inference, as extreme observations can arise from small values of τ.

where $\delta_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Psi}(\mathbf{x} - \boldsymbol{\mu})$, from the current estimates $\widehat{\boldsymbol{\mu}}^{(t)}$ and $\widehat{\boldsymbol{\Psi}}^{(t)}$, we compute $\widehat{\tau}^{(t+1)}$ for the $(t+1)$th iteration:

$$\widehat{\tau}_i^{(t+1)} = \frac{v + m}{v + (\delta_{\mathbf{X}_i}(\widehat{\boldsymbol{\mu}}^{(t)}, \widehat{\boldsymbol{\Psi}}^{(t)}))} \qquad i = 1, \ldots, n. \tag{3.5}$$

- **M-step**
  - Compute the estimates at iteration $t+1$:

$$\widehat{\boldsymbol{\mu}}^{(t+1)} = \frac{\sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)} \mathbf{X}_i}{\sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)}}, \tag{3.6}$$

$$\widehat{\boldsymbol{S}}^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)} [\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}][\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}]'. \tag{3.7}$$

  - Then, the estimate $\widehat{\boldsymbol{\Psi}}^{(t+1)}$ is computed by solving the following optimization problem:

$$\widehat{\boldsymbol{\Psi}}^{(t+1)} = \arg\max_{\boldsymbol{\Psi}} \left(log|\boldsymbol{\Psi}| - tr(\boldsymbol{\Psi}\widehat{\boldsymbol{S}}^{(t+1)}) - \lambda||\boldsymbol{\Psi}||_1 \right), \tag{3.8}$$

  which relates to (3.2) for *glasso* and can be solved using the algorithm from Friedman, Hastie, and Tibshirani, 2008.

The E and M steps are sequentially iterated until a convergence criterion is satisfied, that is, until the maximum term in absolute value of the matrix difference between $\widehat{\boldsymbol{\Psi}}$s in two consecutive iterations is smaller than a given threshold. While convergence to a stationary point is guaranteed in the penalized versions of EM (McLachlan and Krishnan, 2007), the algorithm is not guaranteed to find the global maximum since the penalized log-likelihood function to be maximized is not concave (Finegold and Drton, 2011).

3.2.3 The regularization parameter

The regularization parameter λ controls the intensity of the shrinkage towards zero of individual elements of the precision matrix: the larger λ, the more sparse is the precision matrix estimate. Different approaches, such as cross-validation and information criteria, can be used to select the optimal regularization parameter λ (Finegold and Drton, 2011). We rely on the Bayesian

Information Criterion (BIC) for the calibration of λ, as the literature highlights that it performs better than cross-validation (Foygel and Drton, 2010).

For *tlasso*, we consider a grid of values of λs, that is $\lambda_i \in \mathcal{C} = [\lambda_{min}, \ldots, \lambda_{max}]$ and estimate the *tlasso* network for each λ value. Then, we choose the value of λ that solves the BIC optimization problem:

$$\min_{\lambda \in \mathcal{C}} BIC = -2log(\hat{L}_{\lambda_i}) + k \times log(n), \tag{3.9}$$

where $\hat{L}_{\lambda_i}$ is the value of the likelihood function for the multivariate t-Student distribution with the parameters computed using the corresponding λ_i parameter, k is the number of active edges in the network, and n is the number of observations. The calibration procedure is analogous for *glasso*, with the only difference being the use of the likelihood function for a multivariate Gaussian distribution instead of a multivariate t-Student distribution. In preliminary analyses on simulated data, we also compared the BIC with the Extended Bayesian Information Criterion (EBIC) and the Akaike Information Criterion (AIC), finding that the BIC provides more accurate results than the others.

3.3 Simulation analysis

As a first step, we test the *tlasso* on simulated data using several network configurations and distributions, and we compare its performances to the ones of the *glasso*. We consider 16 settings with different combinations of network structures and underlying distributions of the data (see Section 3.3.1). For each combination, we perform 30 Monte Carlo runs and estimate the networks using *tlasso* and *glasso* on a grid of 30 different λ values ranging from $\lambda_{min} = 0.007$ to $\lambda_{max} = 1$. We set the number of degrees of freedom to $v = 3$, the lowest possible value that yields a finite covariance matrix, to test for the influence of fat tails. We also investigate how *tlasso* and *glasso* estimates are affected by the sample size, comparing estimates obtained for n = 100, 200, 500 and 1000. Finally, we evaluate the performance of the BIC for the optimal choice of the regularization parameter λ.

3.3.1 Simulation set-up

In the simulation set-up, we consider four distributional assumptions and four network configurations, chosen from commonly used benchmark cases and more complex structures that are

typical of financial networks. We then run the simulation analysis for the sixteen combinations of network structures and distributions. We describe the structure of the network in terms of the precision matrix $\mathbf{\Omega}$, which makes it possible to easily derive the partial correlation matrix $\boldsymbol{P}$ as in (3.1).

a. **Chain graph**: This simple graph structure can be represented by a matrix $\mathbf{\Omega}_a$, where all the non-diagonal entries are zero except the elements on the first sub- and super-diagonal. In particular, these entries have been set equal to -0.3, while diagonal elements are set equal to 1.

b. **Sparse random graph**: The matrix $\mathbf{\Omega}_b$ is defined as follows: first a value of -1 is assigned independently with probability 0.05 to the elements in the lower diagonal, then the matrix is mirrored to obtain a symmetric adjacency matrix. A value equal to the negative of the sum of the elements in the respective columns minus one is assigned to the elements on the diagonal. This guarantees that the matrix is positive definite.

c. **Core-periphery matrix:** The matrix $\mathbf{\Omega}_c$ is consistent with a core-periphery structure, that is, it has a group of strongly interconnected nodes (the core) and a periphery connected to nodes within the core, but not among nodes within the periphery. The core is composed of 10 elements and the percentage of the active edges between the core and the periphery has been set to 30%.

d. **Block diagonal matrix**: The matrix $\mathbf{\Omega}_d$ is characterized by a strong block diagonal structure with five distinct blocks. This structure is consistent with a network composed of five disconnected sub-networks. In particular, the blocks are modelled independently and in each of them the time series are generated as a linear combination of two independent factors plus some random noise.

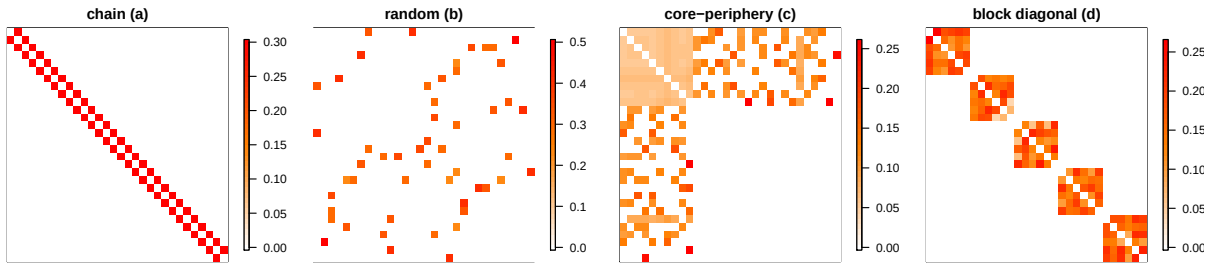
We assume the following four underlying distributions:

1. **Multivariate Gaussian distribution**: $X \sim \mathcal{N}_m(0, \mathbf{\Omega}^{-1})$.

2. **Multivariate t-Student distribution**: $X \sim t_m(0, \dfrac{v-2}{v}\mathbf{\Omega}^{-1}, v), \qquad \text{with } v = 3.$

3. **Contaminated Gaussian distribution**: Gaussian-mixture distribution where 85% of the observation are distributed as a centered Gaussian distribution with covariance matrix $\mathbf{\Omega}^{-1}$,

FIGURE 3.1: Alternative network structures based on partial correlation matrices



The figure represents the color coded heatmaps of the adjacency matrices $\boldsymbol{P}$ corresponding to network structure (a - d). The structures are in order: the chain graph (a), a random graph with high sparsity (b), a core-periphery graph (c) and a block diagonal graph (d).

and the remaining 15% is drawn from independent centered Gaussian distributions with larger variances:

$$X \sim \mathcal{N}_m(0, \boldsymbol{\Omega}^{-1}) \times b + \mathcal{N}_m(0, 2 \times diag(\boldsymbol{\Omega}^{-1})) \times (1 - b),$$

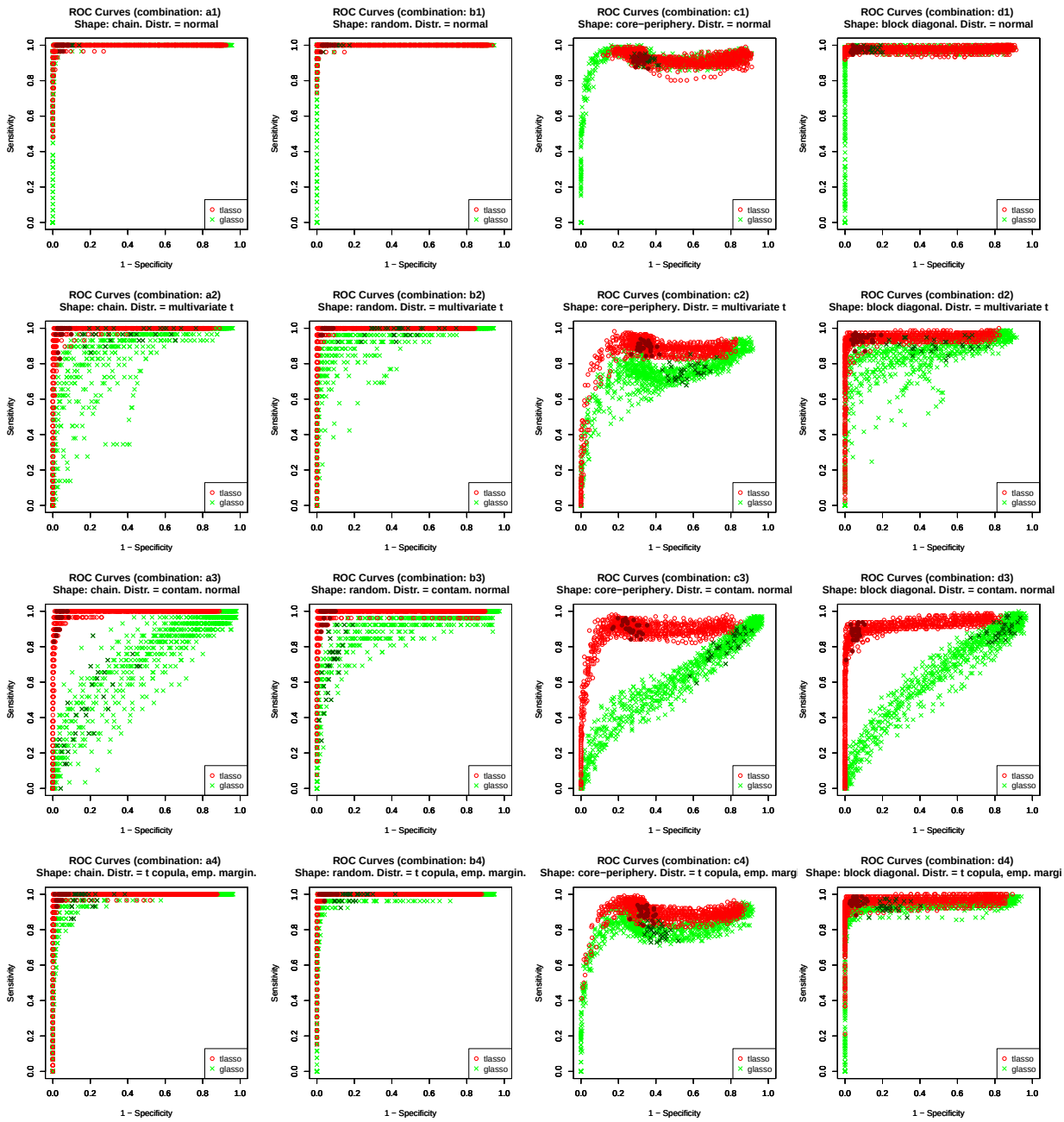
$$\text{with } b \sim Bernoulli(p), \text{and} \quad p = 0.85.$$

4. **Empirical marginals from real CDS data and *t-copula***: Distribution characterized by a *multivariate t-copula* and marginals estimated from real CDS data using a Gaussian kernel. The data are described in Section 3.4 and refer to the period 01/01/2009 – 30/06/2016.

3.3.2 Performance measures

We first evaluate the performances of *tlasso* and *glasso* estimators in terms of ROC (Receiver Operating Characteristics) curves, measuring their ability to identifying correctly the edges in the network. A good estimator should have a high sensitivity (high ratio of true positives to all positives), while having a high specificity (high ratio of true negatives to all negatives), that is, it should identify all the active edges in the original network without introducing spurious ones. ROC curves are a powerful and commonly used tool for the identification of these features. For binary classifiers, they consist in a plot of the sensitivity rate over (1 - specificity) for different thresholds (see for instance Hanley and McNeil, 1982). We use them for the performance analysis of *glasso* and *tlasso*, comparing network estimates with different levels of sparsity as determined for a grid of λ values. In order to provide better representation of the dispersion of

FIGURE 3.2: ROC curves for 16 simulation set-ups

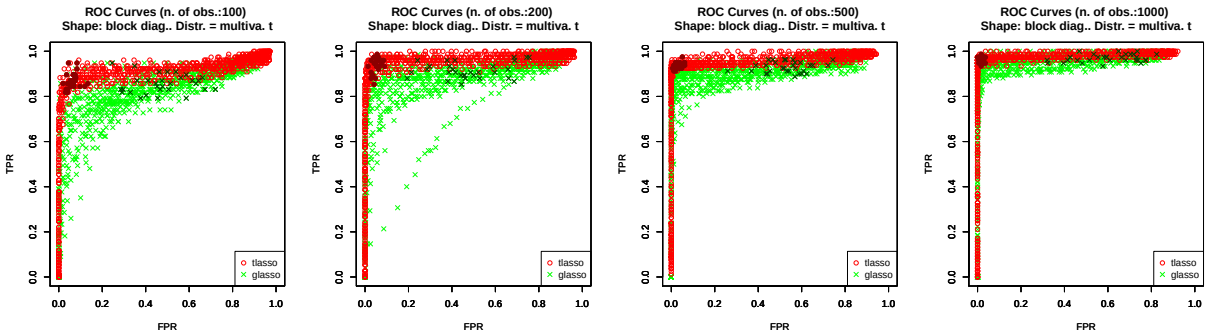


ROC curves for *tlasso* and *glasso* estimation with different network structures and underlying distributions, as defined in Section 3.3.1 (structures (a - d), from left to right; distributions (1 - 4), from top to bottom). Red circles refer to *tlasso* and green crosses to *glasso*. Dark green circles and dark red crosses refer to the optimal estimates identified using the BIC for *tlasso* and *glasso*, respectively.

the data, instead of computing the average value over several runs and drawing an interpolated ROC curve, we represent all the individual data points.

In addition to ROC curves, we evaluate the quality of the estimates of the partial correlation matrix $\boldsymbol{P}$ using the Frobenius norm as distance measure between the optimal estimate of the

FIGURE 3.3: ROC curves for different sample size



ROC curves for *tlasso* and *glasso* estimation with block diagonal network structure (d) and multivariate t-Student distribution with 3 degrees of freedom (2). Green points refers to *glasso* and red points to *tlasso*. Darker points are the optimal estimates identified using the BIC. From left to right the four graphs display results for sample size equal to 100, 200, 500 and 1000.

adjacency matrix $\widehat{\boldsymbol{P}}$ and the known adjacency matrix $\boldsymbol{P}$, as follows:

$$D_F(\widehat{\boldsymbol{P}}, \boldsymbol{P}) = ||\widehat{\boldsymbol{P}} - \boldsymbol{P}||_F = \sqrt{(tr((\widehat{\boldsymbol{P}} - \boldsymbol{P})(\widehat{\boldsymbol{P}} - \boldsymbol{P})'))}, \tag{3.10}$$

where $\widehat{\boldsymbol{P}}$ is the estimate of the partial correlation matrix and $\boldsymbol{P}$ is the one used for the generation of the data. In particular, we use the Frobenius norm to measure the accuracy of the estimation of the optimal networks identified by BIC.

TABLE 3.1: Gini coefficients for ROC curves

	Network structure			
Distribution	Chain (a)	Random (b)	Block diag. (c)	Core-periphery (d)
Panel a - *tlasso*				
Multivariate normal (1)	0.992	0.997	0.702	0.952
Multivariate t (2)	0.979	0.986	0.752	0.959
Contam. normal (3)	0.972	0.984	0.746	0.921
t-copula + emp. marg. (4)	0.984	0.992	0.760	0.930
Panel b - *glasso*				
Multivariate normal (1)	0.986	0.988	0.793	0.943
Multivariate t (2)	0.757	0.873	0.528	0.812
Contam. normal (3)	0.385	0.812	0.179	0.244
t-copula + emp. marg. (4)	0.934	0.965	0.616	0.869

Gini Coefficient for different network structures and underlying distributions.

3.3.3 Simulation results

Figure 3.2 displays the ROC curves for all the four combinations of network configurations (from a to d) and distributions of the data (from 1 to 4), described in Section 3.3.1. Each data point refers to one estimate of the network structure for a different value of λ. Red circles refer to *tlasso* networks, while green crosses indicate *glasso* ones. The optimal networks identified by the BIC are highlighted with darker colours. The more concentrated the data are in the upper left corner, the better the performance, as measured by a high number number of true positives and a low number of false positives. We see that the quality of the estimates varies drastically across different combinations of distributions and network structures. Moving from top to bottom, we analyse the performance of *glasso* and *tlasso* on data with different distributions. In the case of data with a Gaussian distribution (first row), the performances of *glasso* and *tlasso* are similar, though the first model is correctly specified while the second is not. When the distribution is not Gaussian (rows 2, 3 and 4), *tlasso* performs consistently better than *glasso*, especially in the case

TABLE 3.2: Average Frobenius distance for *tlasso*

	Network structure			
Distribution	Chain (a)	Random (b)	Block diag. (c)	Core-periphery (d)
Panel a - *tlasso*				
Multivariate normal (1)	1.22	1.05	1.11	0.71
(s.d.)	(0.10)	(0.12)	(0.06)	(0.04)
Multivariate t (2)	1.29	1.09	1.13	0.79
(s.d.)	(0.22)	(0.08)	(0.06)	(0.08)
Contam. normal (3)	1.51	1.32	1.16	0.87
(s.d.)	(0.15)	(0.12)	(0.06)	(0.06)
t-copula + emp. marg. (4)	1.24	1.11	1.17	0.76
(s.d.)	(0.13)	(0.11)	(0.06)	(0.05)
Panel b - *glasso*				
Multivariate normal (1)	1.15	0.95	1.12	0.73
(s.d.)	(0.11)	(0.09)	(0.05)	(0.05)
Multivariate t (2)	1.88	1.67	1.91	1.72
(s.d.)	(0.31)	(0.28)	(0.30)	(0.40)
Contam. normal (3)	2.21	1.95	3.29	3.45
(s.d.)	(0.09)	(0.15)	(0.35)	(0.37)
t-copula + emp. marg. (4)	1.38	1.28	1.53	1.15
(s.d.)	(0.10)	(0.11)	(0.11)	(0.13)

Average Frobenius distance (and standard deviation in parenthesis) between optimal estimators $\widehat{\boldsymbol{P}}$ and known network structure $\boldsymbol{P}$ for *tlasso* (Panel a) and *glasso* (Panel b) for all network structures and distributions presented in Section 3.3.1. Optimal networks are selected using BIC.

of the distribution characterized by the presence of outliers (row 3).

When studying the effect of the partial correlation structure (Figure 3.2 from left to right), we notice that the topology of the network strongly affects the quality of the estimates provided by the *tlasso* and *glasso*. We see that both models perform better in case of simple structures like the chain graph (a) or a very sparse random graph (b). They have problems, however, in correctly identifying more complex structures such as the core-periphery (c) or the block diagonal configurations (d).

The optimal networks identified by the BIC are highlighted with darker circles and crosses. We see that, in the case of the *tlasso*, the BIC is highly effective for all the specifications considered, as it often selects optimal λs corresponding to good estimates of the true network structure (high sensitivity and specificity, lying in the upper-left corner of the graph). In the case of the *glasso*, the BIC works well for the selection of the optimal parameter λ when the model is correctly specified (row 1), but performs poorly when the distribution is not Gaussian or in presence of outliers (rows 2, 3 and 4).

We integrate the visual analysis of the ROC curves with the computation of the Gini coefficient. In particular, we fit a smoothing cubic spline on the data in order to obtain a unique ROC curve, we then compute the Area Under the Curve (AUC) as in Hanley and McNeil, 1982. We then compute the Gini coefficient as: Gini $= 2 \times$ AUC $- 1$, that ranges from 0 to 1. A Gini coefficient close to 0 denotes a low ability of a model to identify correctly the edges of the network, while a coefficient close to 1 denotes a good ability to identify the edges. The value of the Gini coefficient for each model is shown in Table 3.1, confirming the previous results: *tlasso* outperforms *glasso* in all except one of the simulation settings, and has a stronger advantages in the case of non-Gaussian distributions (in particular the *contaminated normal* case).

Figure 3.3 illustrates as a robustness check how the quality of the estimates changes with the sample size n = 100, 200, 500 and 1000. From all the combinations, we report only the case of a known block diagonal network structure and multivariate t-Student distribution with 3 degrees of freedom (corresponding to combination (d2) in Figure 3.2). Other pairs provide similar results, which are available upon request. As expected, the larger the sample size n, the lower the estimation error, as indicated by the concentration of the points in the upper left corner in Figure 3.3 for $n = 1000$. We see that *tlasso* consistently outperforms *glasso*, which fails to provide accurate estimates, even with large sample size, due to the model misspecification and its lack of robustness.

Finally, Table 3.2 reports the average Frobenius distance of the optimal networks from the true networks for *tlasso* (Panel a) and *glasso* (Panel b). The results confirm those of the ROC curve analysis and show that *tlasso* outperforms *glasso* in most of the specifications, obtaining smaller distances. The difference between the models is particularly relevant for the distribution 3, which is characterized by outliers. The only exceptions are the simulations with normally distributed data, where *glasso* and *tlasso* show similar performance.
In summary, the simulation study shows that *tlasso* generally outperforms *glasso*, in particular under non-normality, model misspecification, and in presence of outliers.

3.4 Empirical analysis

Here, we present the application of the *tlasso* model to the problem of estimating the network structure of a sample of large European banks. Our goal is to infer the network structure from the partial correlations between CDS to capture dependence related to credit risk, and to describe the properties of the system and its evolution over time. We further provide insights on the network estimates by the *tlasso* when compared to those obtained using the *glasso*. Finally, we introduce a decomposition of *strength centrality* that allows us to characterize accurately the role of each bank in the network and to highlight the most relevant channels for the transmission of financial distress.

3.4.1 Banking data

The dataset consists of 31 weekly time series of CDS spreads (5 years maturity, quoted in Euros) of European banks from 12 countries. The data have been downloaded from Thomson Reuters Datastream and span the time period from 01/01/2009 to 30/06/2016. 20 out of 31 banks belong to countries within the Eurozone, while the other 11 are located in the United Kingdom, Sweden and Denmark. Our database includes 85% of those banks with total assets over EUR 500 billion under the ECB's supervision in 2016 and 47% of the banks involved in the European Banking Authority (EBA) stress-test exercise of 2016. The complete list of the banks is reported in Appendix C.2. For the analysis, we consider the log-differences of the CDS spreads.
The *glasso* and *tlasso* models assume that the data follow a multivariate Gaussian and t-Student distribution, respectively. Due to complexity in testing the multivariate goodness of fit in high

dimensionality (see e.g., Justel, Peña, and Zamar, 1997; McAssey, 2013), we do not test directly for the multivariate distributional assumption, but instead we focus on the marginal univariate and bivariate distributions.[5] Results for individual banks are reported in C.2. On average, the marginal distributions show low skewness and high excess kurtosis. We test for two parametric distributions: the t-Student and the Gaussian, fitting them using maximum likelihood. We then evaluate the goodness of fit using the Kolmogorov-Smirnov test and the χ^2 goodness of fit test, both at a significance level of $\alpha = 1\%$. According to the Kolmogorov-Smirnov test, we do not reject the null hypothesis of a t-Student distribution for any of the banks except for one (i.e., Bayerische Landesbank), while the assumption of Gaussianity is rejected for 8 of the 31 banks. The χ^2 goodness of fit test rejects the null hypothesis of a t-Student and Gaussian distribution for 4 and 18 banks, respectively. Concerning the fitting of t-Student marginal distributions, the estimated numbers of degrees of freedom are low, ranging from 1.75 to 8.80 with mean value equal to 5.37. The statistical tests thus provide evidence in favour of t-Student marginal distributions.

Concerning the dependence structure, we analyse 465 bivariate distributions for each pair of banks, fitting different copula families and selecting the best-fitted copula model for each pair using the Bayesian Information Criterion (BIC). The copulas we consider are *Gaussian*, *t* (with 3, 6 and 10 degrees of freedom), *Gumbel*, and *reverse Gumbel*. This selection encompasses models with different levels of tail dependence and allows us to model both symmetric and asymmetric relationships. Table 3.3 reports the percentages of the bivariate distributions for which each copula is the best fit according to the BIC. The vast majority of the tests provide evidence in favour of the *t-copula*. In fact, the *t-copula* with 6 degrees of freedom turns out to be optimal for more than half of the pairs (54.4%), and *t-copulas* with 3, 6 or 10 degrees of freedom are optimal in 93.2% of the cases. The bivariate data therefore appear to display a relatively high level of tail dependence.

The results obtained for the marginal and bivariate distributions suggest the need to consider a methodology that moves beyond the Gaussian distributional and dependence assumption. We therefore expect the *tlasso* model to be a more appropriate tool than the *glasso* due to empirical properties of the data and the *tlasso*âĂŹs robustness to outliers and model misspecification.

[5]Following Kotz and Nadarajah, 2004, any partition of a random variable that follows a multivariate t-Student distribution is also distributed as a t-Student. By testing the distribution of marginals we find a necessary, although not sufficient, condition for the entire distribution to follow a t-Student distribution. Such property holds also for multivariate Gaussian distributions as a special case.

TABLE 3.3: Bivariate copula fitting for pairs of banks.

Copula	Percentage
t-copula (3 degrees of freedom)	10.8%
t-copula (6 degrees of freedom)	54.4%
t-copula (10 degrees of freedom)	28.0%
Gaussian copula	4.7%
Gumbel copula	1.9%
Reverse Gumbel copula	0.2%

Percentage out of the 465 bivariate relationships between each pair of banks' CDS spreads, for which each copula is the best fit according to BIC (01/01/2009 – 30/06/2016).

The analysis of the autocorrelograms of the log-differences of CDS spreads and their squared returns (not reported for brevity) does not highlight relevant evidence of serial correlation and heteroskedasticity. In the rest of the analysis we therefore consider the variables as independent and identically distributed (i.i.d.). As a robustness check, in C.1 we repeat the analysis on the residuals of an ARMA-GARCH model using a static conditional correlation model (CCC-GARCH) and a dynamic conditional correlation one (DCC-GARCH) (see C.1). The results obtained with these specifications are very similar to the baseline scenario.

We finally study the model after controlling for conditioning variables, considering as factors a set of 11 European sovereign CDS spreads (i.e., Austria, Belgium, Denmark, France, Germany, Italy, Netherland, Portugal, Spain, Sweden and United Kingdom), and re-estimating the model after controlling for their effect. Also in this case, the results are consistent to ones obtained with raw data. The results are reported in C.1.

3.4.2 Stability and robustness to outliers

As illustrated in Section 3.3, *tlasso* outperforms *glasso* in the majority of the simulation set-ups we consider. Here, we test the performance of *glasso* and *tlasso* on real data, using them to estimate the sparse correlation network underlying the European banking system.

Since the true network structure is unknown, we cannot use the performance measures described in Section 3.3 for the simulated data. Instead, we compare the *glasso* and *tlasso* focusing on the stability and robustness to outliers. In particular, we consider the stability of the network estimates in a rolling window analysis, computing measures of similarity between estimated networks on different time-windows. The rationale behind this approach is that a robust method

should return very similar network estimates on largely overlapping time-windows and the presence of sudden changes in the network structure would indicate that the model is not robust to outliers and prone to estimation errors.

We consider the period from 01/01/2008 to 30/06/2016, with rolling windows of 100 weekly observations for a total of 343 windows.[6] In order to allow for a more meaningful comparison between time periods, we do not calibrate the penalization parameters λ for *tlasso* and *glasso* in each window, but instead we set the parameter in such a way that the average density of the network across the time windows is equal for the two models. For the estimation of *tlasso* we set the degrees of freedom to 5, close to the mean value of the marginal distributions of the data (see C.2).

We use the Frobenius norm of the difference between the two adjacency matrices, defined as in (3.10) as an indicator for the distance between networks. Figure 3.4 represents graphically the distance between each pair of rolling windows and table 3.4 reports the average distance computed on windows 5, 10 and 15 weeks apart.

The results show that the Frobenius distance between networks tends to increases with the distance in time of the estimation windows, which is in line with our expectations. This can be seen in Figure 3.4, where the points close to the diagonal (which represent windows close to each other) have smaller Frobenius distance. Similarly, we see in Table 3.4 that networks estimated on windows 15 weeks apart from each other differ more than those estimated on windows 10 and 5 weeks apart. Comparing the two models, we notice that, on average, the distance between networks estimated on different windows is smaller for *tlasso* than for *glasso*, resulting in a more stable structure over time for *tlasso*. Interestingly, the *tlasso* networks are not only more stable on average, but they are also characterized by a smoother transition of the network over the rolling windows. This can be observed in Figure 3.4, where *tlasso* has a less block-diagonal structure than *glasso*. Such result is confirmed in Table 3.4, where the value of the standard deviation of the Frobenius distances for *tlasso* is much smaller than for *glasso*. The absence of sudden changes suggests that *tlasso* is less affected by outliers and extreme events, which leads to more reliable and stable estimates.

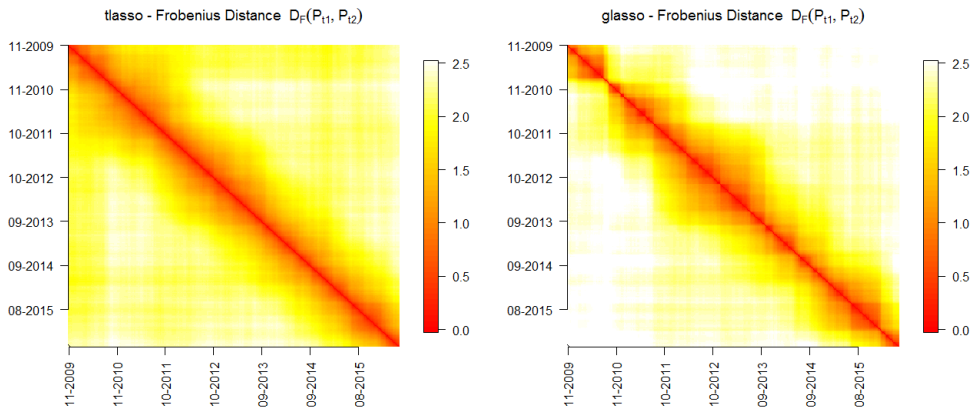
[6]For the rolling analysis the sample is reduced to 29 banks, since CDS series for *Intesa San Paolo* and *Unicredit* are not available for the entire period.

TABLE 3.4: Frobenius distance on rolling window networks

	Frobenius distance	
	tlasso	*glasso*
5 weeks	0.55 (0.06)	0.64 (0.23)
10 weeks	0.78 (0.07)	0.91 (0.26)
15 weeks	0.95 (0.09)	1.11 (0.27)

Average Frobenius distance and standard deviation (in brakets) between networks computed on rolling windows respectively 5, 10 and 15 weeks apart.

FIGURE 3.4: Frobenius distance between networks on different time windows.



Frobenius distance between networks computed on different time windows. Lower distance denotes more similar networks. The horizontal and vertical axes represent the last day in each time window.

3.4.3 Structural analysis of the European banking system network

As a next step, we focus on the identification of the structural properties of the European banking system by computing relevant network indicators and monitoring their changes in two sub-periods: *during-crisis* (01/01/2009 – 31/12/2012) and *post-crisis* (01/01/2013 – 30/06/2016). We focus on the meso-scale properties of the network, such as the presence of a core-periphery or a community structure (see Borgatti and Everett, 2000; Fortunato, 2010).

Most of the available literature analyse the banking systems in individual countries, for which the bilateral exposures in the interbank markets are more easily available (e.g., Craig and Von Peter, 2014 for Germany, Lelyveld and Veld, 2014 for the Netherlands, and Mistrulli, 2011 for Italy). The network structures in these studies are typically characterized by a high level of sparsity, a core-periphery structure, a scale-free configuration (i.e., networks where degree and strength

distributions are power-law), the presence of hubs (i.e., nodes with a number of edges that greatly exceed the average) and *disassortative mixing* patterns (i.e., highly connected banks tend to connect to peripheral ones). The properties of international financial systems are less well-studied, and seem to be characterized by more complex structures. Aldasoro and Alves, 2018 analyse the interbank market bilateral exposures of 53 large European banks and identify a highly interconnected and layered structure with a core-periphery configuration, characterized by a large core. Craig and Saldías, 2016 use equity data to analyse the global banking system and find a complex hierarchical structure, with a relevant regional homophily, a rich-club phenomenon (highly connected nodes tend to be mutually linked) and a core-periphery structure. Here, we provide further evidence in estimating and interpreting dependence networks based on CDS data in international banking systems.

We first look at the descriptive network statistics, reported in Table 3.5. We then compute a set of indicators useful to describe the meso-scale properties of the network (see core-periphery structure, presence of geographical communities and assortative mixing). A description of the network indicators is reported in Section 1.3.1.

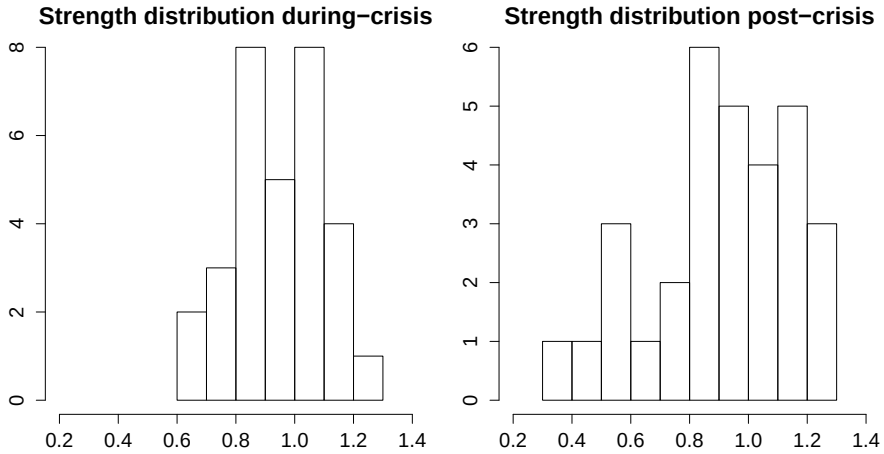
TABLE 3.5: Descriptive network statistics for CDS Networks

	during-crisis	post-crisis
Density	49.25%	52.26%
Average strength	0.95	0.91
Largest connected component	100%	100%
Positive edges %	87.77%	79.42%
Positive edges % (weighted)	96.59%	89.28%

Descriptive network indicators computed on *tlasso networks* for the *during-crisis* and the *post-crisis* periods. See Section 1.3.1 for the definition of network indicators.

In Table 3.5 we notice that for both sub-periods the network is connected (it exists at least a path that connect any two nodes in the network) and by similar values of *density* (49.25% and 52.26%), *average degree* (14.77 and 15.68), and *average strength* (0.95% and 0.91%), suggesting a highly interconnected structure. Edges with positive weights are the majority, especially in the *during-crisis* period and when computed in weighted terms. Figure 3.5 reports the *strength distribution* for the sub-periods. Concerning the *during-crisis* period (left plot), we notice that the support is narrower than in the *post-crisis* period (right plot) and that the distribution is approximately symmetric, suggesting that the networks do not have a scale-free structure and are not characterized by the presence of hubs. In the *post-crisis* period, the *strength distribution*

FIGURE 3.5: Strength distribution of the estimated networks



Strength distribution in the *during-crisis* (left) and *post-crisis* (right) periods.

is negatively skewed, indicating the presence of some banks that are less interconnected to the system compared to the *during-crisis* period.

TABLE 3.6: Network indicators for *tlasso*

	during-crisis		post-crisis	
	tlasso	random network	*tlasso*	random network
Freeman Centralization	0.47**	0.66	0.52*	0.69
Coreness	0.11***	0.34	0.15***	0.31
Modularity (wrt country)	0.36***	-0.04	0.29***	-0.03
Clustering coefficient	2.70%**	2.12%	2.94%***	2.39%
Assortativity (wrt eigen. centr.)	0.61***	0.35	0.35	0.34
Assortativity (wrt str. centr.)	-0.05	0.08	0.06	0.09
Assortativity (wrt betw. centr.)	-0.05	-0.07	-0.08	-0.07
Assortativity (wrt Bonacich centr.)	0.03	0.08	0.11	0.08

Network indicators computed on *tlasso* networks for the *during-crisis* and the *post-crisis* periods. The Table also reports the median value of the indicators computed on 1000 random rewirings of the networks. The stars denote the statistical significance based on the empirical distribution of the indicators on the random networks (***, **, * denote p-value smaller than 1%, 5% and 10%, respectively).

Table 3.6 reports a set of network indicators chosen to highlight the meso-scale structural properties of the networks, focusing in particular on the presence of core-periphery and community structure. As a reference, Table 3.6 also shows the median value of the indicator for a random network and the statistical significance computed on the quantiles of the distribution of the indicators on a random network.[7]

[7]The random networks have been computed by a *rewiring* procedure that consists in randomly selecting two edges (a,b) and (c,d) and substituting them with (a,d), (b,c). The rewiring procedure is repeated iteratively until the network is completely randomized (see for instance Fortunato, 2010). This procedure allows us to destroy existing structural properties while maintaining the *degree* of the nodes and other features such as the *average strength*. The

The *Freeman centralization* provides information about the extent to which the network is centred around its most central node. A high centralization is typically associated with a tiered structure or with the presence of strong hubs. In the *during-crisis* period, this indicator is statistically significantly lower than in the random networks in both sub-periods, suggesting the absence of a *core-periphery* structure. This is confirmed by the low *coreness*,[8] equal to 0.11 in the *during-crisis* period and to 0.15 in the *post-crisis* period.

The following two indicators, *modularity* and the *clustering coefficient*, allow us to measure the tendency of the networks to be organized in communities. In particular, the *modularity* computed on the country partition describes the tendency of the networks to form communities aligned with national borders. The indicator has a high and statistically significant value, especially in the *during-crisis* periods, where it takes a value of 0.36. The presence of communities is also suggested by the high value of the *clustering coefficient*,[9] that is typically associated to networks with communities (Fagiolo, 2007).

Finally, we compute the *assortativity*, an indicator that measures the tendency of nodes to connect to similar ones. In particular, we consider the *assortativity* with respect to four different centrality measures (*eigenvector-*, *strength-*, *betweenness-* and *Bonacich power centrality*) to check the tendency of nodes to be linked to nodes with the same level of importance. Negative assortativity is typical of network with hubs, and such systems are typically resilient to random failures, but fragile to specific attack to hubs Watts and Strogatz, 1998. Positive assortativity, instead, is typical of social networks. Overall, the network does not present neither an assortative or disassortative behaviour. The only exception is the *eigenvector centrality* in the *during-crisis* period, where the indicator is statistically significantly higher than in a random network. The absence of disassortative mixing may seem in contrast with the literature, that shows disassortativity in financial networks (see e.g., Mistrulli, 2011; Hurd, 2016). This may be explained considering that we included in our analysis only the largest European banks, discarding the smallest, and arguably less interconnected institutions. The result may also be related to the presence of a community structure. In fact, Newman and Park, 2003 shows that communities can generate a higher than expected *assortative mixing* since nodes in larger and denser groups have a higher

significance level is obtained from the quantiles of the empirical distribution of the indicators computed on 1000 randomly rewired networks.

[8] *Coreness* is an indicator that denotes the tendency of a network to have a strongly connected core and a sparsely connected periphery (Borgatti and Everett, 2000).

[9] The *clustering coefficient* is an indicator that denotes the tendency of the network to "create triangles", that is, when a node a is connected to nodes b and c, there is a high chance that b and c are connected.

number of connections relative to nodes than in smaller communities. This would lead to highly interconnected nodes being connected to other highly connected nodes, increasing the *assortativity*.

Figure 3.6 shows a visual representation of the networks using a force layout color-coded by country. The representation confirms the structural properties suggested by the indicators: the banking system is characterized by a high interconnection and a relevant community structure aligned with national division, especially in the *during-crisis* period.

The comparison with existing literature is complicated by the differences in methodology and sample selection. As we mentioned before, the structure identified here differs significantly from that identified in studies that focus on actual bilateral exposures in national banking systems. They describe much sparser networks, scale-free distributions for the degree and strength, core-periphery configurations and disassortative mixing (Mistrulli, 2011). Such divergence can be explained by the fact that in our analysis we include only large banks that are part of the *core* of the network in each individual country. The comparison with studies focused on international banking networks shows more consistent results. Aldasoro and Alves, 2018, using the algorithm from Anand, Craig, and Von Peter, 2015 on a dataset of bilateral exposure between large European banks, find a large core composed by twenty to thirty banks, depending on the type of interconnections considered and Craig and Von Peter, 2014 underline the presence of a strongly interconnected structure in Europe.

To our knowledge, our work is the first to put emphasis on the presence of geographical communities in the European banking system. Aldasoro and Alves, 2018 did not provide indicators to measure this feature and Craig and Von Peter, 2014 only describe qualitatively the presence of regional homophily and the higher density of the network structure within countries. From an economic point of view, the identification of national communities is consistent with the different national legal and economic systems, and possibly related to the European sovereign debt crisis and CDS pricing.

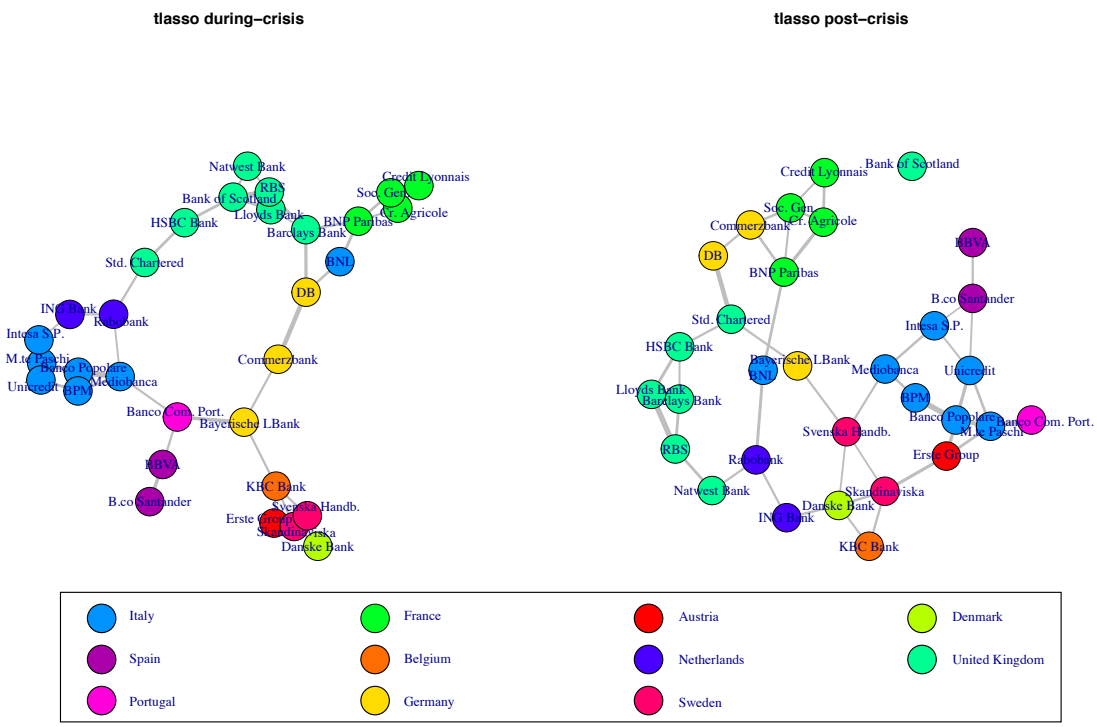
It is well known in the complex network literature that, in general, structural and topological properties influence the diffusion of processes on a network and strongly affect the vulnerability to failures and attacks (Watts and Strogatz, 1998; Newman, 2002; Fortunato, 2010). Concerning banking systems, the structural properties have clear implications in terms of diffusion of financial contagion and the assessment of the network properties is useful for both financial institutions and policy makers. The literature agrees that the presence of a highly interconnected

structure in the interbank market increases the stability of the system, but that in presence of liquidity shocks, it makes the system less stable, leading to a *robust-yet-fragile* property. Moreover, the relationship between structural properties and stability is complex and non linear (Hurd, 2016; Chinazzi and Fagiolo, 2015). Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015, for instance, shows that more densely connected financial networks enhance financial stability but, beyond a certain point, dense interconnections lead to a less stable financial system. Hurd, 2016 shows that the presence of *assortative mixing* facilitates the diffusion of contagion in the system, and Hübsch and Walther, 2017 show how the presence of inhomogeneities among banks can enhance contagion risk. The modelization of the diffusion of financial contagion in networks partitioned in communities is still an open issue in the literature. Applications in other fields show that the presence of communities influences the diffusion of processes over a network (see Weng, Menczer, and Ahn, 2013; Galstyan and Cohen, 2007). We thus expect that this feature may be relevant also in financial applications. Finally, we underline that the presence of communities, similarly to other structural network properties, may have an influence on the interpretation of centrality indicators and raises new questions on how to measure the importance of a node in a system. In the next section, we try to address these questions by introducing a decomposition of *strength centrality*.

3.4.4 Centrality measures for networks with communities

In the previous section, we analysed the structural properties of the estimated networks of the European banking system using the *tlasso* model. Here, we shift the focus to the analysis of individual institutions, studying their interconnectedness and their systemic relevance. Following the definition reported by Schwarcz, 2008, systemic risk is closely related to the propagation and reverberation of risk in a system, and in the literature it has been typically measured either by econometric indicators or by a network approach. Concerning the econometric approach, several studies developed indicators that measure risk trasmission and spillover to the system and from the system, such as *marginal expected shortfall* (Engle and Brownlees, 2010), *expected capital shortfall* (Acharya, Engle, and Richardson, 2012), *CoVaR* and $\Delta CoVaR$ (Adrian and Brunnermeier, 2016). Concerning to network approaches, they typically aim at assessing the relevance of an institution in a system using traditional centrality measures such as *eigenvector-*, *strength-* and *betweeness centrality* (see for instance Alter, Craig, and Raupach, 2015), or by

FIGURE 3.6: *tlasso* networks - graphical representation



Graphical representations of the *tlasso* networks using a force layout. For clarity, only the edges corresponding to a partial correlation larger than 0.1 in absolute value are depicted.

introducing specific measures such as *debtrank* by Battiston et al., 2012a.

A common feature of most of the measures in the literature, either econometric or network-based, is that they summarize in a single measurement the relationship between a given institution and the system, but often do not allow to characterize the relationship between an individual bank's systemic risk and the topological features of the entire network. We introduce a methodology that allows to take explicitly into consideration the network topology, in particular the presence of geographical communities while assessing the role of each individual institution. To clarify the concept, we can think of a network with a geographical community structure. In such network, a bank may be characterized by high centrality because it has strong international interconnections or because it is strongly connected to banks in the same country. We can imagine that these two situations can be very different for regulators who want to measure the systemic importance of a bank or understand the level of EU integration. Our proposed method provides a flexible tool for dealing with this situation by introducing a decomposition of *strength centrality*. This measure, when computed on partial correlation networks, has an interesting interpretation in

terms of shock diffusion, that we discuss below.

We first recall the interpretation of the partial correlation matrix in terms of *hedge regressions*, that are a set of regressions in which each asset is regressed against all others (see Stevens, 1998). Assuming that $\mathbb{E}[\mathbf{X}] = \mathbf{0}$ (where $\mathbf{0}$ is a conformable vector of zeros), the hedge regressions are expressed as:

$$X_t^{(i)} = \sum_{j \neq i} \beta_j^{(i)} X_t^{(i)} + \varepsilon_t^{(i)} \qquad i = 1, \ldots, m, \tag{3.11}$$

where $X^{(i)}$ is the ith asset, $\beta_j^{(i)}$ the coefficient for the jth asset in the ith regression and $\varepsilon^{(i)}$ the residuals. Or in matrix form

$$\mathbf{X} = \boldsymbol{B}\mathbf{X} + \boldsymbol{\varepsilon}, \tag{3.12}$$

where $\boldsymbol{B}$ is a matrix with off-diagonal elements ij equal to $\beta_j^{(i)}$ and diagonal elements equal to 0.

We can relate the elements ω_{ij} of $\boldsymbol{\Omega}$, the precision matrix of the data, to the βs as

$$\omega_{ij} = \begin{cases} -\dfrac{\beta_j^{(i)}}{\nu_i} & \text{if } i \neq j \\ \dfrac{1}{\nu_i} & \text{if } i = j \end{cases}, \tag{3.13}$$

where ν_i is the variance of $\varepsilon^{(i)}$. Using (3.12), (3.13), and (3.1), after some algebra we have

$$\mathbf{X} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \boldsymbol{P} \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}} \mathbf{X} + \boldsymbol{\varepsilon}. \tag{3.14}$$

In order to account for the different conditional variances in the $X^{(i)}$s, we introduce the rescaled variable $\mathbf{x} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}} \mathbf{X}$, in such a way that the rescaled residuals $\mathbf{e} = \boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}} \boldsymbol{\varepsilon}$ have unit variance. By premultiplying the terms in (3.14) by $\boldsymbol{D}_{\boldsymbol{\Omega}}^{\frac{1}{2}}$ we have

$$\mathbf{x} = \boldsymbol{P}\mathbf{x} + \mathbf{e}. \tag{3.15}$$

The rescaling allows to remove the effects of variables endogenous to the network by expressing the extent of the shock in terms of the conditional variances (see Anufriev and Panchenko, 2015). Considering the regression interpretation presented here, we now use the network structure estimated by *tlasso* to study the systemic risk and financial contagion. As we mentioned before, the setting described so far does not allow to identify direct causation between variables, since

it considers only synchronous co-movements. Similarly to Anufriev and Panchenko, 2015, we then consider an observational interpretation of (3.15), where the residuals $\mathbf{e}$ can be seen as an exogenous shock, and the partial correlation network edges as the channels of propagation: a shock $\mathbf{e}$ hitting a set of nodes will have an effect on the other nodes in virtue of (3.15) and the effect on the direct neighbours of the hit nodes is computed as $\boldsymbol{P}\mathbf{e}$ (*first round effect*).

We can then compute the effect on the entire system in terms of a unitary shock hitting the ith node in terms of *first round effect* as

$$c_i^S = \sum_{i \neq j} \alpha_j \rho_{ij} = \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i, \tag{3.16}$$

where $\mathbf{e}_i$ is a unitary shock hitting node i and $\boldsymbol{\alpha}$ is a vector denoting the weight of each node in the system. Notice that in case of $\boldsymbol{\alpha} = \mathbf{1}$, where $\mathbf{1}$ is a conformable vector of ones, the quantity c_i^S consists in the *strength centrality* of node i. For the symmetry of $\boldsymbol{P}$, the network is not directed and the quantities can also be interpreted as the effect of a unitary shock hitting all the nodes on the node i.

As (3.16) is an homogeneous functions of degree 1 with respect to $\boldsymbol{\alpha}$, we can then use the following Euler decomposition:

$$c_i^S = \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i = \sum_{j=1}^{m} \boldsymbol{\alpha}_j \frac{\partial \boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i}{\partial \boldsymbol{\alpha}_j} = \boldsymbol{\alpha}' \nabla (\boldsymbol{\alpha}' \boldsymbol{P} \mathbf{e}_i), \tag{3.17}$$

where ∇ denotes the gradient with respect to vector $\boldsymbol{\alpha}$. We emphasize that this framework accounts only for the immediate transmission of shock. Once the shock is transmitted to the banks connected to the first one, it could be propagated further, generating a cascade effect. We could account for these propagation effects by considering different *rounds* of transmission of the shock, similarly to Anufriev and Panchenko, 2015 or Battiston et al., 2012a. Here, we consider only the first *round* of contagion transmission, and focus on the short term effects. We leave the modeling of successive steps for further research, also considering that the structural properties of the network may change after a shock hits due to reactions of banks and interventions from banking authorities.[10]

[10]The effect of the shock after all the reverberations can be measured as the *Bonacich power centrality* (Anufriev and Panchenko, 2015), and it can be decomposed as in (3.17). However, due to the reverberations, the shocks diffuse more uniformly to the entire networks, making the decomposition less informative. The results are not reported for brevity, and are available upon request.

These components can then be aggregated according to meaningful attributes of the data, in our case the geographical location of the banks.

We apply the decomposition of *strength centrality* to the networks estimated using *tlasso*. In order to make the results easier to read, and to allow for a better comparison between banks, we perform the strength decomposition according to three broad geographical groups and not on individual countries. The regions we consider are *Southern European* countries (i.e., Italy, Spain and Portugal), *Central European* countries (i.e., Germany, France, Austria, Belgium and Netherlands) and the countries *outside the Eurozone* (i.e., United Kingdom, Sweden and Denmark).[11]

Table 3.7 shows the average decomposition of the strength in each geographical area. On average, the largest part of the *strength centrality* of a bank can be attributed to connections with banks in the same geographical region. This confirms the presence of a community structure aligned to geographical divisions, especially in *Southern Europe* and for the group of countries outside the Eurozone, for which the edges within the same group in the crisis-period account respectively for 69.19% and the 68.83% of the total *strength centrality*. Instead, the banks in *Central European countries* seem to be characterized on average by a slightly less regional interconnectedness, as for them edges directed to banks in the same area account on average for 57.98% of the *strength centrality*. The result would be even more pronounced if we exclude the French banks, that are tighly connected to each other. The decomposition yields similar results for the two periods, denoting stability in the structure of European banking network. Still, in the *post-crisis* period the banks show on average a higher international connectivity, quantified by a lower share of *strength centrality* attributable to banks in the same geographical group, especially for the banks outside the Eurozone.

Figure 3.7 provides the relative decomposition of *strength centrality* for individual banks (top panels), the value of *strength centality* (middle panels) and *eigenvector centrality* (bottom panels). The decomposition allows us to highlight the specific position of each bank in the system, and to describe more accurately its properties. We see for instance that Italian and French banks are characterized by a localized interconnectivity, while German, Danish and Dutch banks have on average stronger connections to banks in other geographical areas. The representation allows also to easily identify outliers and banks with specific features, such as the Italian bank *BNL*, whose connections are mostly towards *Central European* banks. This is not surprising, as BNL

[11]The decomposition of *strength centrality* for individual countries gives results consistent with those reported below. They are available from the authors upon request.

is part of the French *BNP Paribas* group. The identification of the most internationalized banks can be used to detect potential *bridges* for the diffusion of contagion between distant areas. Such banks may then be monitored more accurately by regulators. We also show how the decomposition of the *strength centrality* can be used to highlight specificities of each node related to the structural properties of the network. For instance, focusing on *RBS* (Royal Bank of Scotland), we notice that in the *during-crisis* period, it is one of the institutes with the highest *eigenvector centrality*, but its interconnectivity structure spans mostly a limited geographical area. According to the estimated network structure, a potential shock hitting RBS would therefore spread initially to a localized neighbourhood, while a shock on a bank with a wider interconnectivity structure may potentially affect international institutions faster.

TABLE 3.7: Decomposition of *strength centrality* - average % of strength directed to each geographical area

	During-crisis			Post-crisis		
	Southern	Central	Outside	Southern	Central	Outside
Southern	69.19 %	23.42 %	7.84 %	67.20 %	21.17 %	12.06 %
Central	21.05 %	57.98 %	21.82 %	20.19 %	56.52 %	27.85 %
Outside	7.60 %	22.21 %	68.83 %	10.85 %	23.44 %	60.75 %

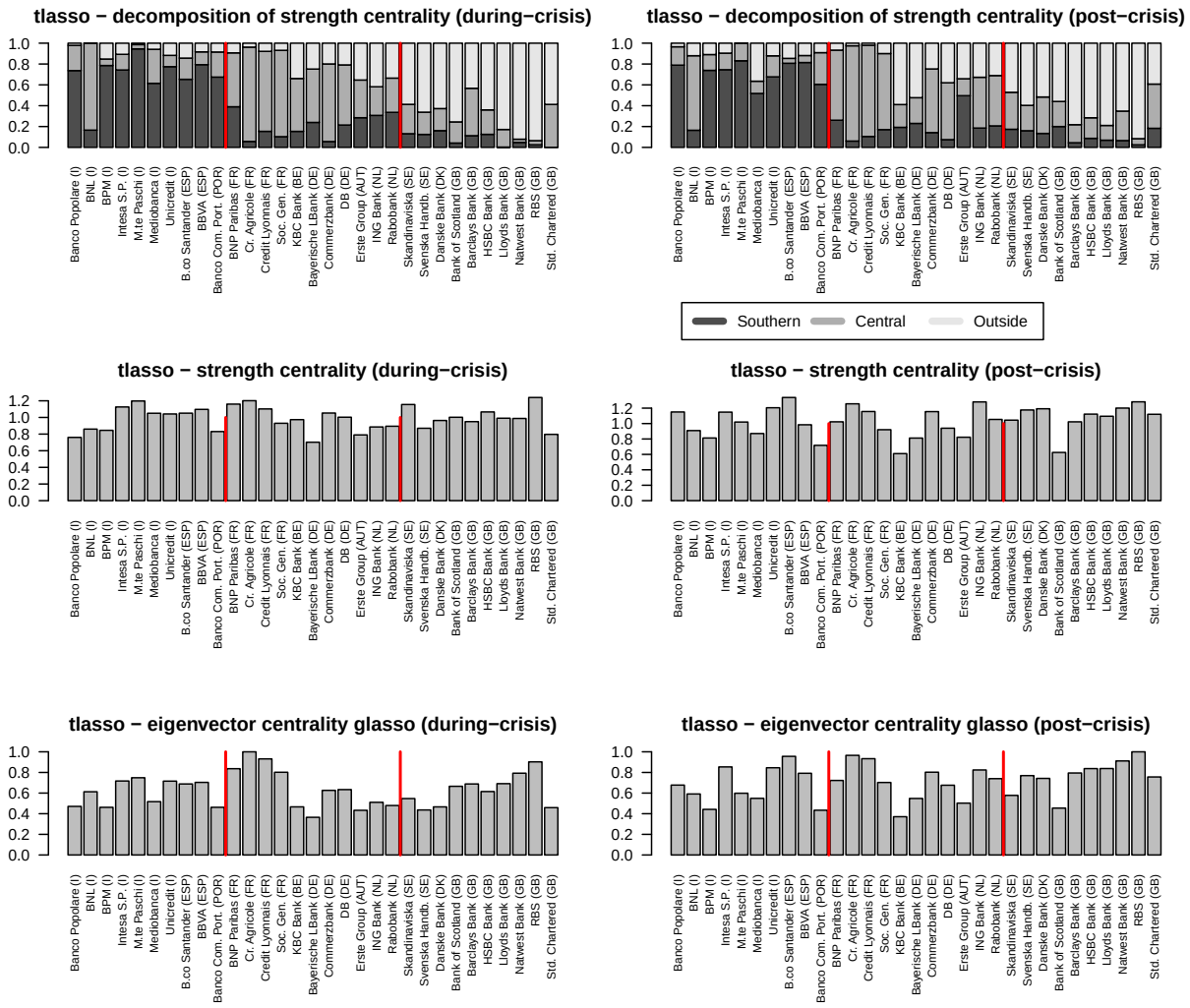
Decomposition of strength centrality. The Table reports the average value of the strength directed to each geographical area. *Southern*: Southern European countries (i.e., Italy, Spain and Portugal). *Central*: Central European countries (i.e., Germany, France, Austria, Belgium and Netherlands). *Outside*: Countries outside the Eurozone (i.e., United Kingdom, Sweden and Denmark).

We point out that the interaction between the structural properties of the network and the local centrality measures assumes a particular relevance for the development of macroprudential regulation and supervision toolkits, that aim to protect the financial system as a whole, in contrast to microprudential regulation, which focuses on the stability of individual institutions, as discussed by Freixas, Laeven, and Peydró, 2015.

In this context, the *strength centrality* decomposition that we propose can be used to complement other network based centrality measures in order to define a new multi-dimensional approach for the identification of systemic relevance, in a spirit similar to the G-SIB framework, an indicator-based assessment methodology introduced in November 2011 by the Financial Stability Board as a response to the 2008 global financial crisis (FSB, 2013).

The question whether a bank with more localized interconnections poses a smaller or greater threat to the system in terms of contagion risk is still an open issue in the literature and should be further studied. On the one hand, a localized system may slow down the diffusion of distress,

FIGURE 3.7: Decomposition of *strength centrality* and centrality measures.



Geographical decomposition of *strength centrality* (upper panels) and value of *strength centrality* (middle panels) and *eigenvector centrality* (lower panels). The three geographical areas (*Southern Europe*, *Central Europe* and *outside Eurozone*) are highlighted with red vertical bands. *Southern*: Southern European countries (i.e., Italy, Spain and Portugal). *Central*: Central European countries (i.e., Germany, France, Austria, Belgium and Netherlands). *Outside*: Countries outside the Eurozone (i.e., United Kingdom, Sweden and Denmark).

allowing the contagion to dissipate before it reaches all the nodes in the network. On the other hand, the concentration of distress in a limited group of banks may reduce the system's ability to dissipate shocks and may trigger feedback effects that would amplify the initial shock. Still, the *strength centrality* decomposition that we propose can provide a useful tool for the characterization of the contagion channels for each bank. The study of contagion mechanisms in networks with geographical communities is high on our agenda.

3.5 Conclusion

In this work, we use partial correlation networks to estimate the structure of the European banking system from CDS data, introducing the *tlasso* model in the financial network literature. Such model estimates efficiently sparse partial correlation networks under the assumption of a multivariate t-Student distribution, inducing sparsity by means of a 1-norm matrix penalization. We show by a simulation analysis and on real CDS data that the *tlasso* is a suitable tool for the estimation of the European banking system network, as it performs better than the alternative *glasso* model when the data are non Gaussian and it is robust to model misspecification and to the presence of outliers.

We then study the topological properties of the financial networks estimated using *tlasso*, focusing on the meso-scale properties. We find evidence of a highly interconnected network characterized by a relevant community structure based on geographical divisions, especially during the crisis period. The network is also characterized by the absence of a core-periphery structure and limited evidence of *assortative mixing*. The results contribute to the growing literature that analyses the network topology of the European banking systems and is consistent with previous works (Craig and Saldías, 2016; Aldasoro and Alves, 2018). Instead, as expected, there are relevant differences compared to studies that focus on national banking systems, where banking networks are typically characterized by the presence of hubs and core-periphery structures.

Finally, we propose a novel decomposition of the *strength centrality* measure, that can be used to assess the contagion risk exposure of financial institutions and to account for the topological features of the network. We use the decomposition to analyse the centrality of financial institutions in presence of geographical communities, and we observe that banks from *Southern Europe* and from *Outside the Eurozone* on average have a more geographically localized interconnectivity structure, while *Central European* banks have on average a more internationalized set of interconnections. The presence of geographical communities emerges both during and after the crisis, and appears to be more prominent in the crisis period.

Our study is, to our knowledge, one of the first to put such emphasis on the presence of a geographical community structure, to quantify the geographical centrality of each bank in an international banking system. Further research will focus on the analysis of the mechanisms of contagion diffusion in a network with communities, and the implications in terms of macroprudential supervision policies.

Chapter 4

Network-ΔCoVaR – Parametric and non-parametric conditional tail risk estimation

Acknowledgement

This chapter is partially based on the article "Network tail risk estimation in the European banking system" co-authored by Rosella Giacometti and Tomáš Tichý, submitted for publication to an academic journal. (Torri, Giacometti, and Paterlini, 2019).

4.1 Introduction

Since the global financial crisis in 2008, the relevance of systemic risk became clear to scholars, regulators and investors, and several approaches have been defined to measure and regulate it. One of the main challenges in the analysis of this phenomenon is the lack of a common definition, that leads to very different modelization approaches, each one focused on different aspects of systemic risk.

Some works interpret systemic risk in terms of potential for the spreading of financial distress, by gauging this increase in tail comovement. These works focus on the analysis of tail risk under stress scenario, and, treating the system as a portfolio of assets, measure either the effect of a systemic shock to the value of an institution, or the effect of the distress of an institution to the entire system. The most known approaches are probably CoVaR and ΔCoVaR, two measures based on the tails of bivariate distributions introduced by Adrian and Brunnermeier, 2016. In particular,

CoVaR measures the Value at Risk (VaR) of the system conditional to a particular asset being distressed, and ΔCoVaR compares the CoVaR to its VaR in a non distressed situation. Similar indicators are the Marginal Expected Shortfall (MES) that measures the marginal constribution of each asset to the expected shortfall of the system (Acharya, Engle, and Richardson, 2012), and the SRISK, that provides a prediction of the level of capital shortfall of an institution based on their MES, leverage and regulatory capital requirements (Brownlees and Engle, 2016).

Another strand of literature focuses more on the interdependence among assets, using a network approach, embracing a vision of systemic risk more focused on the interdependence between institutions, and on the presence of risk spillover and contagion (see e.g. the definition of systemic risk given by Schwarcz, 2008). These approaches, that model institutions as nodes in a network, and their relations as edges, can then uncover structural features of the system that may not emerge from aggregated data. A main challenge is the estimation of the network, that can be modelled in several ways, either considering physical measures of interconnectedness, such as interbank bilateral lendings (Iori et al., 2015), or with statistical measures based on time series, such as Granger causality (Billio et al., 2012), transfer entropy (Bekiros et al., 2017), partial correlations (Torri, Giacometti, and Paterlini, 2018) or empirical tail dependence measures (Poon, Rockinger, and Tawn, 2003).

In this work we merge these two approaches (network indicators and CoVaR) by studying systemic risk from a network perspective, extending some of the results developed in the CoVaR framework to a multivariate setting, studying the theoretical properties of these networks and estimating them in an empirical application focused on the European market. We underline that our methodology aims at providing regulators a more complete view on systemic risk in order to implement more effective policies, complementary to other approaches.

The extension of ΔCoVaR relies on the concept of quantile graphical model, recently introduced in the statistical literature, that models the quantiles of a variable in a system conditional to the value of the other variables (see Belloni, Chen, and Chernozhukov, 2016; Ali, Kolter, and Tibshirani, 2016; Chun et al., 2016). These models are typically estimated using quantile regression and provide a rich and flexible modelization of a multivariate system. Based on these, we introduce here network-ΔCoVaR (also denoted in matrix form as $\mathbf{\Delta CoVaR}$) and we show how such model can also be considered an extension of partial correlation networks, in which we consider conditional quantiles instead of the conditional means. Then, using a parametric specification that assumes a multivariate t-Student distribution of the data, we shed some light on the systemic

risk transmission channels. In particular, we see that the distress of some assets is transmitted to the system primarily in terms of conditional value at risk and standard deviation (instead of conditional mean), even in a homoskedastic setting. In contrast to this, in a Gaussian setting this "contagion" effect that amplifies the initial distress is not present, and the transmission of distress is limited to movements in conditional means. The results presented here, derived under the t-Student hypothesis, are consistent with the idea proposed in several papers, according to which during financial crises tail risk and volatility tend to spread across financial institutions (see e.g. Adrian and Brunnermeier, 2016; Brownlees and Engle, 2016; Engle and Susmel, 1993; Diebold and Yılmaz, 2014), and we provide a simple explanation of this transmission mechanism.

Concerning real world applications, t-Student distribution presents some limitations, as it implicitly assumes the same level of tail dependence among each couple of assets due to the presence of a single value for the degrees of freedom. We therefore rely on a quantile regression approach as Belloni, Chen, and Chernozhukov, 2016 and Hautsch, Schaumburg, and Schienle, 2014. Differently from previous literature, we consider an estimation procedure that allows us to focus on the lower tail of the conditioning distribution by using a custom calibration set. Furthermore, we differentiate from the prominent quantile graphical model literature by using a different technique to induce sparsity in the network: typically it is obtained using an L_1 penalty (lasso), while we consider the SCAD penalty (see Fan and Li, 2001; Wu and Liu, 2009).

We finally propose an empirical application focused on the European banking system, in which we study the network properties of the system, we estimate centrality measures aimed at identifying systemic relevance and systemic fragility, and we propose a methodology that integrates network-ΔCoVaR and the non-performing loans (NPL) ratio, a commonly used balance sheet indicator of credit risk.

Our work contributes to the literature in several aspects: first, it extends to a network dimension of CoVaR and ΔCoVaR, discussing their properties, highlighting the relation with partial correlation network and addressing their estimation in both parametric and non-parametric frameworks. Then, concerning the estimation of the quantile graphical model, we introduce the use of SCAD-penalized quantile regression in alternative to the more common lasso penalized regression, and we propose an estimation procedure focused on the tail of the conditioning assets. Finally empirical application highlights some structural features of the European banking system, especially the role of conditional tail risk as a channel of transmission of financial distress.

The chapter is structured as follows: Section 4.2 discusses the tail-based systemic risk and graphical models, Section 4.3 defines network-CoVaR and network-ΔCoVaR, studying their properties under two parametric distributions, Section 4.4 describes the estimation procedure of ΔCoVaR, Section 4.5 describes the xnetwork indicators used to study the structure of the networks, Section 4.6 presents an empirical application to the European financial system and Section 4.7 concludes.

4.2 Market based systemic risk measurement

4.2.1 CoVaR and ΔCoVaR

The Value at Risk (VaR) is a commonly used risk indicator that expresses the potential loss of a position in a given time period with a level of confidence $1 - \tau$ (typically 95% or 99%). It is implicitly defined as the τ-quantile:

$$Pr\{X \leq \text{VaR}_\tau^X\} = \tau. \tag{4.1}$$

After the global financial crisis in 2008, it become clear that the risk measures based on univariate distributions were not sufficient to assess the risks in the markets, as they failed to consider the interaction among assets, in particular ther joint tail risk. New measures based on bivariate tails have then been developed. Among these, one of the most popular is CoVaR, introduced by Adrian and Brunnermeier, 2016.[1] It is a measure that considers potential spillover risks to the system and allows to capture the contribution to systemic risk of a financial institution. CoVaR measures what happens to the system's VaR when one institution is under stress.
After defining the quantile $\tau = Pr\{X_i \leq \text{VaR}_\tau^{X_i}\}$, we denote CoVaR as a value such that

$$Pr\{X_{sys} \leq \text{CoVaR}_\tau^{sys|X_i} | X_i = \text{VaR}_\tau^{X_i}\} = \tau, \tag{4.2}$$

where X_{sys} is the return of the entire system, and X_i the returns of the ith institution.
A high CoVaR is not necessarily a sign of strong influence of asset i on the tail risk of the system, as it may simply depend by an overall high VaR of the system. It is therefore necessary to compare the CoVaR to a reference value, typically the VaR of the system when the ith institution

[1] The measure has been introduced first in 2008, and then revised and discussed several times over the years. In this work we refer to the version presented in Adrian and Brunnermeier, 2016.

is in a *normal* state (i.e. its median value). This leads to the construction of the ΔCoVaR:

$$\Delta\text{CoVaR}_{\tau}^{sys|X_i} = \text{CoVaR}_{\tau}^{sys|X_i} - \text{CoVaR}_{50\%}^{sys|X_i}. \tag{4.3}$$

ΔCoVaR is a popular measure of systemic risk. The estimation is commonly performed using quantile regression, due to the simple implementation and the relatively little assumptions made on the distribution of the data (see Adrian and Brunnermeier, 2016, for more details). Under the quantile regression framework, ΔCoVaR can be expressed as follows:

$$\Delta\text{CoVaR}_{\tau}^{sys|X_i=\text{VaR}_{\tau}^{X_i}} = \beta_{\tau}^{sys|X_i}(\text{VaR}_{\tau}^{X_i} - \text{VaR}_{50\%}^{X_i}), \tag{4.4}$$

where $\beta_{\tau}^{sys|X_i}$ is the τ-quantile regression parameter of X_{sys} over X_i. Alternative approaches for the estimation of CoVaR and ΔCoVaR include the use of copulas (see Mainik and Schaanning, 2014) or bivariate GARCH models (Girardi and Ergün, 2013). Such models typically rely on parametric assumptions on the distribution of the data, and on the computation of the quantiles of the conditional distribution. We will discuss in greater detail the computation of CoVaR under the assumption of multivariate t-Student distribution later.

In order to avoid to isolate the conditioning effect from the one due to the variability of the asset, it is common to consider an alternative, standardized, specification of the indicator (see e.g. Girardi and Ergün, 2013):

$$\Delta\text{CoVaR}^{\%}{}_{\tau}^{sys|X_i} = \left(\text{CoVaR}_{\tau}^{sys|X_i} - \text{CoVaR}_{50\%}^{sys|X_i}\right)/\text{CoVaR}_{50\%}^{sys|X_i}. \tag{4.5}$$

Finally, we underline how the idea of CoVaR can be extended to other risk measures. One example is the conditional Expected shortfall (CoES), that is computed as the expected shortfall of the market conditional to the fact that one asset is in distress, for which it has been recently proposed an estimation methodology based on super quantile regression (Huang and Uryasev, 2018).

Issues with the ΔCoVaR framework

Although very popular in the literature, CoVaR and ΔCoVaR may present some issues as systemic risk indicators. First, as highlighted by Benoit et al., 2017, ΔCoVaR can be interpreted

as a measure of *systematic risk* rather than *systemic*: in the Gaussian framework the ΔCoVaR measure is a multiple of the traditional regression beta multipled the VaR of the conditioning institution (see Equation 4.4). As a consequence, ΔCoVaR fails to consider the contagion component of systemic risk. Moreover, since in this setting ΔCoVaR is simply a proxy for systemic beta (weighted by the VaR), it would be more logic to estimate directly the betas using either standard OLS regression, or more robust methods such as regression under t-Student distribution (Lange, Little, and Taylor, 1989). In this case, the use of ΔCoVaR is only a source of estimation error.

Another issue with CoVaR is that by using the VaR(X_i) as a conditioning variable, it does not consider the most extreme stress events. Girardi and Ergün, 2013 introduced an alternative specification that uses a different conditioning set, considering the cases where the conditioning asset is *lower than* its VaR. An additional issue is that CoVaR and ΔCoVaR are not monotonically increasing with the level of interdependence. Indeed, as shown by Mainik and Schaanning, 2014, even in Gaussian settings, a stronger correlation may result in lower ΔCoVaR.

A more subtle argument regards the shape of the bivariate distribution, and the characteristics of the lines of conditional quantile functions. The properties of the quantile functions have surprising effects on the calculation of CoVaR and ΔCoVaR. We can divide these effects in two parts: the first is related to the estimation procedure, and in particular to the linearization of the conditional quantile functions estimated with quantile regression. As an example, concerning the t-Student distribution, the conditional quantile functions for $\tau < 0.5$ are concave in X_i, while the functions estimated using linear quantile regressions are parallel lines (these results are shown in greater detail in Appendix D.1 for the multivariate case, and the bivariate setting is a subcase). In this case the estimation method masks the fact that the conditional quantiles are not linear, leading to systematic underestimation of ΔCoVaR. The problem could be addressed by considering non-parametric quantile regression techniques, using smoothing splines such as in Ali, Kolter, and Tibshirani, 2016, or smoothing kernels (e.g. Bondell, Reich, and Wang, 2010). Such approches however increase the number of parameters to estimate, requiring more data points, with a relevant impact on the computational burden.

A second issue is related to the slope of the quantile functions in some distributions characterized by tail dependence, for instance with Clayton copulas. In such cases CoVaR may be unrelated to the level of tail dependence. Intuitively, if two variables have a strong tail dependence, the regression lines for different quantiles will be very close to each other in the lower tails of the

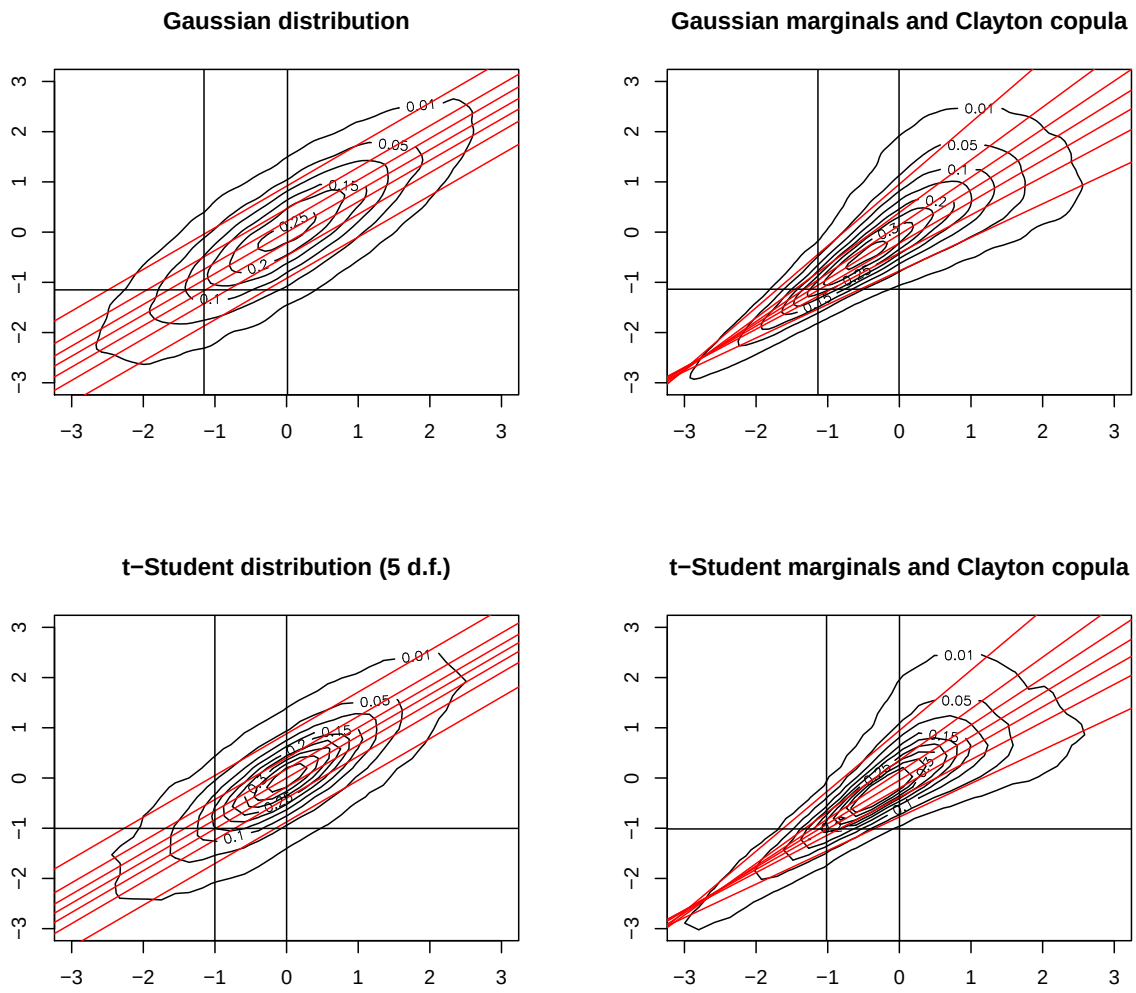


FIGURE 4.1: Simulation (50,000 datapoints) of different bivariate distributions with linear conditional quantiles highlighted ($\tau \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$). Each variable is characterized by different distribution. Note that in the Clayton copula cases, the lower quantiles are less steep than the median regression line, while in elliptical distributions (Gaussian and t-Student) they are parallel.

distribution, and will spread out when moving right and up. As a consequence, the regression lines for the lower quantiles will be less steep (i.e. the β_q^{sys,X_i} will be smaller). In Figure 4.1 we represent the linear conditional quantile functions estimated using quantile regression for four different distributions (see caption for details). We see that higher tail dependence (as in the case of the Clayton copula) may have flatter conditional quantile lines and, consequently, lower ΔCoVaR for a given level of correlation. Indeed, from Table 4.1, we see that for the same level of correlation between variables, the ΔCoVaR is higher for the Gaussian and t-Student settings, in which the tail dependence is smaller compared to the setting characterized by a Clayton copula.

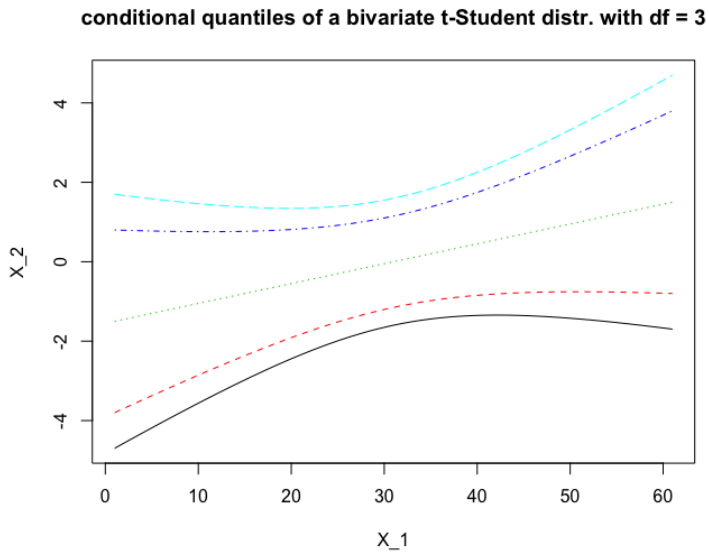


FIGURE 4.2: Conditional quantiles of a bivariate t-Student distribution with degrees of freedom equal to 3. Note that the slope of the quantiles for $\tau < 0.5$ is higher than the best linear approximation (parallel to the β_{mean}) in the region $X_1 < \mathbb{E}[X_1]$.

Finally, the ability of conditional tail measures to predict actual risk of financial institutions has been questioned by empirical analyses. For instance Idier, Lamé, and Mésonnier, 2014 show that Marginal Expected Shortfall (a measure similar in spirit to CoVaR) has smaller predictive power of future losses in periods of crisis, compared to balance sheet metrics. Moreover, Benoit et al., 2017 show that under certain assumptions ΔCoVaR is a multiple of the conditioning asset's VaR, making the forecasting of ΔCoVaR equivalent to the forecasting of the VaR of the asset in isolation.

Model	CoVaR	ΔCoVaR	ρ	$\sigma^2_{X_1}$	$\sigma^2_{X_2}$
Gaussian	2.29	1.38	0.83	1.01	1.00
Gaussian marg., Clayton copula	1.85	1.08	0.83	0.98	0.98
t-Student	2.17	1.30	0.83	0.97	0.99
t-Student marg., Clayton copula	1.82	1.05	0.80	1.01	1.00

TABLE 4.1: Systemic risk measures for different bivariate distributions. The table reports CoVaR, ΔCoVaR (using the quantile $\tau = 0.05$), correlation coefficient (ρ) and the variances of both assets ($\sigma^2_{X_1}$ and $\sigma^2_{X_2}$). The parameter for the Clayton copula is equal to 4, and the dispersion parameters in the Gaussian and t-Student distributions are set in such a way that the correlation is equal for all the settings (except for the t-Student + Clayton copula, where it is slightly smaller).

4.2.2 Systemic risk and network models

Network models are increasingly popular for the analysis of systemic risk, as they allow to consider the component of contagion across companies or financial institutions. In particular, several works study the systemic relevance of institutions based on measures of centrality on the network (e.g. Puliga, Caldarelli, and Battiston, 2014; Alter, Craig, and Raupach, 2015). In general, we can model a network $\mathcal{G}$ as a set of nodes and a set of edges: $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. We can also represent conveniently the network in terms of its adjacency matrix A, that is a $p \times p$ matrix, where p is the number of nodes, and in which non-zero elements ij of the matrix are the edges connecting nodes i and j, and the value is the corresponding weight. Concerning the estimation of the network structure, the literature describes several approaches. A relevant strand of literature focuses on measures of co-movement of time series (see e.g. Billio et al., 2012; Puliga, Caldarelli, and Battiston, 2014; Poon, Rockinger, and Tawn, 2003; Torri, Giacometti, and Paterlini, 2018). Here we discuss in particular the approach of partial correlation networks.

4.2.3 Partial correlation networks

Partial correlation networks are increasingly common in the analysis of systemic risk, as they allow to model the interactions among nodes in a system while controlling for the influence of other nodes (Torri, Giacometti, and Paterlini, 2018). The edge ij of the network is defined as the partial correlation between nodes i and j, conditional to all the other variables in the system. We can represent them using the weighted adjacency matrix P, that can be derived from the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, (i.e. the inverse of the covariance matrix) as $\{\boldsymbol{P}\}_{ij} = \rho_{ij} = -\dfrac{\omega_{ij}}{\omega_{ii}\omega_{jj}}$ for $i \neq j$, and $\rho_{ij} = 0$ for $i = j$, where ω_{ij} is an element of $\boldsymbol{\Omega}$. In matrix form we can write:

$$\boldsymbol{P} = \boldsymbol{I} - D_{\boldsymbol{\Omega}}^{-1/2} \boldsymbol{\Omega} D_{\boldsymbol{\Omega}}^{-1/2}, \tag{4.6}$$

where $D_{\boldsymbol{\Omega}}$ is a matrix with the diagonal elements of $\boldsymbol{\Omega}$, and $\boldsymbol{I}$ is a conformable identity matrix.[2] Commonly used partial correlation networks are Gaussian graphical models that, under the assumption of jointly normally distributed data, imply that two nodes not connected by an edge are

[2]We follow the convention of setting the diagonal elements of $\boldsymbol{P}$ equal to zero in line with Anufriev and Panchenko, 2015, as it provides a more convenient network representation. Alternatively, we could have set the diagonal elements equal to -1.

conditionally independent (Højsgaard, Edwards, and Lauritzen, 2012; Murphy, 2012). Concerning the estimation, in order to overcome issues related to the high dimensionality of the datasets and to improve the interpretability of results, it is common to induce sparsity using penalization techniques. In the Gaussian setting we can rely on the *glasso* model, that induces sparsity in the partial correlation graph by penalizating the L_1 norm of the precision matrix in a quasi maximum likelihood procedure (Friedman, Hastie, and Tibshirani, 2008). Other approaches include folded concave penalties, a class that have better asymptotic properties, such as Smoothly-Clipped Absolute Deviation (SCAD) (Fan and Li, 2001) and the Minimax Convex Penalty (MCP) (Zhang et al., 2010). An alternative is represented by two-steps techniques, such as the adaptive lasso (Zou, 2006).

However, the Gaussian assumption is unrealistic in many fields. In particular, financial time series tend to have fat tails and other well known characteristics that limit the potential application of *glasso* model, leading to estimation problems. Finegold and Drton, 2011 propose an extension of the *glasso* model, the so called *tlasso*, that allows to deal with data characterized by fat tails by relying on a multivariate t-Student distribution. Furthermore, the model shows empirically to be more robust than *glasso* in presence of model misspecification or outliers. Two financial application of *tlasso* can be found in Torri, Giacometti, and Paterlini, 2019 and Torri, Giacometti, and Paterlini, 2018, focusing on equity portfolios and banking system credit networks, respectively. Differently from the Gaussian case, zero partial correlation among two nodes does not implies conditional independence, but only implies that two nodes are partially uncorrelated, as shown by Finegold and Drton, 2011. In particular, as we will see later for the multivariate setting in Section 4.3.2, conditional variances is not constant, and conditional value at risk is not an affine function of the conditioning asset.

We underline that partial correlation networks, by modeling the conditional means of the variables, do not allow to consider conditional tail risk or conditional volatility, limiting their ability to assess systemic risk.

4.3 Network-ΔCoVaR

We propose here the extension of the $\Delta\mathbf{CoVaR}$ measure to the network case, and we discuss its implementation considering a quantile regression framework.

4.3.1 Network-ΔCoVaR and quantile graphical models

As described in Section 4.2.1, CoVaR is defined as the quantile of the distribution of an asset conditional to the distress of another asset. In a multivariate setting, we can extend the concept by modeling the quantile of an asset as a function of all the others, and assessing the marginal effect of distress in another node. We can link this approach to the framework of Quantile Graphical Models (QGM), recently introduced in the statistical literature to describe the dependency structure of multivariate random variables. Several specifications have been discussed, in particular Belloni, Chen, and Chernozhukov, 2016; Ali, Kolter, and Tibshirani, 2016; Chun et al., 2016.

In general, we can express the quantiles of the distribution of an item conditional to the others as

$$Q_{\mathbf{X}_i,\tau} = f(\mathbf{X}_{\backslash i}, \tau), \tag{4.7}$$

where the function f is a generic function. The identification and estimation of the function $f(\mathbf{X}_{\backslash i}, \tau)$ is a crucial step, and several approaches can be chosen. In theory, the function could be determined in a fully non-parametric form but, due to curse of dimensionality, some simplifying assumptions have to be made in order to allow estimation in finite samples. A common choice is to define an additive form such as $f(\mathbf{X}_{\backslash i}, \tau) = \sum_{j \neq i} f_j(\mathbf{X}_j, \tau)$, where $f_j(\mathbf{X}_j, \tau)$ are smooth functions (see e.g. Ali, Kolter, and Tibshirani, 2016). A further simplification is to assume that $f_j(\mathbf{X}_j, \tau)$ are linear functions (Belloni, Chen, and Chernozhukov, 2016); in such case the function f can be expressed in terms of linear quantile regression coefficients: $f(\mathbf{X}_{\backslash i}, \tau) = \alpha_i + \boldsymbol{\beta}_\tau^{i|\cdot} \mathbf{X}_{\backslash i}$, where α_i is a constant, and $\boldsymbol{\beta}_\tau^{i|\cdot}$ is a $1 \times (p-1)$ vector. Such coefficients can be conveniently represented by the matrix $\boldsymbol{B}_\tau$ such that:

$$\{\boldsymbol{B}_\tau\}_{ij} = \begin{cases} \beta_\tau^{i|j} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}. \tag{4.8}$$

In this way we can express the conditional quantile as:

$$f(\mathbf{X}_{\backslash i}, \tau) = \alpha_i + \{\boldsymbol{B}_\tau\}_{i,\cdot} \mathbf{X}. \tag{4.9}$$

An alternative approach is to consider parametric specifications for the distribution of the underlying process, such as a multivariate t-Student or a Gaussian. In this way the conditional quantile function $f(\mathbf{X}_{\setminus i}, \tau)$ can be modeled parametrically and computed exactly.
Using the definition, we can express CoVaR in terms of the conditional quantile function: under the bivariate case we have that $\text{CoVaR}_\tau^{sys|X_i} = f(\text{VaR}_\tau^{X_i}, \tau)$. If the quantile function is assumed to be linear we obtain the original CoVaR formulation proposed by Adrian and Brunnermeier, 2016, while relaxing the linearity assumption we obtain alternative CoVaR specifications, for instance the QL-CoVaR proposed by Bonaccolto, Caporin, and Paterlini, 2018.
Analogously to the bivariate setting, we can extend the approach to the multivariate setting by considering a p-variate distribution of returns. We measure the quantile function of asset i conditional of asset j in distress, and the other assets in the system in their normal (i.e. median) state. More formally we have:

$$Pr\{X_i \leq \text{CoVaR}_\tau^{X_i|X_j} | X_j = \text{VaR}_\tau^{X_j}, X_{\setminus\{i,j\}} = \text{VaR}_{50\%}^{X_{\setminus\{i,j\}}}\} = \tau, \tag{4.10}$$

where $\text{VaR}_{50\%}^{\setminus\{ij\}}$ is the $p-2$ vector of the VaRs of all the assets except ith and jth.[3]
It follows the definition:

Definition 1.

$$\textit{CoVaR}_\tau^{X_i|X_j} := \textit{VaR}_\tau \left(X_i \middle| X_j = \textit{VaR}_\tau(X_j), X_{\setminus\{ij\}} = \textit{VaR}_{50\%}^{\setminus\{i,j\}} \right). \tag{4.11}$$

Once computed the network-CoVaR, we compute network-ΔCoVaR analogously to the bivariate case:[4]

Definition 2.

$$\Delta\textit{CoVaR}_\tau^{X_i|X_j} = \textit{CoVaR}_\tau^{X_i|X_j} - \textit{CoVaR}_{50\%}^{X_i|X_j}. \tag{4.12}$$

The set of all bilateral $\Delta\text{CoVaR}_\tau^{X_i|X_j}$ constitute the edges of a directed and weighted network, that can be represented as the square adjacency matrix $\mathbf{\Delta CoVaR}_\tau$.

[3]To avoid notation clutter we use the notation $\text{CoVaR}_\tau^{X_i|X_j}$ similar to the standard bivariate CoVaR, while a more accurate description would be $\text{CoVaR}_\tau^{X_i|X_j=\text{VaR}_\tau^{X_j}, X_{\setminus\{i,j\}}=\text{VaR}_{50\%}^{X_{\setminus\{i,j\}}}}$.

[4]With a slight abuse of notation we define $\text{CoVaR}_{50\%}^{X_i|X_j}$ as the $\text{VaR}_\tau^{X_i}$ when the asset X_j is in the median state.

If $f(\mathbf{X}_{\backslash i,\tau})$ is an affine function, we can express $\mathbf{\Delta CoVaR}_\tau$ as:

$$\mathbf{\Delta CoVaR}_\tau = \boldsymbol{B}_\tau(\text{VaR}_\tau^{\mathbf{X}} - \text{VaR}_{50\%}^{\mathbf{X}}), \tag{4.13}$$

where $\text{VaR}_\tau^{\mathbf{X}}$ is the $p \times 1$ vector of the VaRs of each variable in $\mathbf{X}$.

4.3.2 Parametric examples

We consider QGM and $\mathbf{\Delta CoVaR}$ under two parametric approaches (Gaussian and t-Student), and we discuss their properties under these distributional assumptions.

Parametric QGM and network-$\mathbf{\Delta CoVaR}$ – Gaussian distribution

In this Section we show how in the Gaussian setting network-ΔCoVaR in strictly related to partial correlation networks, representing a simple standardization of these. Indeed, we show how to express $\mathbf{\Delta CoVaR}$ as a function of the partial correlation matrix and the VaR of the conditioning assets.

Let $\mathbf{X} \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$ be a p-variate Gaussian distribution, and consider the following partition:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\backslash i} \\ \mathbf{X}_i \end{bmatrix} \sim \mathcal{N}_p \left(\begin{bmatrix} \mu_{\backslash i} \\ \mu_i \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{\backslash i,\backslash i} \mathbf{\Sigma}_{\backslash i,i} \\ \mathbf{\Sigma}_{i,\backslash i} \mathbf{\Sigma}_{i,i} \end{bmatrix} \right), \tag{4.14}$$

where i denote the ith variable, and $\backslash i$ the set of all the variables except i. The conditional distribution of $\mathbf{X}_i|\mathbf{X}_{\backslash i}$ can be expressed analytically, and is itself Gaussian (see e.g. Tong, 2012). To simplify the exposition, we consider the case with $\mu = \mathbf{0}$ and we have:

$$\mathbf{X}_i|\mathbf{X}_{\backslash i} \sim \mathcal{N}(\mu_{i|\backslash i}, \mathbf{\Sigma}_{i,i|\backslash i}), \tag{4.15}$$

$$\mu_{i|\backslash i} = \mathbf{\Sigma}'_{\backslash i,i} \mathbf{\Sigma}^{-1}_{\backslash i,\backslash i} \mathbf{X}_{\backslash i},$$

$$\mathbf{\Sigma}_{i,i|\backslash i} = \mathbf{\Sigma}_{i,i} - \mathbf{\Sigma}'_{\backslash i,i} \mathbf{\Sigma}^{-1}_{\backslash i,\backslash i} \mathbf{\Sigma}_{\backslash i,i}.$$

We can then use (4.15) to study the quantiles of the conditional distributions of i as a function of $\mathbf{X}_{\backslash i}$:

$$\begin{aligned} Q_\tau(X_i|X_{\backslash i},\tau) &= \Phi^{-1}(\tau)\boldsymbol{\Sigma}^{1/2}_{i,i|\backslash i} + \mu_{i|\backslash i} \\ &= \Phi^{-1}(\tau)(\boldsymbol{\Sigma}_{i,i} - \boldsymbol{\Sigma}'_{\backslash i,i}\boldsymbol{\Sigma}^{-1}_{\backslash i,\backslash i}\boldsymbol{\Sigma}_{\backslash i,i})^{1/2} + \boldsymbol{\Sigma}'_{\backslash i,i}\boldsymbol{\Sigma}^{-1}_{\backslash i,\backslash i}\mathbf{X}_{\backslash i}, \end{aligned} \tag{4.16}$$

where $\Phi^{-1}(\tau)$ is the inverse of the cumulative density function of a univariate standard Gaussian distribution. The conditional quantile is then an affine function in $\mathbf{X}_{\backslash i}$, and we see that the quantile level τ does not influence the slope of the quantile function. It follows that the conditional quantile functions are parallel. We also notice that the quantity $\boldsymbol{\Sigma}'_{\backslash i,i}\boldsymbol{\Sigma}^{-1}_{\backslash i,\backslash i}$ corresponds to the multiple OLS regression $\boldsymbol{\beta}^{i|\backslash i}_{\text{mean}}$.[5] (4.16) can then be expressed as:

$$Q_\tau(X_i|X_{\backslash i},\tau) = \Phi^{-1}(\tau)\boldsymbol{\Sigma}^{1/2}_{i,i|\backslash i} + \boldsymbol{\beta}^{i|\backslash i}_{\text{mean}}\mathbf{X}_{\backslash i}. \tag{4.17}$$

It follows that all the conditional quantile functions are affine, and parallel to the $\boldsymbol{\beta}^{i|\backslash i}_{\text{mean}}$. This result is analogous to the one presented by Adrian and Brunnermeier, 2016, (Section C) for the bivariate case.

As a consequence of the linearity of the conditional quantile function, the marginal effect of any individual variable in $\mathbf{X}_{\backslash i}$ on the conditional quantile of $\mathbf{X}_i$ does not depend on the level of other variables. It is therefore possible to decompose the effect of the system on the conditional quantile function linearly as $f(\mathbf{X}_{\backslash i},\tau) = \sum_{j\neq i} f_j(\mathbf{X}_j,\tau)$.

Focusing on two assets i and j, we can use (4.16) to describe the network-CoVaR$^{X_i|X_j}$ and network-ΔCoVaR$^{X_i|X_j}$ in terms of the β_{mean} coefficients:[6]

$$\begin{aligned} \text{CoVaR}^{X_i|X_j}_\tau &= \Phi^{-1}(\tau)\boldsymbol{\Sigma}^{1/2}_{i,i|\backslash i} + \left\{\boldsymbol{\Sigma}'_{\backslash i,i}\boldsymbol{\Sigma}^{-1}_{\backslash i,\backslash i}\right\}_j \text{VaR}^{X_j}_\tau \\ &= \Phi^{-1}(\tau)\boldsymbol{\Sigma}^{1/2}_{i,i|\backslash i} + \beta^{i|j}_{\text{mean}}\text{VaR}^{X_j}_\tau, \end{aligned} \tag{4.18}$$

[5] We use the notation $\boldsymbol{\beta}^{i|\backslash i}_{\text{mean}}$ to indicate ordinary least square estimates (instead of $\hat{\boldsymbol{\beta}}^{i|\backslash i}_{\text{OLS}}$) for clarity and greater generality, as in empirical analysis it is common to consider more robust estimators, especially when considering datasets with outliers, distribution mis-specifications or large dimensionality. We also underline that the vector $\boldsymbol{\beta}^{i|\backslash i}_{\text{mean}} = \boldsymbol{\Sigma}'_{\backslash i,i}\boldsymbol{\Sigma}^{-1}_{\backslash i,\backslash i}$ is a row vector and $\mathbf{X}$ is a column vector, while typically in the econometric literature, the opposite convention applies.

[6] we refer to the vector of βs in the multivariate regression of i against all the other variables as $\boldsymbol{\beta}^{i|\backslash i}_{\text{mean}}$, and its jth element as $\beta^{i|j}_{\text{mean}}$.

where $\beta^{i|j}_{\text{mean}}$ is the coefficient for the jth node in the regression of node i on all the others, and $\mathbf{\Sigma}_{i|\setminus i}$ is defined as in (4.15).

From (4.18) we then compute network-ΔCoVaR:

$$\begin{aligned}\Delta\text{CoVaR}^{X_i|X_j} &= \text{CoVaR}^{X_i|X_j}_{\tau} - \text{CoVaR}^{X_i|X_j}_{50\%} \\ &= \beta^{i|j}_{\text{mean}}\left(\text{VaR}^{X_j}_{\tau} - \text{VaR}^{X_j}_{50\%}\right) \\ &= \beta^{i|j}_{\text{mean}}\text{VaR}^{X_j}_{\tau},\end{aligned} \tag{4.19}$$

where in the last passages we used the fact that under a centered Gaussian distribution $\text{VaR}_{50\%}(X_j) = \mathbb{E}(X_j) = 0$.

We also know that VaR_{τ} in a centered Gaussian distribution is multiple of the standard deviation: $\text{VaR}^{X}_{\tau} = \Phi^{-1}(\tau)\sigma(X)$. It follows that we can rewrite (4.19) in terms of the standard deviation of the conditioning asset j as:

$$\Delta\text{CoVaR}^{X_i|X_j} = \beta^{i|j}_{\text{mean}}\Phi^{-1}(\tau)\sigma(X_j), \tag{4.20}$$

where $\sigma(X_j)$ is the (unconditional) standard deviation of asset X_j.

Using the representation in terms of $\beta^{i|j}_{\text{mean}}$ We can also re-write (4.19) and (4.20) in matrix notation, denoting the network structure as the following weighted adjacency matrix:

$$\mathbf{\Delta CoVaR}_{\tau} = \Phi^{-1}(\tau)\boldsymbol{B}\boldsymbol{D}^{1/2}_{\mathbf{\Sigma}}, \tag{4.21}$$

where $\boldsymbol{D}_{\mathbf{\Sigma}}$ is a diagonal matrix with the diagonal elements of the covariance matrix $\mathbf{\Sigma}$ and $\boldsymbol{B}$ is a matrix of the coefficients defined as in (4.8). Note that, as we highlighted before, under the Gaussian hypothesis the mean regression coefficients β_{mean} are equal to the τ-quantile regression β_{τ}.

Finally, we can express $\mathbf{\Delta CoVaR}_{\tau}$ in terms of the partial correlation matrix. The variance $\mathbf{\Sigma}_{i,i|\setminus i}$ in the conditional distribution of $\mathbf{X}_i$, is the variance of the residuals in the regression of $\mathbf{X}_i$ against the other variables $\setminus i$ (known as hedge regression, see Stevens, 1998; Goto and Xu, 2015). This quantity is closely related to the precision matrix. Indeed, as shown by Anufriev and Panchenko, 2015, $\mathbf{\Sigma}_{i|\setminus i}$ is equal to $1/\omega_{ii}$, where ω_{ii} is a diagonal element of the precision matrix

$\mathbf{\Omega} := \mathbf{\Sigma}^{-1}$. We establish therefore the relation between the conditional distribution of $\mathbf{X}_{i|\setminus i}$ and the elements of precision matrix $\mathbf{\Omega}$ as follows:

$$\omega_{ij} = \begin{cases} -\dfrac{\beta^{i|j}_{\text{mean}}}{\mathbf{\Sigma}_{i,i|\setminus i}} & i \neq j \\ \dfrac{1}{\mathbf{\Sigma}_{i,i|\setminus i}} & i = j \end{cases}, \tag{4.22}$$

that in matrix form is:

$$\mathbf{\Omega} = \boldsymbol{D}_{\mathbf{\Omega}}(\boldsymbol{I} - \boldsymbol{B}), \tag{4.23}$$

or equivalently

$$\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{D}_{\mathbf{\Omega}}^{-1}\mathbf{\Omega}. \tag{4.24}$$

We can then represent the network-ΔCoVaR in terms of precision matrix $\mathbf{\Omega}$. In particular, by considering the representation of $\beta^{i|j}_{mean} = -\omega_{ij}/\omega_{ii}$, where ω_{ik} is an element of the precision matrix $\mathbf{\Omega}$, we can see that network-ΔCoVaR is simply a rescaled version of the precision matrix:

$$\mathbf{\Delta CoVaR} = \Phi^{-1}(\tau)(\boldsymbol{D}_{\mathbf{\Sigma}}^{1/2} - \boldsymbol{D}_{\mathbf{\Omega}}^{-1}\mathbf{\Omega}\boldsymbol{D}_{\mathbf{\Sigma}}^{1/2}). \tag{4.25}$$

In general, the ΔCoVaR networks in the Gaussian case do not carry any additional information compared to β_{mean} and can be directly derived from the covariance matrix of the distribution. In the following section we focus on a t-Student distribution setting, showing how in this case the relation between quantile graphical models, ΔCoVaR and partial correlation networks is more complex.

Parametric QGM and network-$\mathbf{\Delta CoVaR}$ – t-Student distribution

In the multivariate t-Student setting, the conditional quantile functions are not linear (with the exception of the one for $\tau = 0.5$). We can still compute esplicitly the conditional distribution that, as shown in Ding, 2016 is still a t-Student distribution. We show here some characteristics and we analyse the issues arising from the linear approximation of the quantile functions, common in the linear quantile regression estimation procedures.

Let $\mathbf{X} \sim t_p(\mu, \mathbf{\Sigma}, \nu)$ where $\mu, \mathbf{\Sigma}, \nu$ are the location, scale and number of degrees of freedom of the distribution, respectively. We also consider the partition

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\backslash i} \\ \mathbf{X}_i \end{bmatrix} \sim t_p \left(\begin{bmatrix} \mu_{\backslash i} \\ \mu_i \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{\backslash i,\backslash i} \mathbf{\Sigma}_{\backslash i,i} \\ \mathbf{\Sigma}_{i,\backslash i} \mathbf{\Sigma}_{i,i} \end{bmatrix}, \nu \right). \tag{4.26}$$

We focus on the case with $\mu = \mathbf{0}$ as in the previous section and, following Ding, 2016, we have that:

$$\mathbf{X}_i | \mathbf{X}_{\backslash i} \sim t_{p_2}(\mu_{i|\backslash i}, \frac{\nu + d(\mathbf{X}_{\backslash i})}{\nu + p - 1} \mathbf{\Sigma}_{i,i|\backslash i}, \nu + p - 1), \tag{4.27}$$

$$\mu_{i|\backslash i} = \mathbf{\Sigma}'_{\backslash i,i} \mathbf{\Sigma}^{-1}_{\backslash i \backslash i} \mathbf{X}_{\backslash i}, \tag{4.28}$$

$$\mathbf{\Sigma}_{i,i|\backslash i} = \mathbf{\Sigma}_{i,i} - \mathbf{\Sigma}'_{\backslash i,i} \mathbf{\Sigma}^{-1}_{\backslash i,\backslash i} \mathbf{\Sigma}_{\backslash i,i}, \tag{4.29}$$

where $d(\mathbf{X}_{\backslash i}) = \mathbf{X}'_{\backslash i} \mathbf{\Sigma}^{-1}_{\backslash i,\backslash i} \mathbf{X}_{\backslash i}$ is the squared Mahalanobis distance of $\mathbf{X}_{\backslash i}$ from the origin with scale matrix $\mathbf{\Sigma}_{\backslash i,\backslash i}$.

We can then compute the conditional quantile of $\mathbf{X}_i$ as function of $\mathbf{X}_{\backslash i}$ obtaining:

$$Q(\mathbf{X}_i | \mathbf{X}_{\backslash i}, \tau) = Q_{t,\nu+p-1}(\tau) \left(\frac{\nu + d(\mathbf{X}_{\backslash i})}{\nu + p - 1} \mathbf{\Sigma}_{i,i|\backslash i} \right)^{1/2} + \mu_{i|\backslash i}, \tag{4.30}$$

where $Q_{t,\nu+p_1}(\tau)$ is the quantile function of a t-Student distribution with $\nu + p - 1$ degrees of freedom.

The function is not linear in $\mathbf{X}_{\backslash i}$ due to the term $d(\mathbf{X}_{\backslash i})$. Differently from the Gaussian case, the conditional quantile cannot be computed simply as a translation of the β_{mean}.

Figure 4.3 provides a graphical representation of the conditional quantiles: we plot the conditional quantile of a univariate marginal $\mathbf{X}_i$ versus another univariate marginal $\mathbf{X}_j$, keeping all the other variables equal to zero, for multivariate t-Student distributions with different dimensions. (We use 5 degrees of freedom, and we consider distributions with dimension 5, 10, 20, 30, 50 and 1,000, respectively). We see that the curvature of the conditional quantile functions reduces when the dimension of the multivariate distribution increase (i.e. when the quantile is conditioned by more assets). Still, the non linearity remains significant also for a relatively high number of assets. We also see that the conditional quantiles are closer to the mean for high dimensionality, meaning that under t-Student distribution, the presence of a large number of assets that assume

mean values will reduce the dispersion of the conditional distributions, flattening them to their expected value.

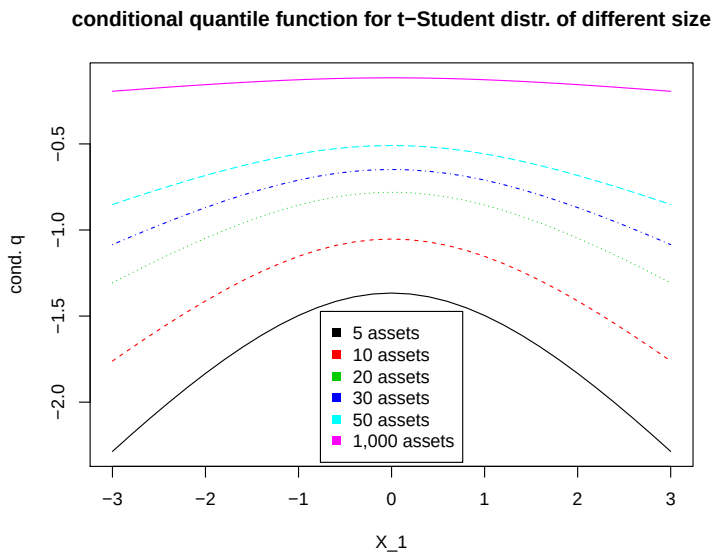


FIGURE 4.3: Conditional quantiles ($\tau = 0.05$) of multivariate t-Student distributions $t_p(\mathbf{0}, \boldsymbol{I}, 5)$ with dimensionality p = 5, 10, 20, 30, 50 and 1,000, respectively. Note that the curvature of the quantile reduces when p increases, but that it remains significant also for relatively high levels of p (e.g. a system of 50 assets).

Equation 4.30 shows that, differently from the Gaussian case, the quantile function under a t-Student distribution is not linear. The assumption of linearity however is common in the empirical estimation based on quantile regression. We can compute the best linear approximation as follows, in order to assess the effect of such assumption on conditional quantiles and $\mathbf{\Delta CoVaR}_\tau$ estimates.

Proposition 1. *Let* $\mathbf{X} \sim t_p(\mu, \mathbf{\Sigma}, \nu)$*, given a partition* $\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\backslash i} \\ \mathbf{X}_i \end{bmatrix}$*, the linear approximation of the conditional quantile* τ *of* $\mathbf{X}_i | \mathbf{X}_{\backslash i}$ *as a function of* $\mathbf{X}_{\backslash i}$ *can be expressed as:*

$$\bar{Q}(X_i | X_{\backslash i}, \tau) = g(\mathbf{X}_{\backslash i}) = Q_{t,\nu+p-1}(\tau) \mathbb{E}_{X_{\backslash i}} \left[\left(\frac{\nu + d(\mathbf{X}_{\backslash i})}{\nu + p - 1} \mathbf{\Sigma}_{i,i|\backslash i} \right)^{1/2} \right] + \mathbf{\Sigma}_{i,\backslash i} \mathbf{\Sigma}_{\backslash i,\backslash i}^{-1} \mathbf{X}_{\backslash i} \tag{4.31}$$

$$= Q_{t,\nu+p-1}(\tau) \mathbb{E}_{X_{\backslash i}} \left[\left(\frac{\nu + d(\mathbf{X}_{\backslash i})}{\nu + p - 1} \mathbf{\Sigma}_{i,i|\backslash i} \right)^{1/2} \right] + \mu_{i|\backslash i}. \tag{4.32}$$

see proof in Appendix D.1.

Remark. *Note that, differently from the true conditional quantiles (4.30), the linear approximations in (4.31) are all parallel to the mean-regression line as in the Gaussian case.*

Remark. *The τ-quantile function for $\tau = 0.5$ is a linear function and coincides with the linear approximation since $Q_t(0.5) = 0$. It follows that, if $\beta^{i|j}_{0.5} = 0$, the two variables are conditionally uncorrelated, consistently to Finegold and Drton, 2011, Proposition 1.*

Given the linear approximation in (4.31), we can quantify the approximation error compared to the real quantile function, and we have the following result:

Proposition 2. *Given $\mathbf{X}$ and the partition as defined in (4.26), The slope of the τ-conditional quantile function for $\tau < \frac{1}{2}$ is higher than the one of the linear approximation for $\mathbf{X}_{\setminus i}|\mathbf{X}_j < 0 \quad \forall j \neq i$.*

See proof in Appendix D.1.

Remark. *The higher the degrees of freedom, the better the linear approximation is, and the smaller the concavity for low quantiles. The limit case is the Gaussian distribution where the quantiles are linear.*

Thanks to equation (4.30), we can compute the CoVaR and ΔCoVaR:

$$\text{CoVaR}_{\tau}^{\mathbf{X}_i|\mathbf{X}_j} = Q_{t,\nu+p-1}(\tau)\left(\frac{\nu + d(\mathbf{X}_{\setminus i}^{d_j})}{\nu + p - 1}\mathbf{\Sigma}_{i,i|\setminus i}\right)^{1/2} + \beta^{i|\setminus i}_{\text{mean}}\mathbf{X}_{\setminus i}^{d_j}, \tag{4.33}$$

where $\mathbf{X}_{\setminus i}^{d_j}$ denotes a state of the system in which $\mathbf{X}_j = \text{VaR}_{\tau}^{\mathbf{X}_j}, \mathbf{X}_{\setminus\{i,j\}} = \text{VaR}_{50\%}^{\mathbf{X}_{\setminus\{ij\}}}$ and $\beta^{i|\setminus i}_{\text{mean}} = \mathbf{\Sigma}'_{\setminus i,i}\mathbf{\Sigma}^{-1}_{\setminus i,\setminus i}$.

$$\begin{aligned}\Delta\text{CoVaR}^{\mathbf{X}_i|\mathbf{X}_j} &= \text{CoVaR}_{\tau}^{\mathbf{X}_i|\mathbf{X}_j} - \text{CoVaR}_{50\%}^{\mathbf{X}_i|\mathbf{X}_j} \\ &= Q_{t,\nu+p-1}(\tau)\mathbf{\Sigma}^{1/2}_{i,i|\setminus i}\left(\left(\frac{\nu + d(\mathbf{X}_{\setminus i}^{d_j})}{\nu + p - 1}\right)^{1/2} - \left(\frac{\nu}{\nu + p - 1}\right)^{1/2}\right) + \beta^{i|\setminus i}_{\text{mean}}\mathbf{X}_{\setminus i}^{d_j}.\end{aligned} \tag{4.34}$$

In the last passage we used the fact that $d(\text{VaR}_{50\%}^{X_{\setminus i}}) = \mathbf{0}$ under a centered t-Student distribution.

Remark. *As a consequence of Proposition 2, we have that for $\tau < 0.5$, the estimation of ΔCoVaR done with linear quantile regression, or a linear approximation of the quantile function, would result in an underestimation of risk.*

$$\beta^{i|j}_{mean}(\text{VaR}^{\mathbf{X}_j}_{\tau} - \text{VaR}^{\mathbf{X}_j}_{50\%}) \leq \Delta\text{CoVaR}^{\mathbf{X}_i|\mathbf{X}_j}. \tag{4.35}$$

Other than the computation of $\mathbf{\Delta CoVaR}$, the explicit representation of conditional distribution of $\mathbf{X}_i|\mathbf{X}_{\setminus i}$ allows to highlight an additional characteristic of QGM in a t-Student setting, related to the standard deviation of the conditional distribution. From (4.27) we see that the standard deviation of the distribution of $\mathbf{X}_i|\mathbf{X}_{\setminus i}$ is an increasing function of $d(\mathbf{X}_{\setminus i})$. It follows that any deviation of the conditioning variable from the center of the distribution (positive or negative) has the effect of increasing the standard deviation of the conditioned variable. The economic interpretation is that, in line with several studies, the volatility is then a channel for the transmission of financial distress (e.g. Engle and Susmel, 1993; Diebold and Yılmaz, 2014). That is, a shock in the price of an asset may have as a consequence a spike in volatility in the neighbouring assets, even without considering an heteroskedastic model.

Finally, we highlight that a limitation of the t-Student framework presented here, is that the distribution has a unique value for the degrees of freedom, resulting in an easier estimation, but in a less flexible framework as it does not allow to model separately the shape of the tails in correspondance to each couple of variables. Still, this parametric framework allows to point out the relevance of non-linearity of the quantile function. In the empirical application we will consider a non-parametric estimation based on quantile regression, described in Section 4.4.

4.3.3 Normalized network-$\mathbf{\Delta CoVaR}$

ΔCoVaR measures the absolute impact on the VaR on an institution. In some cases it may be useful to consider the relative impact. We propose a standardized version of network-ΔCoVaR. For the bivariate case Mainik and Schaanning, 2014 propose to normalize ΔCoVaR dividing it by $\text{CoVaR}^{X_i|X_j}_{50\%}$ (see Eq. 4.5). We extend the same approach to the multivariate case as follows:

Definition 3.

$$\Delta CoVaR^{\%\,\mathbf{X}_i|\mathbf{X}_j}_{\tau} = \frac{\Delta CoVaR^{\mathbf{X}_i|\mathbf{X}_j}_{\tau}}{CoVaR^{\mathbf{X}_i|\mathbf{X}_j}_{50\%}}. \tag{4.36}$$

This quantity represents the relative impact on $\mathbf{X}_i$ of an idiosyncratic shock hitting variable $\mathbf{X}_j$, and will be used in the empirical analysis in Section 4.6.

We can also write (4.36) in matrix form as follows:

$$\mathbf{\Delta CoVaR}_\tau^{\%} = \boldsymbol{D}_{\text{CoVaR}_{50\%}}^{-1} \mathbf{\Delta CoVaR}, \tag{4.37}$$

where $\boldsymbol{D}_{\text{CoVaR}_{50\%}}$ is a diagonal matrix composed of the $\text{CoVaR}_{50\%}$s.
Under the Gaussian setting this standardization helps us further clarify the relation between partial correlation networks and $\mathbf{\Delta CoVaR}$. Indeed, if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, using (4.20), (4.36) and $\Phi^{-1}(50\%) = 0$, we have:

$$\Delta\text{CoVaR}_\tau^{\%,i|j} = \frac{\beta_{\text{mean}}^{i|j} \Phi^{-1}(\tau)\sigma(\mathbf{X}_j)}{\Phi^{-1}(\tau)\mathbf{\Sigma}_{i,i|\setminus i}^{1/2} + \beta_{\text{mean}}^{i|j} \text{VaR}_{50\%}^{\mathbf{X}_j}} \tag{4.38}$$

$$= \frac{\beta_{\text{mean}}^{i|j} \sigma(\mathbf{X}_j)}{\mathbf{\Sigma}_{i,i|\setminus i}^{1/2}}, \tag{4.39}$$

or in matrix form, using (4.22):

$$\mathbf{\Delta CoVaR}_\tau^{\%} = D_{\mathbf{\Omega}}^{1/2} \boldsymbol{B} D_{\mathbf{\Sigma}}^{1/2}, \tag{4.40}$$

where $D_{\mathbf{\Omega}}$ is a matrix with the diagonal elements of $\mathbf{\Omega}$ on the diagonal. For comparison, partial correlation networks are defined as:

$$\boldsymbol{P} = D_{\mathbf{\Omega}}^{1/2} \boldsymbol{B} D_{\mathbf{\Omega}}^{-1/2}, \tag{4.41}$$

we observe that the two matrices differ only by the standardization of the columns: for $\mathbf{\Delta CoVaR}_\tau^{\%}$ the unconditional standard deviations $D_{\mathbf{\Sigma}}^{1/2}$ are used, while for partial correlation the conditional ones $D_{\mathbf{\Omega}}^{-1/2}$. Intutively, we can see the standardization in partial correlation network as a way to assess the effect of an idiosyncratic shock, corresponding the the component of standard deviation not explained by the movement of other assets. In the $\mathbf{\Delta CoVaR}^{\%}$ case instead, we assess the effect on the system of one asset being in distress, quantifying the distress state as its VaR.
If the distribution of the data is not multivariate Gaussian, we do not have the equivalency between $\beta_{\text{mean}}^{i|j}$ and $\beta_\tau^{i|j}$, and therefore partial correlation networks and tail $\mathbf{\Delta CoVaR}^{\%}$ networks differ not only for the standardization.

4.4 Estimation of network-ΔCoVaR

4.4.1 Parametric estimation under Gaussian and t-Student distributions

We first consider a parametric approach for the estimation of network-ΔCoVaR. If we assume a Gaussian or a t-Student distribution of the variables, we can use Equations 4.18 and 4.33 to compute network-ΔCoVaR analytically. Therefore, the estimation problem boils down to the estimation of the parameters of the multivariate distributions, in particular of the dispersion matrix. In settings with a large number of assets compared to the length of the time series, such estimation can become challeging and the sample estimates are typically ill-conditioned. Concerning the Gaussian case, we can use more robust approaches that allow to obtain better conditioned estimates such as *Ledoit-Wolf* shrinkage estimator (Ledoit and Wolf, 2004a), random matrix theory filtering (Laloux et al., 2000) or graphical lasso (*glasso*) (Friedman, Hastie, and Tibshirani, 2008). The advantage of this last approach, is that it provides a sparse estimate of the precision matrix and it has a natural interpretation in terms of Markovian networks (Højsgaard, Edwards, and Lauritzen, 2012).
Under the t-Student distribution it is possible to rely on the *tlasso* model introduced by Finegold and Drton, 2011, and used in the context of banking network estimate by Torri, Giacometti, and Paterlini, 2018. This model allows us to obtain a robust estimate of the scatter parameter, that is also parsimonious thanks to the sparsity induced by the L_1 penalization applied on the precision matrix.

4.4.2 Non-parametric estimation with SCAD-penalized quantile regression

Quantile regression provides a powerful framework for the estimation of non-parametric quantile graphical models, and it can be directly used for the estimation of network-ΔCoVaR. We consider penalized quantile regression, that allows to perform model selection and estimation at the same time, increasing the efficiency of the estimator in setting with a high number of variables in relation to the observations (see e.g. Koenker, 2005), and we introduce the usage of SCAD-penalized quantile regression (Wu and Liu, 2009) in the estimation of quantile graphical models, alternatively to the more traditional *lasso* penalization (Belloni, Chernozhukov, et al., 2011).

The procedure consists in the estimation of the quantile graphical model, that can then be used to compute network-$\mathbf{\Delta CoVaR}$ as in (4.13). Similarly to Belloni, Chen, and Chernozhukov, 2016, we consider a quantile graphical model specification in which the $f(\mathbf{X}_{\setminus i}, \tau)$ is assumed to be an affine function. $\text{CoVaR}^{\mathbf{X}_i|\mathbf{X}_{\setminus i}}$ is therefore an affine function of $\mathbf{X}_{\setminus i}$ too. We estimate the coefficients of the model independently by penalized quantile regression, and we then construct the matrix $\boldsymbol{B}$ of the β coefficients. The use of penalization allows to regularize the solution in large dimensional settings by increasing the efficiency of the estimator (see e.g. Tibshirani, 1996; Fan and Li, 2001; Friedman, Hastie, and Tibshirani, 2008).

Each quantile regression can be computed by minimizing the following function:

$$\min_{\boldsymbol{\beta}_\tau^{i|\cdot}} \mathbb{E}[\rho(\mathbf{X}_i - \boldsymbol{\beta}_\tau^{i|\cdot}\mathbf{X}_{\setminus i})] + \sum_{j=1, j\neq i}^{p} n\sqrt{\tau(1-\tau)}p_\lambda(\boldsymbol{\beta}_\tau^{i|j}) \qquad \forall i = 1, \dots, p, \tag{4.42}$$

where $\rho_\tau(x) = (\tau - \mathbb{I}_{\{x\leq 0\}})x$ is the asymmetric loss function (see Koenker and Hallock, 2001), n is the number of observations, $\boldsymbol{\beta}_\tau^{i|j}$ is the jth element of the vector of coefficients $\boldsymbol{\beta}_\tau^{i|\cdot}$, and $p_\lambda(\beta)$ the penalization function. The introduction of the penalization allows to regularize the solution of the estimation problem, especially in high dimensional settings. In particular, for some penalization functions, such as the lasso, it is possible to induce sparsity in the regressors (i.e. setting some of the coefficients exactly to zero), indentifying only the most relevant ones (see Tibshirani, 1996; Fan and Li, 2001).

For the estimation of the model Belloni, Chen, and Chernozhukov, 2016 propose an algorithm suitable for high-dimensional setting, in which they induce sparsity using lasso penalization (see Algorithm 3.2, in Belloni, Chen, and Chernozhukov, 2016). We consider an alternative estimation algorithm based on the Smoothly Clipped Absolute Deviation (SCAD) penalization (instead of the lasso). SCAD penalty has been introduced in the statistical literature by Fan and Li, 2001 and is defined as:

$$p_\lambda^{\text{SCAD}}(\beta) = \begin{cases} \lambda|\beta| & \text{if } |\beta| \leq \lambda \\ -\dfrac{|\beta|^2 - 2a\lambda|\beta| + \lambda^2}{2(a-1)} & \text{if } \lambda < |\beta| \leq a\lambda \\ \dfrac{(a+1)\lambda^2}{2} & \text{if } |\beta| > a\lambda \end{cases}, \tag{4.43}$$

where λ and a are positive constants. The function is linear and equivalent to a lasso penalization

near the origin, then quadratic in the interval $(\lambda, a\lambda]$ and finally flat. It belongs to the wider class of folded concave penalizations defined by Fan and Li, 2001.
Wu and Liu, 2009 proved that the model estimation performed using SCAD penalty possesses the oracle property (i.e., asymptotically, it identifies the right subset model and it has an optimal estimation rate), while the lasso penalty estimator does not (see Zou, 2006; Lam and Fan, 2009). The non-convexity of the penalization makes the optimization problem much harder to solve, however several specific algorithms have been developed. In the penalized mean regression framework, iterative algorithm based on local linear approximation of the penalty function allow for relatively fast estimation (see Zou and Li, 2008; Fan, Xue, and Zou, 2014). Moreover, Fan, Xue, and Zou, 2014 showed that if such algorithms are initialized at a lasso optimum satisfying certain properties, the two-stage procedure produces an oracle solution, and Loh and Wainwright, 2013 show that under suitable conditions, any stationary point of the objective function will lie within the statistical precision of the underlying parameter vector and propose specific algorithms to obtain near-global optimum solutions.
In the context of penalized linear regression, the non-convexity does not allow a linear programming representation of the problem but, following Wu and Liu, 2009, it is possible to address the problem using a Difference Convex Algorithm (DCA) that uses the representation of SCAD penalty as the difference between a linear and a convex function, and solves a sequence of convex problems to approximate the SCAD problem efficiently.

The SCAD penalization requires the calibration of two parameters: the penalization factor λ and the parameter a that regulates the shape of the penalty. Concerning the parameter a, we follow the suggestion of Fan and Li, 2001, that sets $a = 3.7$, that is the value that minimizes Bayes risk. Concerning the parameter λ, we use Bayesian Information Criterion (BIC) to select the best model across a grid of values for λ.
We further highlight that, in the Gaussian case, the estimation performed using linear quantile regression should give the same result of the parametric model, as the conditional quantile functions are linear, while under the t-Student assumption the quantiles are not linear and the linear approximation leads to an underestimation of tail risk.

Tail-located non-parametric estimation

One of the main drawbacks of the conditional quantile estimation obtained with quantile regression is the implicit linearity of the quantile function. We have seen, as an example, that under the t-Student distribution this assumption does not hold. In low dimensional settings this issue can be addressed by considering non-linear quantile regression, using a basis expansion (Ali, Kolter, and Tibshirani, 2016) or kernel weights (Bonaccolto, Caporin, and Paterlini, 2018).These techniques however typically increase the number of parameters in the model, leading to larger estimation error, especially in highly multivariate setting. We consider an alternative solution, that is to estimate the model on a calibration set defined specifically to approximate a particular section of the conditional quantile function.

Using the terminology of Belloni, Chen, and Chernozhukov, 2016, we need to estimate a $\mathcal{W}$-Conditional quantile graphical model in which the quantile is computed conditional to a family of events $\varpi \in \mathcal{W}$ corresponding to a distress state of the system.

In the bivariate case the conditioning set can be identified as the set of all the observations in which $X_1 < \{\text{VaR}_\tau^{X_1} | X_1 < \kappa\}$, consisting of the observations where the conditioning asset $\mathbf{X}_1$ is below a certain threshold κ (e.g. we may consider $\kappa = \text{VaR}_{50\%}^{X_1}$). In a multivariate setting the selection of the conditioning set is more challenging, as we have multiple conditioning assets. We propose the following two step estimation procedure:

1. Identify the distress set $\mathcal{W}$ as the set of time observations in which a diversified portfolio is below a certain threshold (e.g. the 50th percentile). The portfolio is composed by the equity returns of the banks sample, weighted by their respective standard deviation.[7]

2. Construct the $\mathcal{W}$-Conditional quantile graphical model with the required τ (e.g. 0.05) using as a conditioning set $\mathcal{W}$, and use the estimated coefficients to compute $\mathbf{\Delta CoVaR}$.

This approach allows us to consider the characteristics of the conditional quantile function in situations of downward movements of the overall market, offering a better approximation of the true quantile function.

[7]We used this weighting to avoid to over-represent large or more volatile banks in the sample. We considered alternative distress states, consisting respectively in the days with lowest returns of an equally weighted portfolio of all the assets in the system and a capitalization weighted portfolio of the same assets. In both cases the the results are similar due to the high level of collinearity in the market, leading to high correlation of all these conditioning portfolios. The results are not reported for brevity.

Indeed, by conditioning to the datapoints with the lowest aggregated returns, we can obtain linear approximation of the quantile functions that are more focused on the lower tail of the distribution. This in turn should allow to better reflect the slope of the quantile function in distressed situation. Note that, compared to the bivariate case, in which we can estimate CoVaR using only the lowest returns of the conditioning asset, in a multivariate setting we cannot choose a set for each conditioning asset as they are all estimated together. If the assets in the market are positively correlated, the portfolio that we consider represents a sufficiently good approximation (in the trivial case of perfect rank correlation among all the variables, the bivariate case extends perfectly). The choice of the conditioning set is an heuristic approach to identify the observations that are most likely to influence the behaviour of the tails of the assets.[8]

4.5 Network systemic risk indicators

Once the $\mathbf{\Delta CoVaR}$ networks are estimated, we can use them to assess systemic risk in the system, using specific network measures. Differently from partial correlation networks, $\mathbf{\Delta CoVaR}$ networks are directed. We can therefore compute two sets of indicators, one constructed using incoming edges, and the other based on outcoming edges. Similarly to Diebold and Yılmaz, 2014, we can interpret these two sets as indicators of fragility to shock from the system (systemic fragility) and contribution to systemic risk (systemic relevance), respectively.

In the empirical analysis, we first focus on the structural properties of the network, in particular on the community structure. A community structure is a topological feature that consists in the presence of groups of nodes that are more connected to each other than to the rest of the network. Such feature is commonly found in social networks, where the groups may reflect the sharing of common interests and background, or geographical proximity (see e.g. Girvan and Newman, 2002). Similarly, in a banking system the presence of communities may derive from the preferential relationship among banks in terms of exposures in the interbank networks, similarity of business models or, in the case of Europe, historical presence of national banking systems relatively separated from each other. We can quantify the level of community structure in a network by using the concept of modularity, a quantity that can be optimized to obtain

[8] We also considered an alternative procedure in which the distress set is defined separately for each node, considering a portfolio composed of its neighbours in a QGM with $\tau = 0.5$ (median QGM). The results are rather similar and are not reported for brevity.

the optimal partition for the network.[9] We expect that in the European banking system, despite the increasing integration is still characterized by the presence of strong national communities, due to differences in term of market reach, business practices and, to some extent, regulatory frameworks.

After the study of network properties, we focus on the analysis of individual institutions using network centrality measures. These are commonly used for the identification of the most relevant nodes in a network. In banking applications they have been used to assess the relevance of institutions in the system. From the literature we mention Balla, Ergen, and Migueis, 2014, Diebold and Yılmaz, 2014 and Billio et al., 2012. Network centrality indicators are considered to some extent also in the normative framework: two of the indicators considered in the G-SIB assessment methodology (see FSB, 2013) account for the level of intra-financial system liabilities and assets, respectively, consistently with the in- and out-strength in the network of interbank exposures (Kaltwasser and Spelta, 2018). We consider in particular strength centrality, that is computed as the sum of the weights of the connected edges. Since the network is directed, we compute strength centrality both for incoming and outcoming nodes, obtaining in-strength c^i_{in}, that denotes systemic fragility (incoming edges), and out-strength c^i_{out}, denoting systemic relevance (outcoming edges):

$$c^i_{\text{in}} = \sum_{j=1}^{p} w_{ij}, \tag{4.44}$$

$$c^i_{\text{out}} = \sum_{j=1}^{p} w_{ji}. \tag{4.45}$$

Credit risk weighted interconnectivity

Tipically, centrality measures allow to assess the role of each node in the system purely in terms of the interconnections among the nodes in the network. Still, the degree of systemic risk depends not only on these interconnection, but also on the idiosyncratic characteristics of one nodes, and the ones of their neighbours. For this reason, the inclusion of information on the idiosyncratic characteristics of each bank in network models is a challenging, but relevant topic. Several contagion model for instance use balance sheet data to assess the ability of a bank to absorb a shock or to transmit it to other institutions (Gai, Haldane, and Kapadia, 2011; Hurd, 2016). Benoit et al.,

[9] See Section 1.3.4 for the formal definition of modularity and Newman, 2004 for a greedy optimization algorithm.

2017 highlights how one of the main challenges for the measurement and assessment of systemic risk is to develop measures that combine multiple sources of information, including market data, balance sheet figures, expert assessments and proprietary data on portfolio position.

We propose here to integrate the $\mathbf{\Delta CoVaR}$ measure with balance sheet data. In particular, we use Non-Performing-Loans (NPL) as a proxy of credit risk of an institution. We build for each bank two indicators $c^i_{\text{NPL-in}}$, $c^i_{\text{NPL-out}}$ composed as the average NPL ratio of each neighbouring bank, weighted by the strength of the edges:

$$c^i_{\text{NPL-in}} = \sum_{j=1}^{p} w_{ji}\text{NPL}_j, \tag{4.46}$$

$$c^i_{\text{NPL-out}} = \sum_{j=1}^{p} w_{ij}\text{NPL}_i, \tag{4.47}$$

where w_{ij} is the strength of the edge ij, and NPL_i is the NPL ratio of bank i, computed as the ratio between non performing loans over total loans. Such measure may be computed using several different indicators.

This measure allows us to provide a more reliable and nuanced assessment of systemic risk. Interconnection alone may not be enough to cause financial risk. The credit quality of the neighbours may influence the risk of a bank: a strong connection with a solid bank may indeed not represent a threat, while a fragile company may significantly affect the stability of the neighbours. We underline that, differently from other systemic risk measures (for instance SRISK (Brownlees and Engle, 2016) and to some extent $\mathbf{\Delta CoVaR}$), c^i_{NPL} cannot be translated directly in terms of capital shortfalls, or other monetary indicators. Instead, it can be thought as an overall indicator of systemic relevance/fragility, similarly to the score-based system in the G-SII framework (FSB, 2013). The difference is that in this case the aggregation of these measures is not a linear combination, but is defined as in (4.5). We opted for a simple weighted average, but future work may explore different forms for the aggregation function.

Alternatively to the NPL ratio, we could use a market based indicator of credit risk, such as the spread of credit default sawps (CDS). Such contracts however may already include a componenent referring to systemic risk accounted by the market. Alternatively, we could build a forward looking measure using the determinants of NPL, to overcome the fact that NPL can be considered an ex-post measure of systemic risk. We leave this analysis to future studies.

4.6 Empirical analysis

4.6.1 Data and methodology

In the empirical analysis we estimate network-ΔCoVaR on the equity returns of a set of large European banks. The set includes 36 banks and is largely overlapping to the set of banks included in the EBA stress test exercises. In line with Balla, Ergen, and Migueis, 2014; Billio et al., 2012; Adrian and Brunnermeier, 2016, we consider equity returns as these data are widely available, very liquid and allow us to use daily data or even higher frequencies. Other studies on systemic risk consider CDS spreads, as they may reflect more closely the credit risk of financial institutions (e.g. Puliga, Caldarelli, and Battiston, 2014). However in many cases CDS time series are incomplete and their market are not liquid enough to use daily data.

Due to the presence of heteroskedasticity in the data, we proceed in two steps: first we fit a DCC-GARCH model to the data (Engle, 2002), and then we estimate the conditional quantile model on the residuals. The residual of the GARCH model are standardized to have unitary standard deviation, and this allows to avoid that the size of the variables influences the penalization. The practice of standardizing the variable is indeed common in the estimation of penalized models (see Højsgaard, Edwards, and Lauritzen, 2012).

We consider two different periods: Jan 2007–Dec 2012, roughly corresponding to the crisis period and the one immediately before, and Jan 2013–Apr 2018, corresponding to the period after the crisis. We analyse a sample of 36 banks from 15 countries, for which the time series are available for the entire period. The list of the institutions is available in Appendix D.2. Concerning the NPL ratios, we use from Thomson Reuters Datastream data of end-2012 for the first period, and end-2017 for the second period.

We estimate two different network structures: first we consider $\mathbf{\Delta CoVaR}$ network using the tail-located non-parametric estimation approach outlined in Section 4.4.2. Concerning the value of τ, we opted for a value of 0.1 to achieve a good balance in terms of tail risk analysis, and stable numerical results. Lower values of τ may indeed lead to unreliable estimates in finite samples. We then estimate a quantile graphical model using $\tau = 0.5$, that we call *median-QGM* to describe the interconnections of the conditional medians. This model, has an interpretation similar to partial correlation networks, and allows to describe the interconnections in term of conditional

location of the distributions.[10] We highlight that the computation of the $\mathbf{\Delta CoVaR}$ for $\tau = 0.5$ is not meaningful as it would be always equal to zero by construction. We therefore consider the conditional quantile computed using data standardized by the unconditional variance.

4.6.2 Empirical results

Structural properties of the network

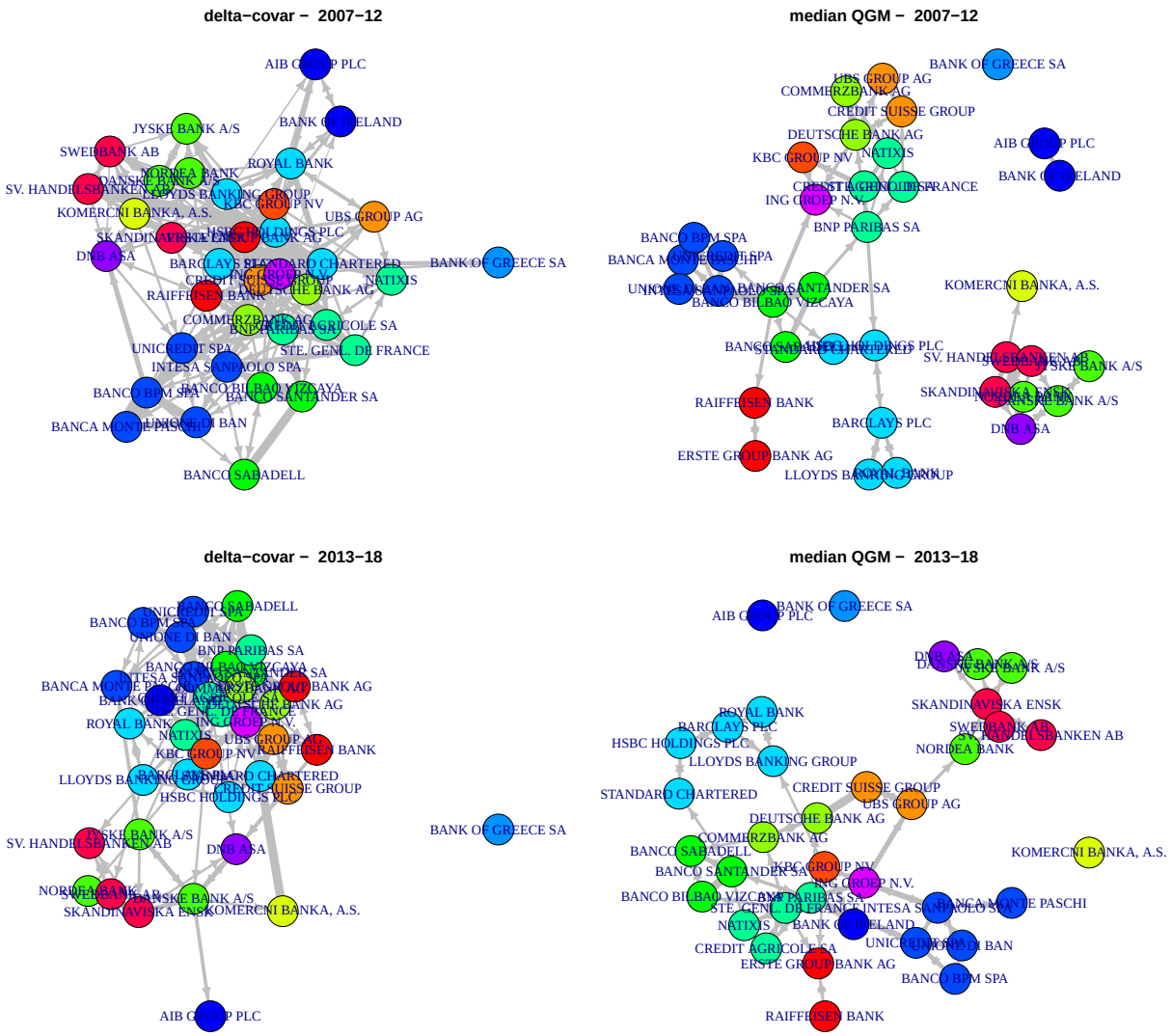


FIGURE 4.4: Network representation of ΔCoVaR ($\tau = 0.1$) and median QGM model for the European banking system computed on the periods 2007-12 and 2013-18.

[10] In a preliminary analysis we compared these networks to partial correlation networks computed using *tlasso* (Finegold and Drton, 2011) and the results are rather similar, confirming the suitability of the t-Student parametric assumption, at least for the analysis of conditional means networks. The results are not reported for brevity.

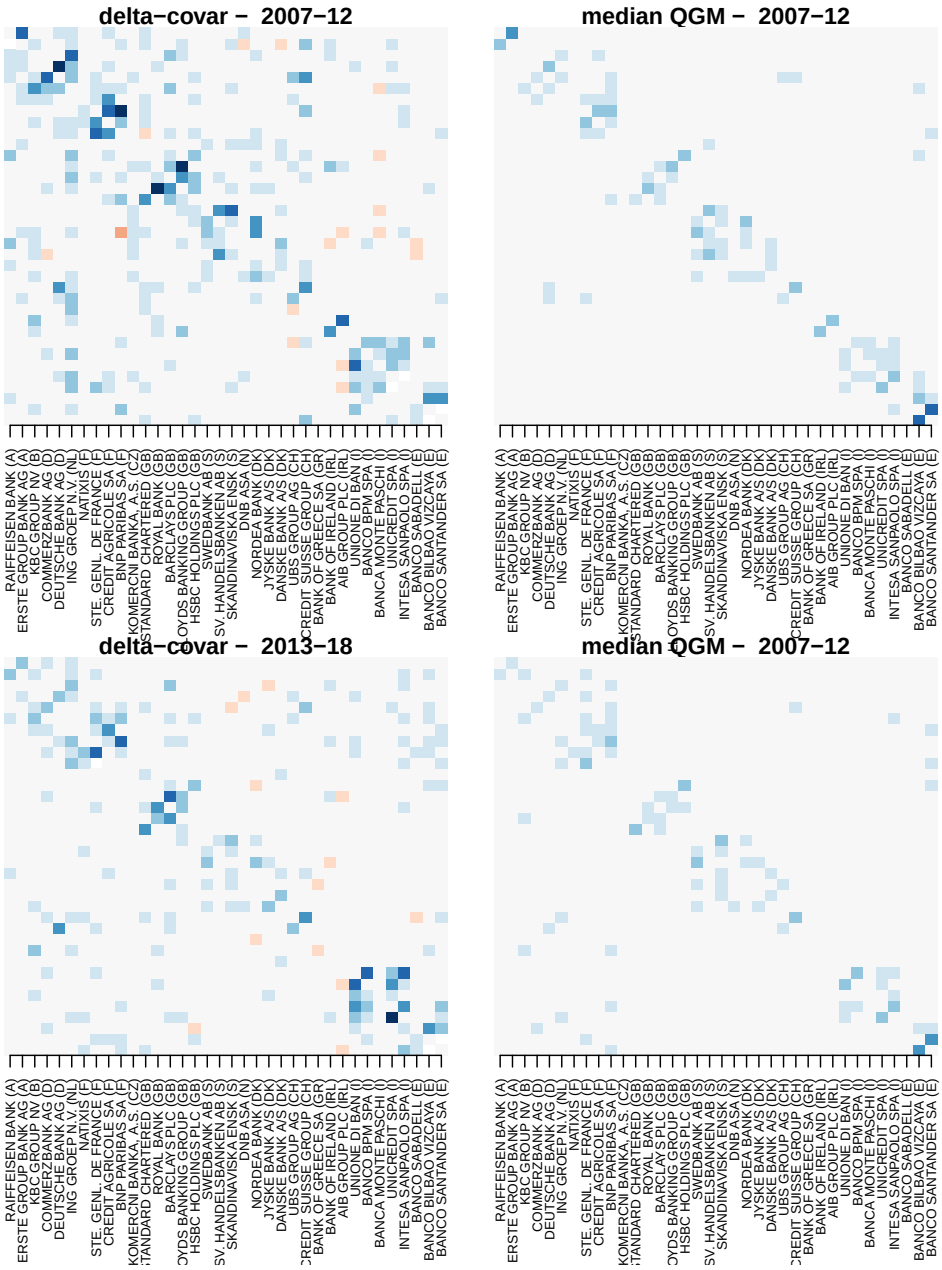


FIGURE 4.5: Adjacency matrix representation of ΔCoVaR ($\tau = 0.1$), median QGM model and *tlasso* for the European banking system computed for the periods 2007-12 and 2013-18.

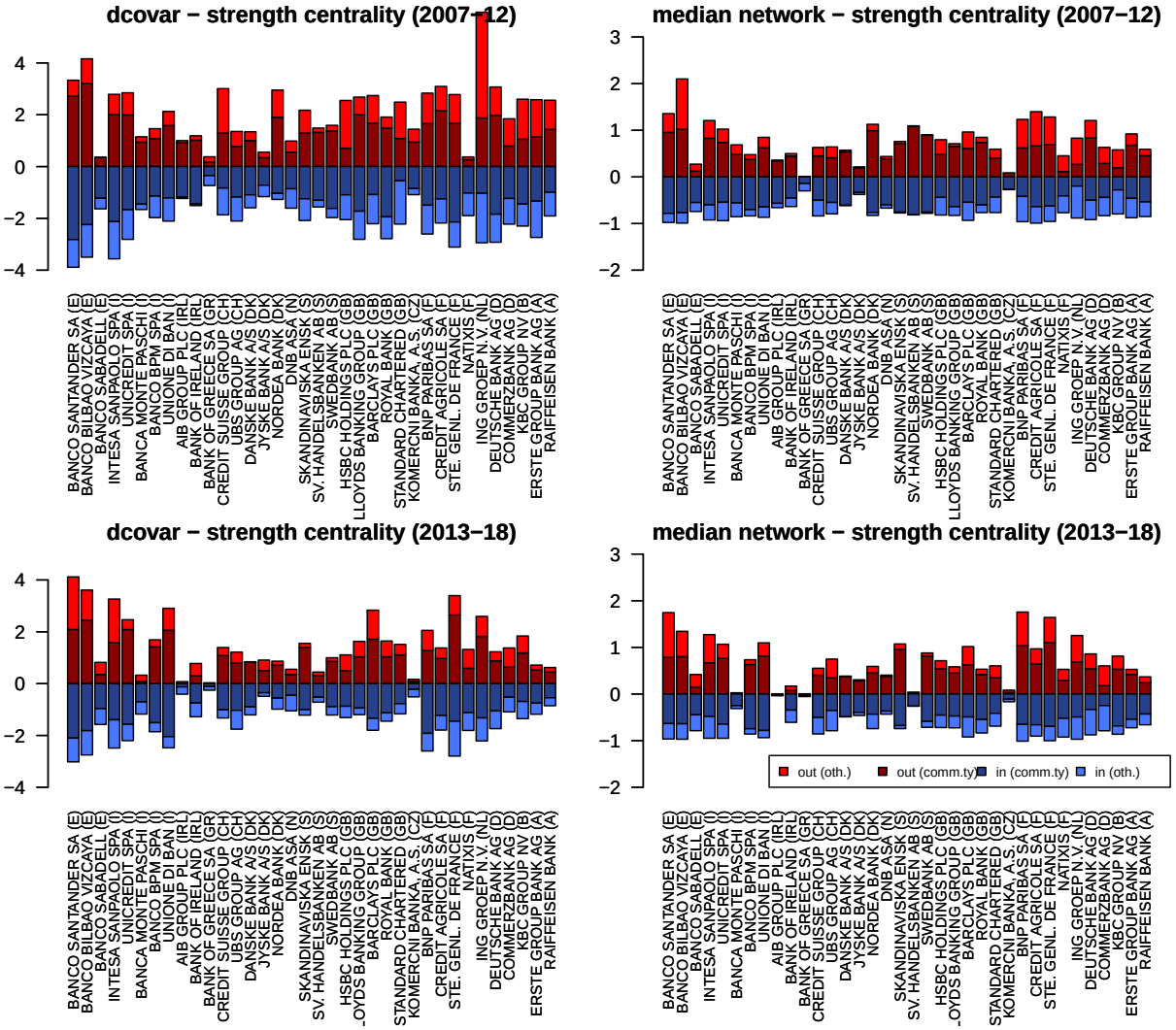


FIGURE 4.6: In- and Out-strength centrality (blue and red, respectively) of the banks in the system for ΔCoVaR networks (left) and median QGM (right), for the periods 2007-12 and 2013-18. We also highlight the decomposition of strength centrality: the darker parts of the bars represent the component of strength centrality attributable to banks in the same community, while the lighter parts to banks in other communities.

We report in Figure 4.4 the $\mathbf{\Delta CoVaR}_\tau$ network ($\tau = 0.1$) and the median QGM ($\tau = 0.5$) for the two periods. In Figure 4.5 we represent the adjacency matrices. The network is computed using daily data. We see that the structures retrived by network-$\mathbf{\Delta CoVaR}_\tau$ are less sparse than the median QGM in both periods. This is aligned to the results presented in Section 4.3.2 for the t-Student distribution case, where the non-linearity of the quantile functions leads to stronger effects in the tails of the distributions compared to the means. We also notice that, in both networks, institutions from the same country tend to stay close to each other. This geographical division is much clearer in the median QGM than in the $\mathbf{\Delta CoVaR}$ networks. The interpretation is that economic shocks are transmitted at international scale in terms of conditional tail risk to a greater extend compared to the conditional means. We also see that the median QGM networks are characterized by a clearer division of the banks according to countries (highlighted with different colors), denoting a more pronounced modular community structure.

We further study the topological properties of the network by identifying the community structure using the modularity maximization algorithm proposed by Newman, 2004. Tables D.2 and D.3 in Appendix D.3, report the composition of the optimal communities for both $\mathbf{\Delta CoVaR}_\tau$ and median QGM networks for the two periods. We see that the communities strongly reflect the nationality of the banks, and are very similar between the $\mathbf{\Delta CoVaR}_\tau$ and the median QGM networks.

We highlight that, in accordance with the visual inspection, the community structure is stronger in the median QGM than in $\mathbf{\Delta CoVaR}_\tau$, as shown by the value of the modularity reported in Table 4.2, where we also report the modularity of the optimal partitions. We see that the median QGM has a stronger community structure compared to the $\mathbf{\Delta CoVaR}_\tau$ network, especially in the period of the crisis. Comparing the two time periods, in accordance to the analysis of CDS spreads partial correlation networks in Torri, Giacometti, and Paterlini, 2018, the median QGM appears to have a stronger community structure in the period 2007-12, that includes the global financial crisis compared to the following period. Instead, the modularity of the tail network remains approximately stable. The combined interpretation of these factors is that, in crisis period, local clusters of banks tend to insulate from each other, but that such segregation is not sufficient to contain tail risk, that can still propagate across the entire network.

TABLE 4.2: Modularity of the optimal network partitions

	2007-12	2013-18
$\mathbf{\Delta CoVaR}$	0.38	0.38
median QGM	0.49	0.42

In- and out-strength centrality analysis and decomposition

Focusing on individual institutions, Figure 4.6 reports the in- and out-strength centrality for each bank. Strength centrality is one of the most common network indicators, and is used to assess the relevance of a node in the system. Since the network is directed, we distinguish between in-strength centrality, that denotes the fragility of an institutions, and out-strength, that denotes the potential of systemic risk transmission. We also represent the component of strength attributable to edges with banks in the same community and banks outside. We see that the centrality measures are rather heterogeneous among the assets in the $\mathbf{\Delta CoVaR}$ networks, especially for out-strength centrality, denoting the presence of systemic risk spreaders. More in detail, we see that in the period 2007–12 the most systemically relevant bank (i.e. the ones with the highest out-strength) is the Dutch *ING*, followed by two Spanish banks (*Banco Santander* and *BBVA*). We also see that many of the most interconnected institutions are settled in Central Europe and United Kingdom, and not in peripheral countries. This suggests that banks in *core* countries are particularly relevant in the transmission of tail risk. Concerning the decomposition of strength directed to nodes in the same communities and to nodes in other communities, we see that in most of the cases the out-stregth is composed for the largest part of links to banks in the same community, but that the component directed to other communities is relatively large for banks in central Europe. A similar situation applies to in-strength. Concerning the median QGM, we see that both in- and out-strength are more homogeneous across banks, meaning that the relative importance of each institutions is more uniform. The decomposition shows a larger relevance of intra-communities edges compared to the $\mathbf{\Delta CoVaR}_\tau$ network, coherently to the stronger community structure for this network identified before.

In the period 2013–18, the results are similar, although the in- and out-strengths are on average smaller. We also notice that some of the financial institutions appear to be significantly disconnected from the system, in particular *Bank of Greece* and the two Irish banks *AIB Group* and *Bank of Ireland*. The strength centralities in the $\mathbf{\Delta CoVaR}_\tau$ network and the median QGM are

rather similar, while in the previous period they show significant differences.
Overall, this evidence suggests that in the first period, that includes the great financial crisis, the transmission of risk was more concentrated on the tails, while in the post crisis period the transmission channels for tail risk are similar to the ones for the conditional mean. The combined results denote a scenario where, in periods of financial crisis, idiosyncratic shocks have a direct effect on the conditional expected return of other institutions only on regional basis, but have a much more widespread effect on the conditional tail risk. The results are consistent with the theoretical analysis performed in Section 4.3.2, where we highlighted how in the case of a t-Student distribution, the tail risk interconnection is pronounced compared to the relation of conditional means.
This shows evidence that the transmission of distress across the network is concentrated on tail risk, and not in the means (indeed networks build on central measures, as QGM and *tlasso*, show an interconnection structure with a strong community structure).

Credit risk weighted interconnectivity

We report in Figure 4.7 the in- and out-centrality adjusted for NPL ratio and, for reference, the plain in- and out- strength centralities. We see several differences between the two sets of indicators. First, we see that banks in Northern and Central Europe, in particular Switzerland, Denmark, Norway, Sweden and UK, have particularly low NPL-adjusted in- and out-strength centrality, especially in the time period 2013-18. This may denote less systemic relevance, and less fragility thanks to the low levels of systemic risk among them and their neighbours. Spanish and French banks maintain a relatively high level of systemic importance. The banks that have the highest NPL-adjusted in- and out-stength centrality are the Italian ones, due to the high level of NPLs, and the strong connectivity among each other. Irish banks, on the other hand, have very high level of both NPL-adjusted in- and out-strength centrality in the period 2007-12, but a low one in the period 2013-18. These results highlight the relevance of peripheral countries in terms of systemic risk management in Europe, but at the same time allows to shift the focus to the interconnectivity of individual banks in the network. Focusing on the Italian banks, we can see for example that Banca Monte dei Paschi di Siena, one of the institutions with the highest level of NPL (i.e. highest credit risk), is not one of the most relevant in terms of NPL-adjusted centrality, due to the limited level of connectivity compared to other banks. On the contrary, the

relevance of ING Bank is relatively high, despite the level of NPL is low, in view of its strong interconnectivity.

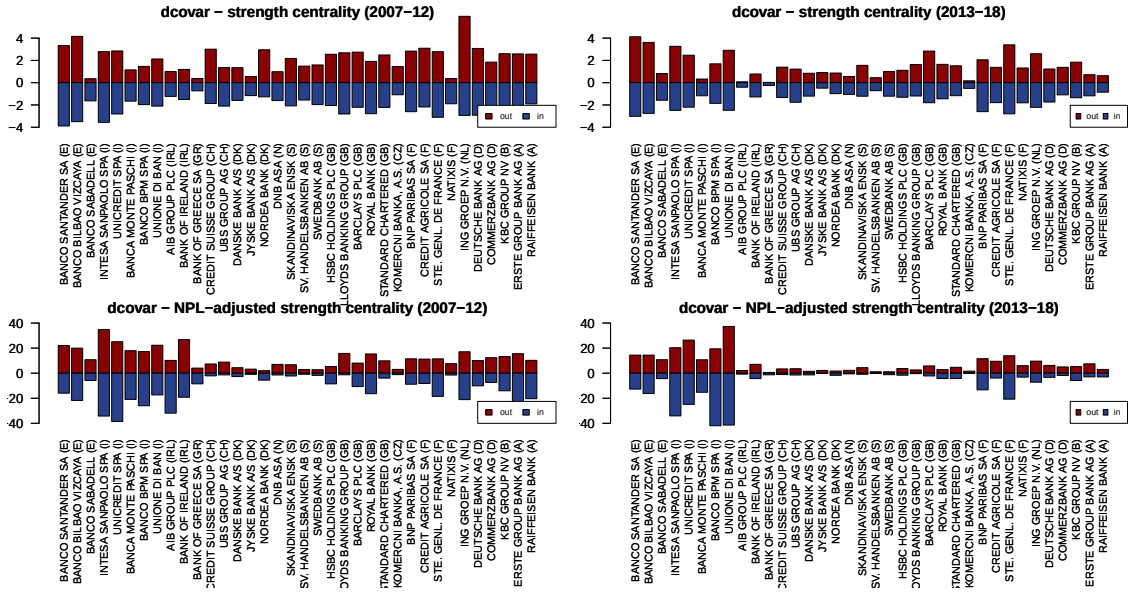


FIGURE 4.7: In- and out-strength centrality (blue and red, respectively) of the banks in the system for $\mathbf{\Delta CoVaR}$ networks in the periods 2007-12 (left) and 2013-18 (right). The top panels represent traditional strength centrality, while the bottom ones are the NPL-adjusted centralities

4.7 Conclusion

In this work we have introduced network-ΔCoVaR (or $\mathbf{\Delta CoVaR}$), a framework that allows us to study systemic risk from a network perspective focusing on conditional tail risk. After introducing CoVaR and ΔCoVaR in the bivariate case, and highlighting their main limitations, we formalize the definition of network-ΔCoVaR in a multivariate framework. We then highlight some study properties of these networks by considering two different parametric settings, characterized by a multivariate Gaussian and t-Student distribution, respectively. In particular, we describe the relationship of ΔCoVaR networks with partial correlation networks (that in the Gaussian case can be obtained by a simple rescaling), and we compute analytical the conditional quantile functions under both distributions, showing that in the latter case the use of linear approximations of the function leads to an underestimation of risk.

We then propose an estimation procedure based on SCAD penalized quantile regression and focused on the estimation of the tails of the distribution.

Finally, we propose an empirical application focused on the European banking system, where we estimate network-ΔCoVaR, we analyse the properties of the network and we compare it to the median QGM (i.e. the network the reflects the conditional medians). We found that ΔCoVaR networks are characterized by a stronger interconnectivity, and a weaker community structure. From a regulators' perspective, these results highlight how conditional tail risk may represent a channel of transmission of systemic shocks, and how network-ΔCoVaR provides different information compared to median quantile regressions and partial correlation networks. Being based on comovement of time series, network-ΔCoVaR does not allow to isolate the mechanisms for the transmission of contagion. However, by allowing to map the channels of tail risk transmission, and to assess the relevance of each institution in the system, may represent powerful tools for regulators, complementar to other indicators of systemic relevance and fragility.

This work contributes to the toolbox of financial regulators by providing a methodology for the construction of market-based tail risk networks, and by suggesting an approach to integrate balance sheet information to market's view on interconnectedness in situation of distress.

Chapter 5

Liquidity contagion in banking networks with community structure

Acknowledgement

This chapter is partially based on the conference proceeding "Financial Contagion in International Banking Networks: A Simulation Study", presented at the 16th International Conference On Finance And Banking, Ostrava, Czech Republic, co-authored by Rosella Giacometti (Torri and Giacometti, 2017), and the proceeding "Systemic Risk and Community Structure in the European Banking System" presented at the MME2017 conference, Hradec Králové, Czech Republic (Torri, 2017).

5.1 Introduction

One of the most striking features of financial crises is that they often spread quickly across countries and institutions, and that small shocks affecting a particular region or a small group of institutions can cause contagion in the rest of the system and infect other economic sectors. The focus on contagion become a fundamental aspect of the analysis of systemic risk, especially after the 2008 crisis. We can refer for instance at the definition of systemic risk provided by Schwarcz, 2008, that underline the presence of transmission mechanism that spread and boost the diffusion of distress in the system. Due to the focus on risk transmission, network theory is a powerful tool, and network-based modeling approach are now increasingly common, and have been extensively applied to model default contagion and liquidity crises in banking systems, see for instance Eisenberg and Noe, 2001 and Hurd, 2016. Within this framework, a fundamental

research question regards the study the effect of the structural properties of the network on the diffusion of financial contagion. Hurd, 2016 study the effect of assortativity on the diffusion of contagion, Gai, Haldane, and Kapadia, 2011 and Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015 analyse the relationship between contagion and the level of network connectivity. Other studies highlighted the role of heterogeneity in the system (e.g. Amini, Cont, and Minca, 2016), information asymmetries Battiston et al., 2012b and moral hazard (Brusco and Castiglionesi, 2007). Overall, the relationship between structural properties and stability is complex and non linear and tend to show a *robust-yet-fragile* configuration (Hurd, 2016; Chinazzi and Fagiolo, 2015).

Here, we focus on a particular network property that has not been studied extensively in the literature: the presence of communities, that are groups of banks with stronger relationships among each other compared to the rest of the element in the system. In an international banking system, we expect these communities to reflect the grouping of banks in countries and geographical areas. Focusing on the European context, the presence of national communities is indeed a rather intuitive feature of a banking system where the merging of the regulatory systems is still an ongoing process and is composed by national banking systems with very a long history. This in turns requires researchers to put greater attention to the international dimension of the network, as the regulation and the supervision is now increasingly centralized and harmonized, and market practices are leading to a greater market integration. Indeed, the available literature shows that the European banking system is characterized by a complex and stratified structure, with a large group of interconnected banks and a high level of interconnectedness (Aldasoro and Alves, 2018; Craig and Saldías, 2016). Still, the studies that focus on international banking systems are limited compared to the ones that study national sectors, probably due to the limited availability of the required datasets, that are typically undisclosed also for researchers.

Several studies show that the presence of communities can influence the diffusion of epidemic diseases among people or rumors in online social networks (e.g. Stegehuis, Hofstad, and Leeuwaarden, 2016 and Salathé and Jones, 2010). Here we fill a gap in the literature on financial contagion by addressing the same question in the context of liquidity banking crises, adopting a simulation scheme similar to Gai, Haldane, and Kapadia, 2011 and modeling the presence of community in the simulated networks in a realistic way, and we connect the results to the network literature developed in other fields. The Chapter is structured as follows: Section 5.2 presents a motivating example for the analysis, based on the partial correlation networks constructed using the *tlasso*

model, Section 5.3 discusses the role of community structure in process diffusion over network, not necessarily related to the financial setting, Section 5.4 describes the characteristics of the model, the structure of the simulated balance sheet, the structure of the simulated networks and the contagion dynamics. Section 5.5 describes the simulation study, presents the results and introduces different policy simulations. Section 5.7 concludes.

5.2 Community detection in partial correlation banking networks in Europe

Here, in a setting similar to Chapter 3, we analyse the presence of a community structure in the European banking system. Using the *tlasso* model, we construct a partial correlation network among large European banks from CDS (credit default swaps) data. We then identify the community structure using a community detection algorithm and we analyse its evolution over time.

Following a relevant stream of literature, instead of focusing on the real exposured between banks to construct the network, we infer the network structure using tools that can capture co-movements and dependence patterns between financial time series to establish the existence of links among banks, assuming that markets are efficients, prices should provide useful information about the interconnection structure. An example of this is the estimation of credit risk networks from credit default swap (CDS) spreads or equity price times series (e.g. Puliga, Caldarelli, and Battiston, 2014, Anufriev and Panchenko, 2015 and Billio et al., 2012). One major advantage of these approaches is that they rely on public data and well known statistical modelling techniques.

5.2.1 Dataset

The dataset consists of 31 weekly time series of CDS (5 years maturity) of European financial institutions settled in 12 countries. They refer to CDS spreads quoted in Euro and they span the time period from January 2009 to June 2016. 20 of the banks in the sample belong to countries in the Eurozone, the other 11 are located in the United Kingdom, Sweden and Denmark. We observe that our database includes 85% of the banks with total assets over 500 billions that are under the European Central Bank (ECB) supervision and it is also consistent with the European Banking Authority (EBA) stress-test exercise 2016, representing 47% of the banks involved. For

the analysis we consider the log-differences of CDS spreads, computing the partial correlation matrix from them using *tlasso* algorithm. We first estimate the network using the data of the entire sample period, and then we analyse the evolution over time using a rolling analysis using windows of 100 weekly observations each.

5.2.2 Community structure and community detection algorithm

A *community* in the field of complex network can be defined as a group of nodes that are more densely connected among themselves than with nodes outside the group. The problem of identifying the best community structure is well studied in the network literature (see for instance Fortunato, 2010).[1] We consider an optimization-based approach in which the optimal community structure is the one associated with the highest *modularity* (Newman, 2004), a quantity defined as follows. Given a partition $G = \{G_1, \ldots, G_p\}$ the modularity Q is:

$$Q = \frac{1}{2m} \sum_{i,j} \left(a_{ij} - \frac{s_i s_j}{2m} \right) \mathbb{I}_{[g_i = g_j]}, \tag{5.1}$$

where a_{ij} is an element of the adjacency matrix $\boldsymbol{A}$, s_i is the strength of node i, $m = \frac{1}{2} \sum_{i,j} a_{i,j}$, g_i is the group in the partition in which the element i belongs and $\mathbb{I}_{[g_i = g_j]}$ is 1 if $g_i = g_j$ and 0 otherwise. Modularity can assume values between -1 and 1, with positive and high values denoting a good division of the network into communities. The procedure proposed in Newman, 2004 identifies the optimal partition using a greedy optimization that, starting with each vertex being the unique member of a community, repeatedly joins together the two communities whose amalgamation produces the largest increase in modularity. This approach can be implemented efficiently on large networks and identifies automatically the optimal number of communities. Note that a positive value of modularity is not a sufficient condition for identifying a network divided in communities, therefore we need to test if it is the modularity is statistically significantly higher than the one of a random network. In particular we generate the random networks using a degree-preserving rewiring procedure Fortunato, 2010.

[1] In the case of non-overlapping communities we can refer to the optimal community structure as *optimal partition*.

5.2.3 Empirical results

Static analysis

Figure 5.1 shows the network represented using a force layout. The visual inspection denotes the clustering of banks in communities aligned with the national groups.

Table 5.1 shows the composition of the optimal communities identified by Newman's algorithm. The partition consists in five communities and it is possible to notice that it roughly overlaps with geographical divisions, confirming the results of the visual inspection of Figure 5.1. In particular, community 1 is composed uniquely by banks from Mediterranean Countries, community 2 by British and German banks, communities 3 and 4 include a more diversified group of banks from United Kingdom (UK), central and northern Europe and finally community 5 is composed by French banks (with the exception *BNL*, that is part of the French group *BNP Paribas* but is Italian).

Table 5.2 reports the value of modularity of the optimal community structure compared to two geographical partitions, one obtained grouping banks by country and one by grouping them in three broad geographical areas: Southern Europe, Central Europe and countries outside Eurozone. For comparison we also consider a partition based on the size of the banks, to check whether banks of similar dimension tend to connect to each other.[2] For each indicator we compute a confidence level based on the empirical distribution of the indicator computed on 1000 random rewirings of the network. We observe that the modularity of the optimal partition is equal to 0.461 and statistically significantly different from the null model with a confidence level higher than 99%, confirming that the network is characterized by a relevant division in communities. We also see that the modularity of geographical partitions (0.352 and 0.335 for the country partition and the area partition respectively), although smaller than the optimal one, are rather high and statistically significant, indicating that the geographical divisions represent a relevant feature of the banking network. Concerning the partition by size, although the modularity is positive and statistically significant, it has a much smaller value compared to the other partitions (0.087), suggesting that is a less relevant factor.

[2] In particular we defined 5 classes of homogeneous size based on the *total assets* based on 2015 balance sheet.

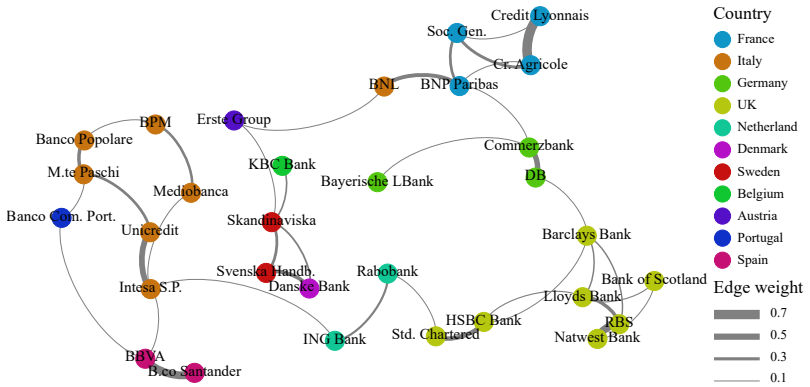


FIGURE 5.1: Graphical representation of the network with a force layout.

Community	Number of banks	Constituents
1	9	M.te Paschi(IT), Banco Popolare(IT), BPM(IT), Intesa S.P.(IT), Mediobanca(IT), Unicredit(IT), B.co Santander(ES), BBVA(ES), Banco Com. Port.(PT)
2	7	Commerzbank(DE), DB(DE), Barclays Bank(UK), Bank of Scotland(UK), Lloyds Bank(UK), Natwest Bank(UK), RBS(UK)
3	6	KBC Bank(BE), Bayerische LBank(DE), Erste Group(AUT), Skandinaviska(SE), Svenska Handb.(SE), Danske Bank(DK)
4	4	Rabobank(NL), ING Bank(NL), HSBC Bank(UK), Std. Chartered(UK)
5	5	BNL(IT), BNP Paribas(FR), Cr. Agricole(FR), Credit Lyonnais(FR), Soc. Gen.(FR)

TABLE 5.1: Constituents of Optimal Communities.

Dynamic analysis

We perform a rolling analysis to monitor the evolution of the community structure over time. In particular, we consider the evolution of modularity as presented in Figure 5.2. We see that the modularity of the optimal partition is highest in the period corresponding to the Sovereign crisis, decreases from mid-2012 and then grows again in recent years. The pattern is similar

Partition	Modularity
Optimal partition	0.461***
Countries	0.352***
Geographical area	0.335***
Total assets	0.087***

TABLE 5.2: Modularity of 4 partitions. ***,**,* refer to confidence level of 99%, 95% and 90% respectively.

for the geographical partitions, while modularity of the partition generated by the size show a moderate increase across the time period. The high level of modularity during the crisis is consistent with the sovereign-driven nature of the European crisis: the increased relevance of country risk leads to a decrease in confidence in the transnational interbank market, and thus to a "flight to safety"and a tightening of national banking systems. The rise in modularity in the last part of the sample may be related to the low level of the interbank interest rates in recent years, that makes less convenient for banks in core countries to lend to banks in peripheral countries, exacerbating the division among national banking systems. In an unreported test, we also measure the stability of the community structure by testing how well the optimal partition in a given estimation window can describe the community structure in a future window. We found that the optimal partition, despite the variations in the modularity over time, is characterized by a great stability. The results are available from the author upon request.

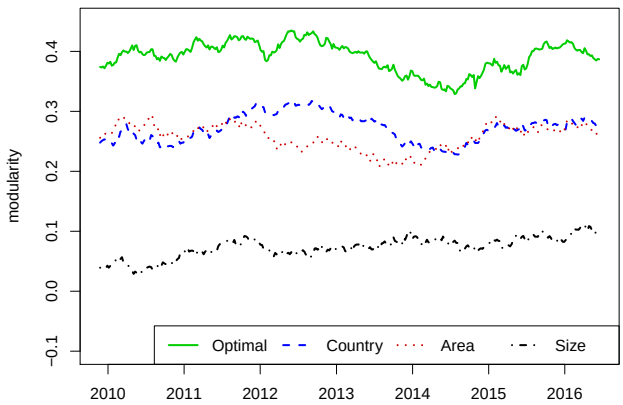


FIGURE 5.2: Evolution of Modularity over Time for Different Partitions.

5.3 Epidemic diffusion in networks with communities

The study of dynamic process on network is of great interest in many scientific fields, including, among the others, the diffusion of diseases, information spread in social network, and adoption of innovations of behaviour (Pastor-Satorras et al., 2015). The characteristics of the diffusion depends on the specificities of the process, as well as the structural properties of the network (Granovetter, 1977; Watts and Strogatz, 1998; Centola and Macy, 2007). The presence of a community structure is known to influence the diffusion of dynamic process, and the effects strongly depends on the nature of the process (see e.g. Fortunato, 2010; Salathé and Jones, 2010). The evidence shows that a community structure can either enforce as well as inhibit diffusion processes (Stegehuis, Hofstad, and Leeuwaarden, 2016), it is therefore of primary importance to study the effect of the community structure in relationship to the specific characteristics of the process and the topology of the network.

Starting from the seminal work of Granovetter, 1977, that describes the role of "weak ties" in the spread of processes in social networks, a large part of the literature focuses on contagion models where processes can diffuse by means of a single contacts between two nodes. In such models, edges connecting distant nodes (i.e. the "weak ties") are the primary channels for the diffusion of process, while connections among dense groups of nodes in the networks are not as relevant, since they represent redundant connections and therefore contribute less to the overall diffusion of the process. Examples of these *simple contagion models* are the susceptible-infected (SI) contagion model and the susceptible-infected-recovered (SIR) model, where a node can be infected as long as it is exposed to at least one affected node (for an overview see Newman, 2010). Such forms of diffusion processes may benefit from the presence of ties among distant nodes in a network, that work as *bridges* for a faster transmission of the process. The presence of tight communities on the other hand, leads to a reduction in the transmission effectiveness by creating redundant patterns between nodes in the same communities, and reducing the number of edges across distant nodes in the network. As an example, Huang and Li, 2007 show how in a SI model, in networks with scale-free distributions the presence of a strong community structure can reduce the danger brought by epidemic prevalence in the network. The interpretation is that the outbreak of a shock tends to be confined in a community, reducing its spread to the rest of the network. Similar results have been obtained by Salathé and Jones, 2010 considering a SIR model, showing how community structure has a hindering effect on spreading of contagious

diseases in human populations, and that target vaccination to individuals bridging communities can be effective in containing the spread of diseases.

Centola and Macy, 2007 study a different form of dynamic processes on networks, that are defined *complex contagion*. In this case a node can be affected only if a certain number (or a certain percentage) of its neighbours is affected, requiring therefore multiple channels for the diffusion of contagion. In this case, the presence of reduntant patterns is necessary for the transmission of the process, and therefore netoworks with high level of clustering coefficient allow a more effective transmission of contagion. They indeed show that the most effective networks for the transmission of complex contagion are regular lattices, while random graphs with weak ties do not allow the transmission of contagion (Centola, Eguíluz, and Macy, 2007). Moreover, Reid and Hurley, 2011 show that complex contagion can spread faster in networks with overlapping community structure, and O'Sullivan et al., 2015 develop an approximation method that allow to study the effect of clustering on the diffusion of complex contagion, confirming the results of Centola and Macy, 2007.

In the analysis of social networks, complex contagion has an explanation in terms of social reinforcement behaviours, in which an individual is likely to adopt an habit if several of his neighbours do it. For instance, Mønsted et al., 2017 study the diffusion of information in social media, proposing an experiment using Twitter bots, and showing that information diffusion can be better described using complex, instead of simple contagion models.

To our knowledge, the concept of complex contagion has not been explicitly applied in the financial literature, however, several financial contagion models may be considered similar to complex contagion, as a single exposure to a distressed institution may not be sufficient to cause the transmission of contagion. The DebtRank methodology proposed by Battiston et al., 2012a, for instance, consists in a contagion model in a financial system, where an institution defaults when losses exceed the capital, transmitting then the distress to other institutions. A single default of a debtor may not be sufficient to cause an institution to go in distress, due to the presence of capital buffers, while multiple debtors' defaults can instead facilitate the transmission of contagion (Battiston et al., 2012a). Similarly, in other contagion models such as Gai and Kapadia, 2010; Gai, Haldane, and Kapadia, 2011; Elsinger et al., 2013 the distress transmission can be fostered by the presence of redundant paths, as multiple distressed neighbours may weaken the protection of capital or liquidity buffers.

In a network with a strong community structure, the relatively high density within communities,

compared to a random network, may increase the ability of the distress to grow and reach a relevant dimension, while in a more randomized network the shock would disperse more easily. On the other hand, the separation of communities may reduce the ability to spread to the entire network.

The balance of these two aspects make the study of diffusion of complex contagion in network with communities particularly challenging, also because many of the common approximations used for the derivation of analytical results require assumptions not compatible with the presence of a community structure (O'Sullivan et al., 2015), and since in financial application it is well known how the introduction of heterogeneity among nodes is crucial for the analysis of contagion mechanisms, reducing the ability to find analytical solutions (Chinazzi and Fagiolo, 2015; Amini, Cont, and Minca, 2016).

Numerical simulations are powerful tools for the analysis of financial contagion, and in the next section we propose a study based on the liquidity contagion framework of Gai, Haldane, and Kapadia, 2011, where we isolate the role of community structure while controlling for other variables in the system and including some level of heterogeneity in the network construction.

5.4 Interbank contagion modelization

The contagion model that we adopt is largely similar to the one proposed by Gai, Haldane, and Kapadia, 2011, and allows us to model the diffusion of liquidity shocks in a banking system. In this Section we present the properties of the model, focusing on three main aspects: the network, the construction of banksâĂŹ balance sheets and the contagion dynamics.

5.4.1 Interbank network simulation

The interbank system is constructed as a directed network in which the edges represent bilateral exposures between two institutions. Here we construct the network as random graphs. The focus of the study is on the presence of a community structure; to model the communities in the network we proceed as follows: first we construct m separate networks (that will be the communities), each of them characterized by the same degree distribution, and then we progressively rewire the communities, blending them together. The result is a set of networks with a progressively weaker community structure. The rewiring procedure allows to maintain several properties of

the network, in particular the degree distribution, allowing to compare the results of networks with different rewiring.[3] We consider three different network specifications characterized by different degree distributions: in the first specification the degree distribution of the nodes follow a Poisson distribution (*Poisson network*), in the second the degree distribution follows a power law (*power-law network*) and in the last specification the distribution is different for the in-degree (Poisson) and for the out-degree (power-law) (*asymmetric network*).[4]

Power law and Poisson networks are known to be characterized by rather different behaviours in presence of crisis: typically power-law networks are more resilient to random failures, but are more fragile to specific failures to central nodes, showing a robust-yet-fragile behaviour. The third network model introduces an asymmetry in the degree distributions and allows to isolate the concentration of assets on the creditor-side while maintaining a relative uniform distribution of the degree for the debtors. Overall, the three models should allow to assess the effect of the degree distribution and allow to obtain robust results. Moreover, the framework is flexible enough to easily accommodate other network configurations. The networks are represented using the adjacency matrix $\boldsymbol{A}$, that is an $n \times n$ square matrix where each element a_{ij} is equal to one if it exists a credit exposure between node i and node j, and zero otherwise. We also define the matrix $\mathbf{Q}$, that is a right stochastic matrix (i.e. a matrix in which all the rows sum to one) in which the elements $q_{ij} = \frac{a_{ij}}{k_j}$, where k_j is the in-degree of element j. in matrix form we can write:

$$\mathbf{Q} = \boldsymbol{D_A} \boldsymbol{A}, \tag{5.2}$$

where $\boldsymbol{D_A} = \text{diag}(\frac{1}{k_1}, \ldots, \frac{1}{k_p})$ is the diagonal matrix composed by the reciprocal of the degrees of the nodes in the network.

Network indicators

In Section 5.5.2 we test whether the introduction of liquidity requirement policies based on network measures influence the diffusion of distress. In particular, we consider the following centrality measures:

- **In-degree** and **out-degree**. Degree is one of the simplest centrality measures, and is computed as the number of edges connected to a node. In signed network in-degree is the

[3]For more on the rewiring procedure see for instance Fortunato, 2010.

[4]The Poisson networks have been generated using the classical ErdősâĂŞRényi model, while the scale-free and asymmetric networks with the fitness model proposed by Goh, Kahng, and Kim, 2001.

number of ingoing edges (liabilities in the interbank market) and out-degree is the number of outgoing edges (assets in interbank market).

- **Eigenvector centrality**. It is a centrality measures that expands degree centrality and accounts for the fact that connections to important nodes should be more important than connections to secondary nodes.

- **Betweenness centrality**. Measure that counts the number of shortest paths between each pair of nodes. Nodes with a high betweenness centrality may have a particular important role in the diffusion of distress, since they represent âĂIJbridgesâĂİ between different communities in the network.

5.4.2 Balance sheet

The structure of the balance sheet is similar to the one proposed by Gai, Haldane, and Kapadia, 2011, and although it makes some simplifications, most importantly the size homogeneity among banks, it allows to isolate the network effects while maintaining some realistic features, such as the option for the banks to obtain liquidity by entering (repurchasement agreements) repo transactions. The liabilities are composed of unsecured liabilities in the interbank market L_i^{IB}; repo contracts L_i^R; retail deposit L_i^D; and capital K_i. The assets are composed by unsecured assets in the interbank market A_i^{IB}; fixed assets such as mortgages and corporate loans A_i^F; reverse repo assets (collateralized lending) A_i^{RR}; assets which may be used as collateral in repo transactions (collateral assets) A_i^C; and fully liquid assets A_i^L. The interbank assets and liabilities are the actual exposures in the network. In order to simplify the analysis and to focus on the network effects, we assign to each bank the same amount of liability in the interbank market (therefore $L_i^{IB} = L^{IB} \quad \forall i$). The interbank assets are then computed endogenously for each bank according to the connections in the network. The balance sheet items, together with the baseline calibration values, are shown in Table 5.3.

5.4.3 Liquidity shortage and contagion mechanism

Our framework studies the diffusion of liquidity shocks. A liquidity shock happens when a bank cannot temporary fulfill its duties due to an idiosyncratic shock, a change in repo haircuts or the hoarding of liquidity from its creditors in the interbank market. The general idea behind the

Parameter	Description	Baseline calibration
Liabilities		
L_i^{IB}	Unsecured interbank liabilities	15% of balance sheet
L_i^R	Repo liabilities (i.e. borrowing secured with collateral)	20% of balance sheet
L_i^D	Retail deposits	Endogenous (balancing item)
K_i	Capital	4% of balance sheet
Assets		
A_i^{IB}	Unsecured interbank assets i	Endogenous (depending on network)
A_i^F	Fixed assets (e.g. individual corporate loans or mortgages)	Endogenous (depending on A_i^{IB})
A_i^C	assets that can be used as collateral in repo transactions	10% of balance sheet
A_i^{RR}	Reverse repo assets (i.e. collateralized lending)	11% of balance sheet
A_i^L	Fully liquid assets	1.2% of balance sheet
ε_i	Idiosyncratic liquidity shock	0.2 for $i = 1$; 0 oth.
Network		
n	Number of banks in the system	100
z	Average in- and out-degree of the nodes	10
c	Number of banks in each community	20

TABLE 5.3: Balance Sheet Items and Baseline Calibration.

contagion mechanism is that, under distress, the banks act preemptively and hoard liquidity from the interbank market shrinking their balance sheets. This is turn may cause its counterparties in the market to have a liquidity shortage and to start a chain reaction.

The liquidity of a bank i is composed by its liquid assets A_i^L and by the funds that can be obtained entering a repo transaction using as a collateral A_i^C and $\frac{A_i^{RR}}{1-h}$ (note that the repo transaction requires an haircut, therefore the liquidity raised is multiplied by $1 - h - h_i$).[5] The total liquidity is then composed by three items: the value of the repo liabilities L_i^R, the amount of uninsured liabilities withdrawn from the creditors in the interbank market $(\mathbf{Q}'\lambda)L_i^{IB}$ (more on this later), and an idiosyncratic shock ε_i. We do not specify the cause of the shock, since we are only

[5]The haircut refers to the difference between the amount of cash lent on a repo transaction and the value of the collateral. It reflects the liquidity and default risk that the lender assumes. Here it is assumed that the total haircut depends on a common component h and an idiosyncratic component h_i specific for every bank.

interested to the effect on the liquidity position. Formally we can write:

$$l_i = A_i^L + (1 - h - h_i)(A_i^C + \frac{A_i^{RR}}{1-h}) - L_i^R - (\mathbf{Q}'\lambda)L_i^{IB} - \varepsilon_i \qquad i = 1, \ldots, n, \tag{5.3}$$

where the balance sheet items are defined as in Table 5.3.[6] A bank i suffers a liquidity crisis when $l_i < 0$. For convenience we define the vector e such that $e_i = A_i^L + (1-h-h_i)(A_i^C + \frac{A_i^{RR}}{1-h}) - L_i^R$. Then, we can rewrite Equation 5.3 as:

$$l_i = e_i - (\mathbf{Q}'\lambda)L_i^{IB} - \varepsilon_i \qquad i = 1, \ldots, n, \tag{5.4}$$

where $\mathbf{Q}$ is the right stochastic $n \times n$ matrix that represents the interbank market (see Section 5.4.1) and λ, a vector of length n, denotes the percentage of liquidity that each bank withdraws from the market as a consequence of a liquidity crisis. $(\mathbf{Q}'\lambda)_i$ is therefore the percentage of interbank liabilities of bank i withdrawn by all its creditors.

The values in the vector λ are crucial for the spreading of contagion in the market: the higher they are, the larger the possibility to start a chain reaction and infect other subjects. The upper bound for each element λ_i is 1, that is, a distressed bank withdraws all its asset in the interbank market. On the other hand, the lower bound consists in the case in which a distressed bank withdraws only the minimum amount of assets required to face its liquidity shortage. It is reasonable to assume that a bank witnessing a liquidity shortage withdraws precautionary more than the minimum required, fearing further liquidity problems in the future. Gai, Haldane, and Kapadia, 2011 propose to use a value of $\lambda_i = 1$ for each bank with a liquidity shortage. Instead, we consider an milder approach, and in our specification a distressed bank withdraws assets from the interbank market in a quantity proportional to the liquidity shortage with a multiplicative factor $\lambda^0 > 1$. This difference allows to model the cascade more realistically, as the assumption of withdraw of the entire liquidity would very often result in extreme liquidity crises even in case of minor liquidity shortage, while in reality policy intervention and proactive reactions by the banks may limit the extent of the crisis. We also impose that $\lambda_i \leq 1 \ \forall i$, so that a bank cannot withdraw

[6]Notice that in our model we implicitly assume that the banks cannot obtain new liquidity by selling its fixed assets or withdraw lending to the real economy. This assumption is motivated on the fact that banks may prefer to hoard liquidity from the interbank market since it is faster and less visible to other market participants (Gai, Haldane, and Kapadia, 2011).

more that its total assets invested in the interbank market:

$$\lambda_i = \min\left(\frac{-\min(l_i, 0)}{A_i^{IB}}\lambda^0, 1\right). \tag{5.5}$$

In our baseline specification we set $\lambda^0 = 2$ (i.e. a bank hoard from the interbank market twice the funds required to cover its liquidity shortage). It is clear that, if a bank with a liquidity shortage withdraws enough assets from the interbank market, may make another bank enter a liquidity crisis itself. The second bank will have to hoard liquidity itself, and may potentially trigger a chain reaction in the network. In Section 5.5 we also consider alternative values for λ^0, in order to assess the sensitivity of the model to this parameter.

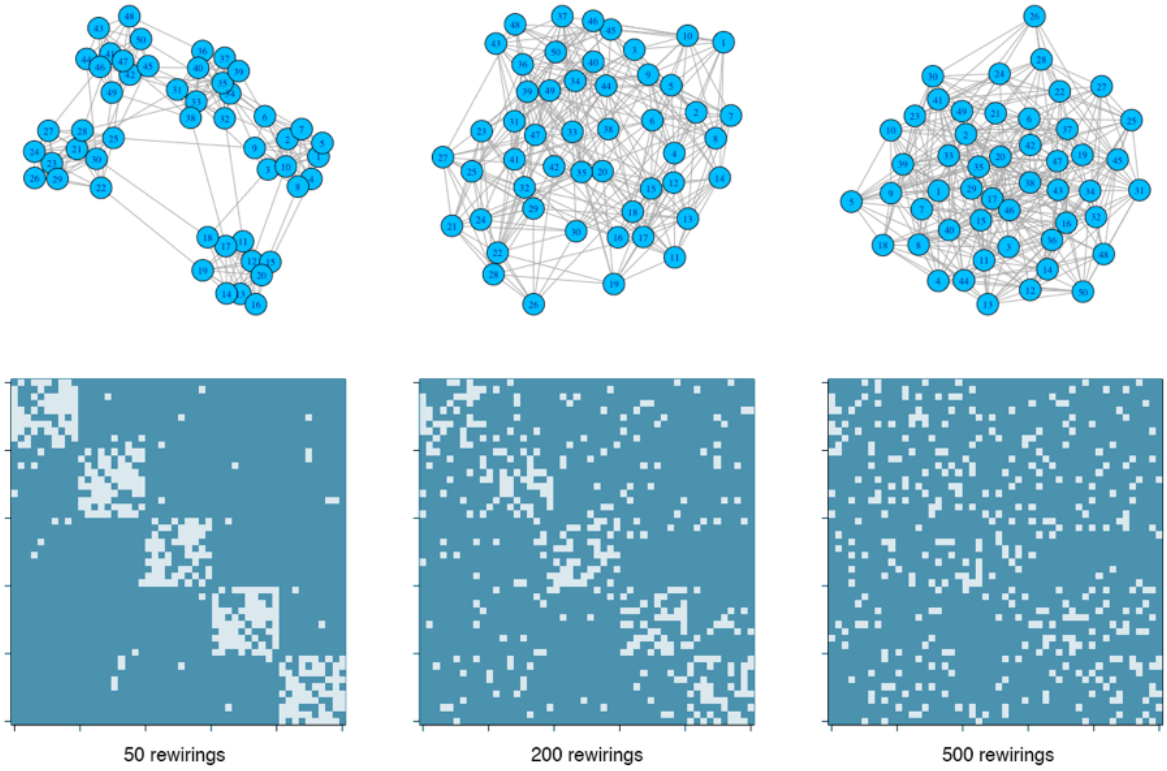
Note that in our framework the transmission of distress is related to liquidity hoarding and not to the presence of defaulting counterparts. This approach, in which the distress moves upstream from creditors to debtors is symmetric to default-based contagion mechanisms such as Eisenberg and Noe, 2001 or Gai and Kapadia, 2010.

5.5 Simulation study

The main goal of this analysis is to assess the effects of the presence of a community structure on the diffusion of liquidity shocks. To isolate the effect of this topological feature, we generate random networks and we control the level of community structure using a rewiring procedure: starting from networks composed by completely separated communities, we progressively "shuffle" the edges in order to remove the community structure while maintaining some properties such as the degree of individual nodes (see e.g. Fortunato, 2010). we consider for each experiment 15 different levels of rewiring, starting from networks with completely separated communities, to ones with almost no community structure. Figure 5.3 reports a representation of the rewiring process for a Poisson network.

The contagion is obtained by applying an idiosyncratic liquidity shock to one of the banks. We point out that the main source of stochasticity in the model is related to the network structure (and by the random selection of the starting node), while the contagion diffusion process is deterministic. The balance sheet structure is also fixed, with the exception of the interbank assets, that are determined endogenously by the network structure.

FIGURE 5.3: Graphical representation of the rewiring process using a force-layout (first row) and the corresponding adjacency matrices (second row).



We test the level of contagion by measuring the average ratio of banks that face a liquidity shortage and the extent of systemic hoarding (i.e. the total amount of interbank assets frozen during the crisis). We are also interested in understanding the extent to which the distress propagates outside the community of the bank shocked, therefore we decompose the first indicator in two parts: one accounts for the banks in the shocked community, the other to the ones in the rest of the network. The results are obtained on the basis of 500 simulation runs for each of the setting.

5.5.1 Scenario simulation

In this simulation setting we assume that the network is composed by five different communities, each of them composed by twenty banks. The community structure is progressively removed using a rewiring procedure, with 15 rounds of rewiring, where in each round 50 rewirings are done (i.e. for 50 times we take two random edges and we swap the destination node). The rewiring procedure does not influence the degree of the nodes, therefore we obtain a set of networks whose community structure is progressively weaker, but with comparable characteristics.

The values of balance sheet items are the ones in Table 5.3, and we repeat the experiment using networks with different degree distributions (the *Poisson*, *power-law* and *asymmetric networks*; see Section 5.4.1). For robustness, we repeated the experiments with networks with different density, characterized by nodes with in-degree equal to 10 and 6 respectively.

We model a liquidity shock for one of the banks by introducing an idiosyncratic shock $\varepsilon = 0.2$ and we simulate the diffusion of contagion over the network.

Figure 5.4 reports the percentage of distressed banks in the system for the six network specifications and different levels of rewiring (in each Panel, from left to right the networks are increasingly randomized. First row represents the networks with in-degree = 10 and second row in-degree = 6). It is possible to see that in most of the setting the number of distressed banks is higher for networks with a stronger community structure. More precisely, the percentage of distressed banks is relatively smaller in the non-rewired network (since the contagion cannot spread outside the first community), then it grows for moderate level of rewiring and finally reduces for high level of rewiring. It is interesting to notice that, in networks with communities, the number of distressed banks is higher not only in the shocked community, but also in the rest of the network, meaning that the contagion spreads faster also *outside* the community where the first shock hit. moreover, while the number of distressed banks in the shocked community decreases with the level of rewiring, the effect in the rest of the network does not change monotonically with level of rewiring and is maximum for a limited number of rewiring. These results, consistent for different network configurations and densities, suggests that the presence of communities fosters the diffusion of distress and amplifies more the initial shock.

An explanation to the results is that the distress can increase its extent by reverberating within a community, and then diffuse in the rest of the network once it assumes a larger size. The results are confirmed by Figure 5.5, that reports the quartiles of the distribution of level of distress introduced in the system by the initial shock. We see that the distress is generally higher for networks with tight communities, both in terms of median value across the iteration and the lowest quartile.

The comparison between banks with different density (first vs. second row of Figures 5.4 and 5.5), shows that the community structure is even more relevant in sparser networks: for the *Poisson* and *scale-free* networks the contagion spreads much faster in networks with a strong community structure compared to the denser networks. This can be explained by the fact that denser networks are typically more resilient to contagion, making more difficult the development

of a chain reaction. The only exception to the previous results is the *asymmetric* and less dense networks, where the community structure does not seem to influence the diffusion of contagion, probably due to the superior resilience to shocks of this particular network configuration.

5.5.2 Policy experiments

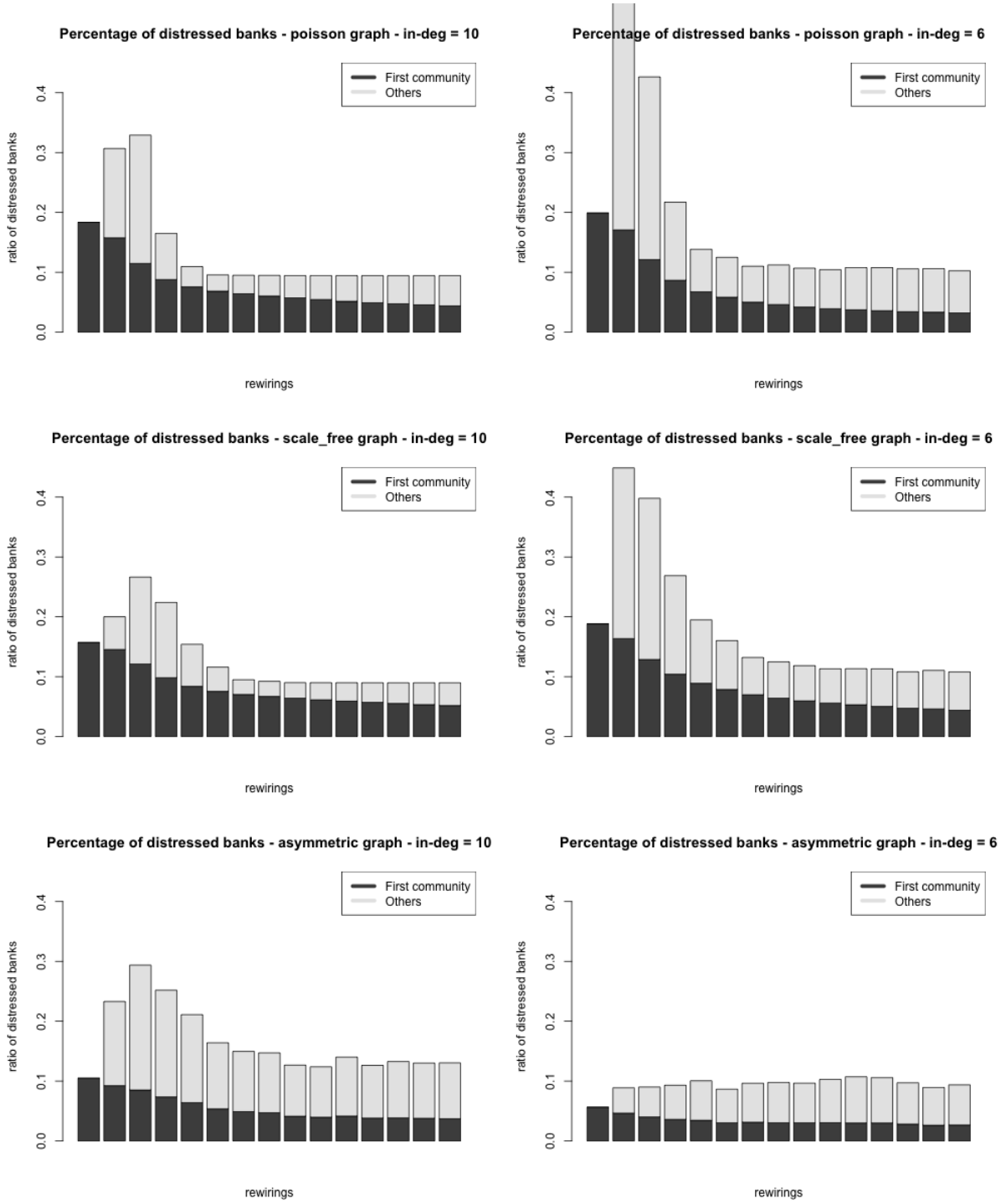
In this section we conduct four policy experiments, testing the effect of different possible precautionary measures to reduce the diffusion of liquidity crises. The experiments that we propose are general in nature, and consist in direct changes of the model parameters. As we will point out however, such changes can be related to specific policies, and the results may help to highlight some stylized facts and to better assess real world scenarios. Examples are the introduction of *liquidity coverage ratio* (LCR) and the liquidity risk monitoring tools introduced in the Basel III framework as a response to the great financial crisis of 2007-08 (see Basel III, 2013).

The first experiment consists in a change ofthe percentage of liquid assets in the balance sheet of every bank, corresponding to a change in the liquidity requirements for the banks.

In particular, we test four different levels: $A^L_{ij} = \{1.1\%, 1.2\%, 1.3\%, 1.4\%, 1.5\%\}\ \forall i$, (the baseline value was 1.2%). The results are presented in Figure 5.6 (first column). In all the settings, the number of distressed banks decreases monotonically with the liquid assets A^L_i, meaning that a higher level of liquidity helps to contain liquidity crises, especially for the networks with a stronger community structure (on the left part of the graph). This experiment denotes a high sensitivity of the contagion effect to the capital liquidity, especially in the case of network with a strong community structure. In a more realistic setting this suggests to pose a particular attention to the level of liquidity in the system, especially in a network with communities.

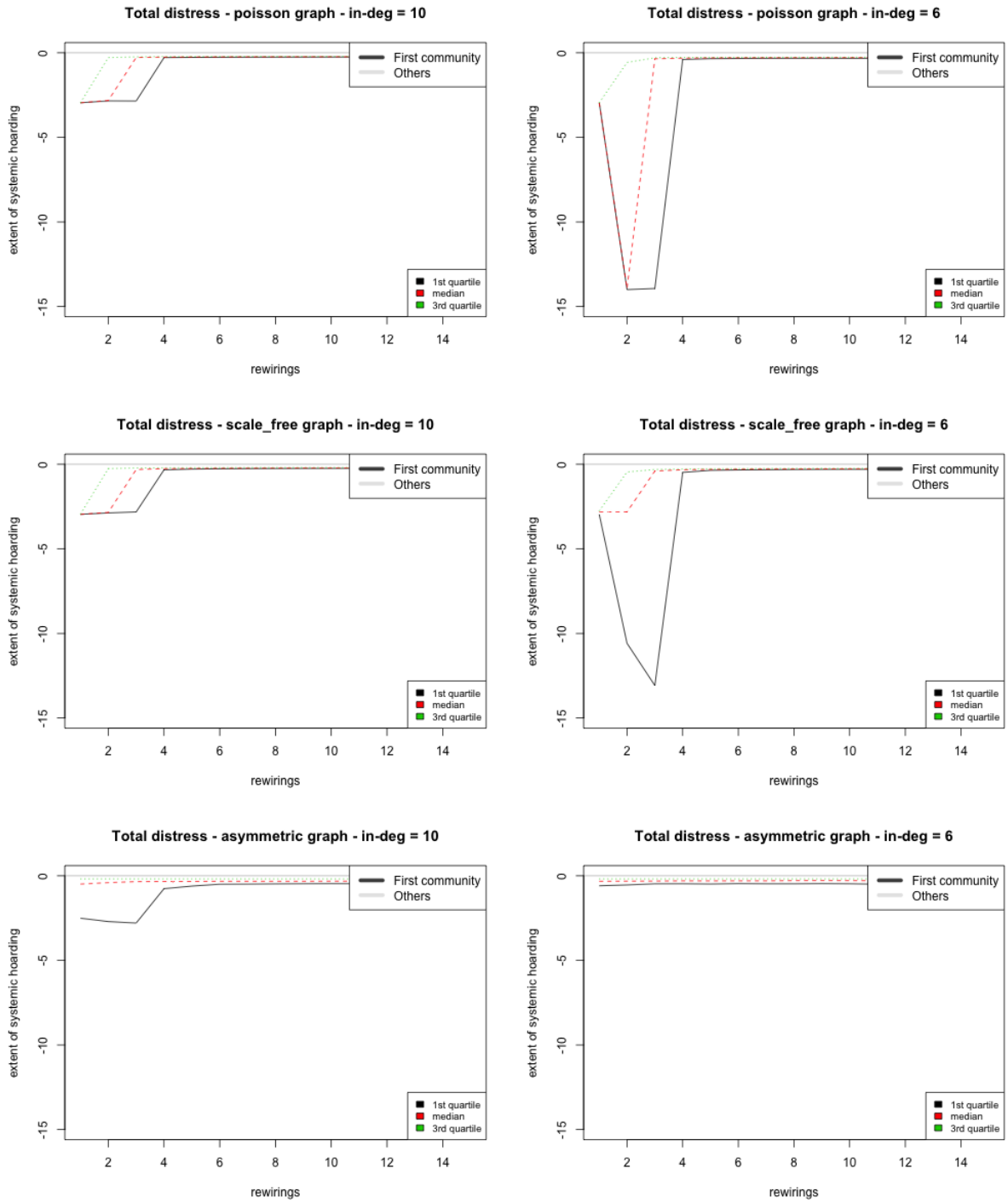
In the second policy exercise, we change the liquid assets only to the banks in the country where the exogenous shock hits, leaving the ones for the other banks to the baseline level of 1.2%. The results are reported in Figure 5.6 (second column). We see that the results are similar to the ones obtained for the previous policy experiment, especially in networks with a stronger community structures. This means that in order to prevent contagion dynamics, an intervention on the banks in the community where the liquidity crisis starts is particularly effective. Rather intutively, the real-world implication would be that in a trans-national context, nation-wide interventions may be a good strategy to contain the spread of a liquidity contagion.

FIGURE 5.4: Average percentage of the banks in the network who suffer liquidity shortages. x axes represent the level of rewiring. Left column: *Poisson*, *scale free* and *asymmetric* network with average in- and out-degree = 10. Right column: networks with average in- and out-degree = 6.



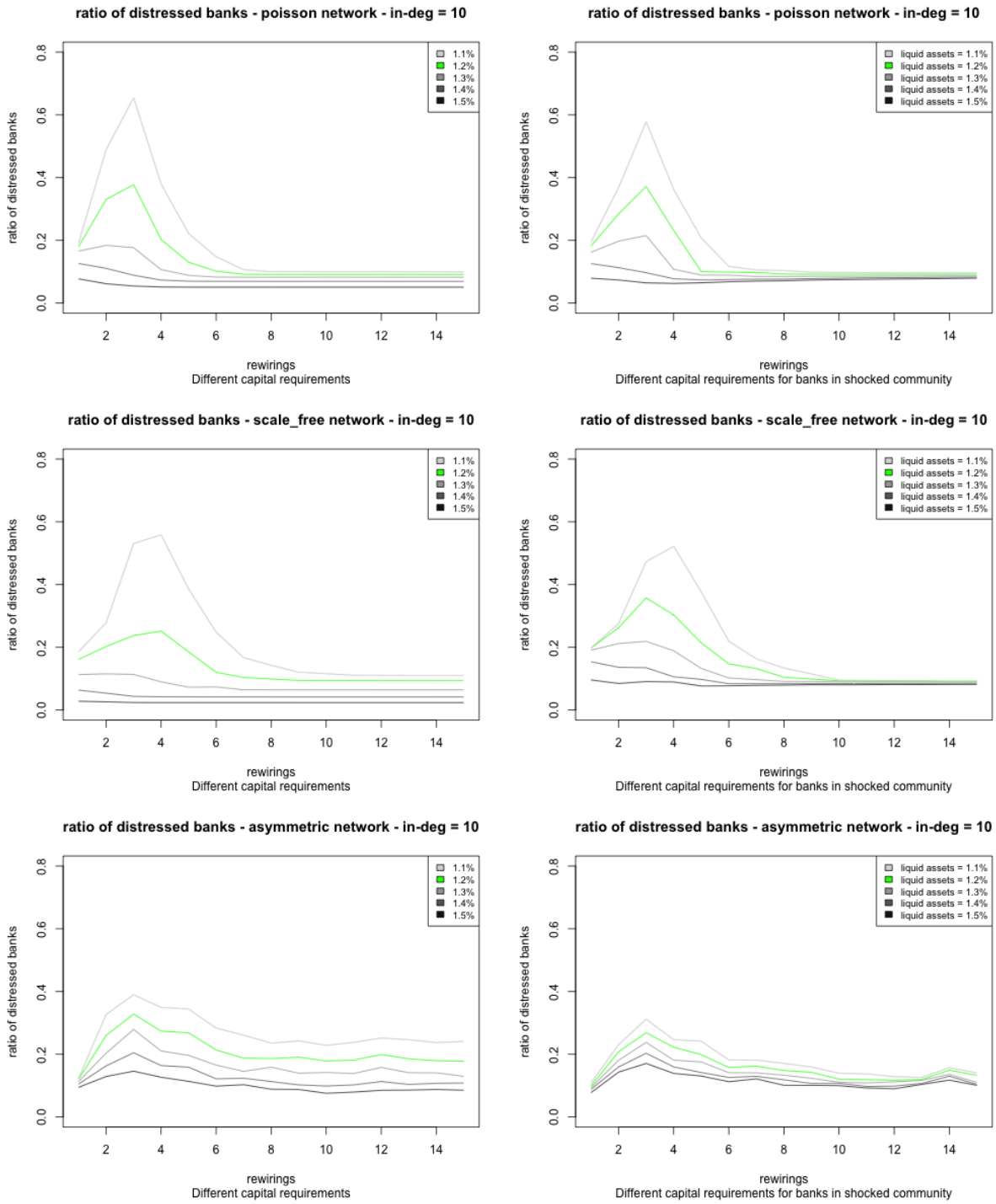
In the third experiment the liquid assets are partially determined using centrality measures. In

FIGURE 5.5: Average percentage of the banks in the network who suffer liquidity shortages. x axes represent the level of rewiring. First column: *Poisson*, *scale free* and *asymmetric* network with average in- and out-degree = 10. Second column: networks with average in- and out-degree = 6.



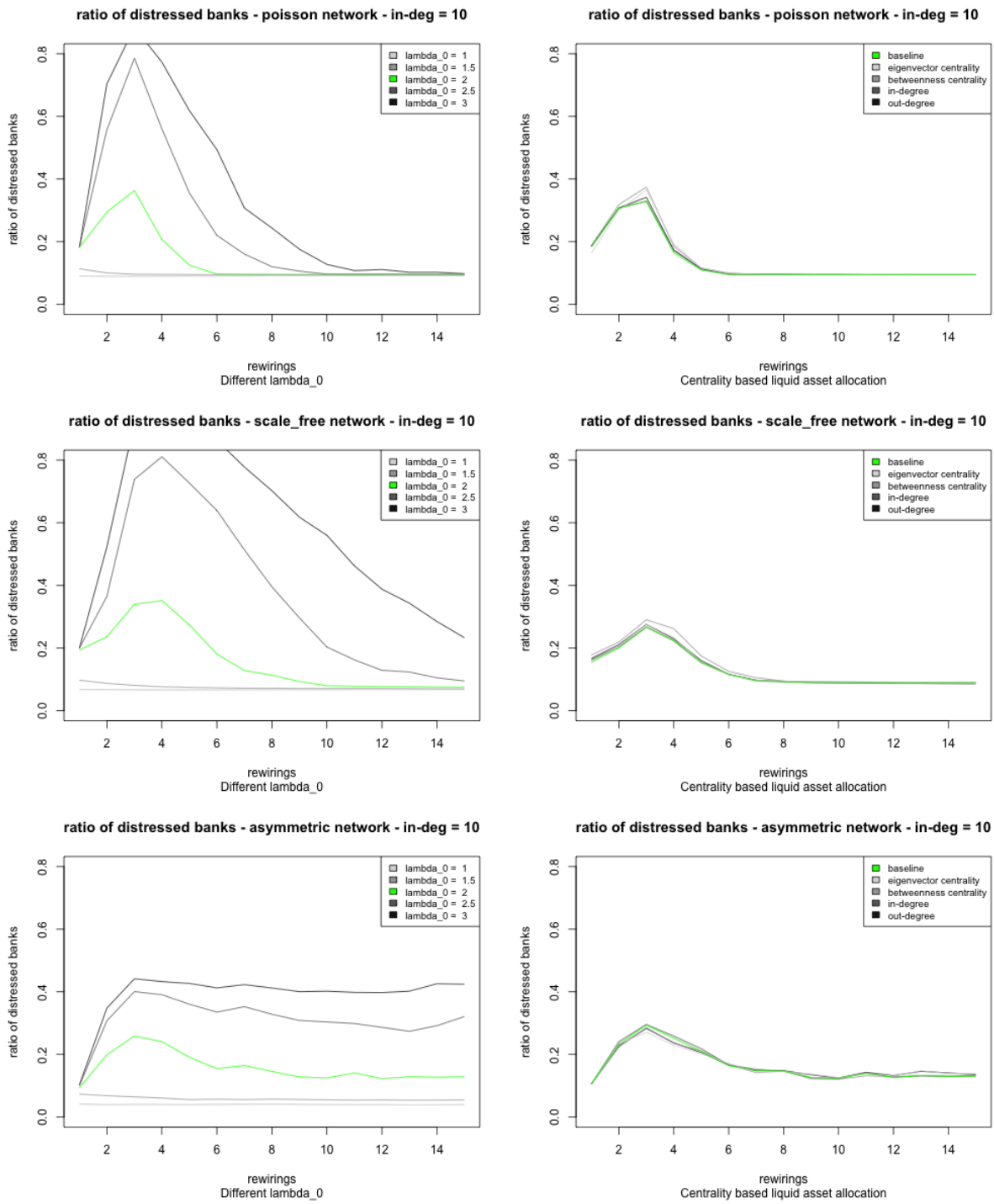
particular, we maintain the total amount of liquid asset in the system equal to the baseline scenario, and we assign it to each bank as follow: each bank will have 1% of the balance sheet of

FIGURE 5.6: Percentage of distressed banks using alternative capital requirements (methods 1 and 2). The first column represents the case in which the liquid assets are changed for all the banks in the network. The second column the case in which is changed only for banks in the communities of the shocked bank.



liquid assets, while the rest (0.2% on average) is assigned proportionally to the centrality measures. The first column of Figure 5.7 reports the results of this experiments in terms of percentage

FIGURE 5.7: Percentage of distressed banks using alternative capital requirements (methods 3 and 4). The first column is the case of centrality-based liquidity requirements. The second column is the case of different hoarding parameters λ^0.



of distressed banks in the system. The green line is the baseline scenario (equal proportion of liquid assets for all the banks), and the others are computed considering four centrality measures: in- and out-degree, eigenvector centrality and betweenness centrality. We see that the allocation

does not influence significantly the percentage of defaults in any of the network configurations. This results may appear puzzling, given the debate on *too-central-to-fail* institutions, that focuses the attention on the identification of systematically important banks and their role in the financial system, as for instance the G-SIB (Globally Systemic Banks), an official list of financial institutions identified by the FSB (Financial Stability Board), that are subject to stricter regulation (Board of Financial Stability, 2013). Our results may be related specifically to our methodology, in which the shock that hits a bank is not proportional to the size of the banks. In this setting, a shock applied to a relevant node would be more easily dispersed and would be less likely to cause chain reactions. Our results therefore do not dismiss the need for regulators to pay a particular attention to the *systemically important banks*. Instead, these results underline the fact that their safety is not a sufficient condition for the stability of the system, that may be fragile also to shocks to more peripheral banks, especially if the network structure is characterized by communities where the effect of a liquidity crisis can be amplified. Moreover, this result underlines how the usage of off-the-shelf centrality measures may have implicit drawbacks. Future studies may further address this issue by comparing the performance of more sophisticated centrality measures such as *Debt Rank* (Battiston et al., 2012a), centrality based on *harmonic distance* (Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015) and *strength between* and *within* country (Torri, Giacometti, and Paterlini, 2018).

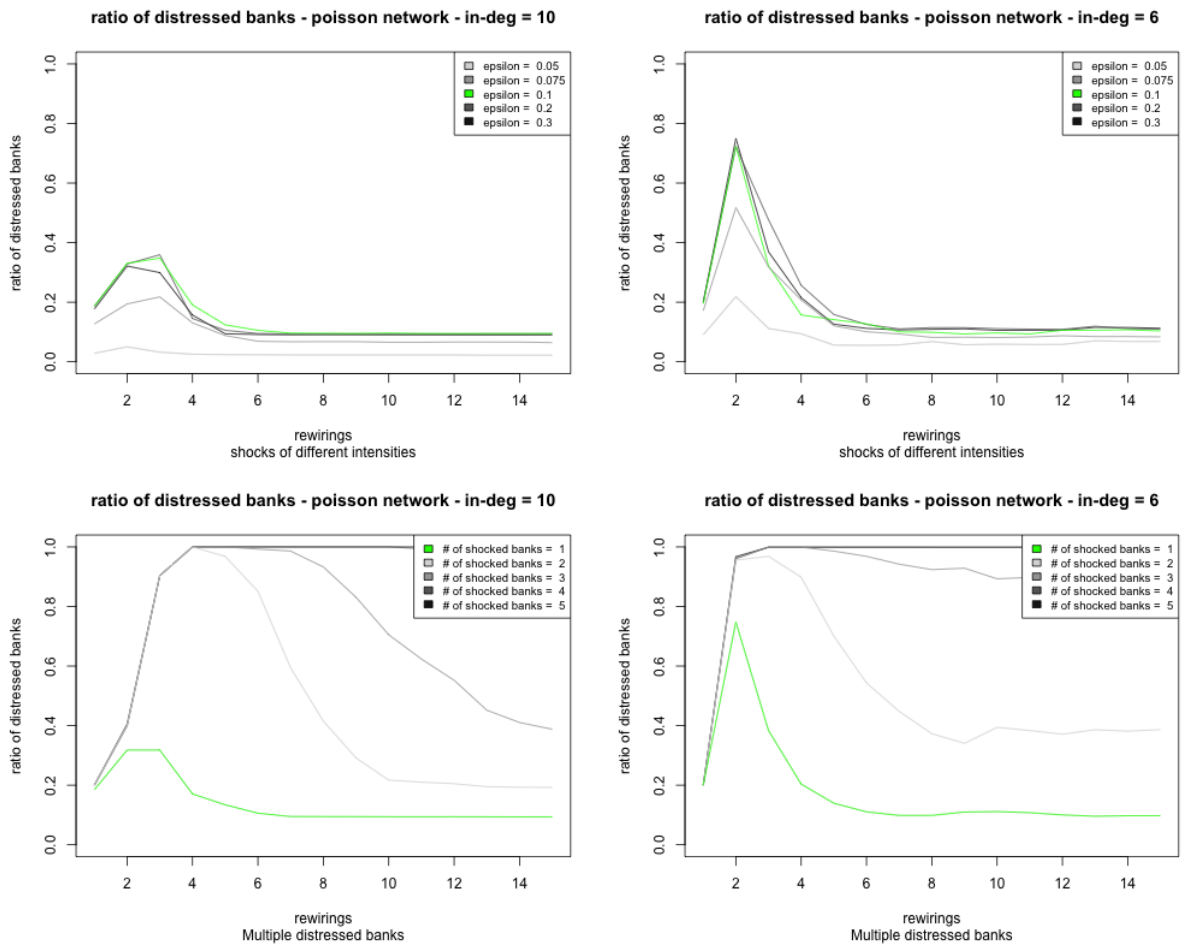
Finally, we model a fourth setting in which the coefficient λ^0 varies from 1 to 3 (the baseline value is 2). This means that a distressed bank withdraws from the system a quantity of liquidity different from the baseline scenario, where the banks hoarded twice the amount of money required to fulfill their liquidity requirements. The increase of λ^0 corresponds to a more conservative approach of the banks in the system, that fear further liquidity constraints, and can be associated to policies that increases the transparency in the system or guarantee credit lines to distressed banks in the event of a generalized liquidity crisis.[7] The second column of figure 5.7 shows the percentage of distressed banks for different levels of λ^0. We see that the extent of liquidity crises increases monotonically in λ^0 for every simulation setting. For $\lambda^0 = 1$ the contagion does not spread at all, when λ^0 increases the contagion is more intense, especially in network with a strong community structure, and finally, for high value of λ^0, the contagion grows significantly and is stronger also in networks without community structure.

[7]Note however that the introduction of guarantee systems can have a positive ex post effect, but risk to mine market discipline, especially in presence of *too-big-to-fail* institutions.

5.6 Robustness analysis

In the previous sections we have seen that the results presented in the chapter are robust to three different network specifications and to two levels of in-degrees of the nodes. We consider here two more robustness checks. First, we consider different levels of idiosyncratic shock (ε equal to 0.05, 0.075 0.1, 0.2, and 0.3, while the baseline level was 0.2). We then consider scenarios where more than one node is in distress, by applying a shock with ε=0.2 to 2,3,4 and 5 banks in the same community. For brevity we report only the results for the *Poisson networks*, that are qualitatively similar to the ones obtained with other network models.

FIGURE 5.8: Percentage of distressed banks using alternative specifications. The top panels refer to scenarios with alternative level of idiosyncratic shock for the distressed bank, while bottom panels refer to scenarios where more than one bank is shocked. Panels on the left refer to networks with in-degree equal to 10, and in panels on the right the in-degree is equal to 6.



We see in the upper panels of Figure 5.8 the effects of changes in the size of the idiosyncratic

shock ε for *Poisson networks* with in-degree of the nodes equal to 10 and 6, respectively. The level of distress introduced in the system, expressed in terms of ratio of distressed banks, grows proportionally to the size of the shock for small εs, but when ε reaches a certain level, it stabilizes around values similar to the baseline scenario. The explanation is that, once the shock is strong enough, the hit bank hoards all its asset in the interbank market, and higher level of shocks do not further affect the system.

The the effect of multiple shocks has drastic consequences on the system. In the bottom panels of Figure 5.8 we see that the extent of the distressed introduced in the system is much higher when more than one bank receives an idiosyncratic shock. Moreover, for the cases with 2-3 shocked banks, the graph is qualitatively similar to the baseline case: the distress first increases with the level of rewiring, but then it tend to decreases when the community becomes weak enough, and the shocks are dispersed in the network. On the other hand, when the number of shocked banks is higher, the level of distress remains high for high level of rewiring of the networks.

Overall, this robustness analysis confirms the results obtained before.

5.7 Conclusion

This work expands the literature on financial contagion in interbank credit markets, modeling liquidity crises is simulation settings similar to Gai, Haldane, and Kapadia, 2011. To our knowledge, our study is the first to analyze the role of a community structure in the diffusion of liquidity contagion in an interbank network. We claim that this feature is particularly relevant for international banking systems where the geographical collocation of banks greatly influences its interactions with other banks, such as the European banking system.

In the analysis, we consider several network specifications, characterized by different degree distributions and density, simulating an idiosyncratic shock to a random bank in the network. We compare the diffusion process in a simulation study where we construct networks with identical degree distribution, but with a different level of community structure, obtained thanks to a rewiring procedure.

The empirical application shows that the presence of communities has a strong influence on the diffusion of contagion, increasing both the number of banks involved and the size of the total liquidity shortage. The results are consistent across different network specifications. We also

introduce several variations to the original model specification that reflect different policy interventions: an increased level of liquid assets (first for all the banks in the system, then only for the community where the crisis bursts) that reflect a tighter regulation on the liquidity requirement; the assignment of liquidity requirements proportional to network centrality measures, that reflects policies that target the so called *systemically important banks*; finally we test the change of the parameter λ^0, that controls the amount of funds that a bank hoards from the interbank market as a consequence of a liquidity shortage, whose changes may be determined by the level of confidence in the market influenced by policies that promote transparency. The results show that the increase of the liquidity requirements and the increase in the confidence in the market (lower λ^0) are effective measures in the reduction of contagion risk in the market, especially in networks with tight communities. The introduction of centrality measures for the determination of liquidity requirements on the other hand seem to have a limited effect on the contagion risk, at least in out empirical setting. We can point out three main policy recommendation from this work: the first is that the prevention of liquidity crises should be focused on the promotion of market transparency and the definition of sound liquidity requirements for all the banks, the second is that the community structure of a network can have a great influence on the issue, fostering the growth of uncontrollable chain reactions, and finally, that the monitoring of *systemically important banks* is not sufficient for the prevention of liquidity crises. Overall, the results presented here contribute to the general network literature that studies the diffusion of dynamic process over networks with communities, showing that, due to the peculiarities of the contagion model presented here, liquidity contagion may be amplified by the presence of communities. This work can be expanded in several directions: the first is to define specific risk measures for the systems that are based on the concept of community structure, the second is to assess the structural properties of the true international interbank network in Europe in order to test the empirical application of the results, and finally to introduce in the model more realistic features such as informational asymmetries, the possibility of gather liquidity from other sources or the public intervention in case of widespread crisis.

Appendix A

Appendix for Chapter 1

In this appendix we describe the algorithms used for the estimation of *glasso* and *tlasso*.

A.1 The *glasso* algorithm

Here we briefly describe the algorithm proposed by Friedman, Hastie, and Tibshirani, 2008 to solve (1.9), the *glasso* model. For convenience, we define X_i as the ith element of X, and and $X_\backslash i$ as the vector of all the elements of X except the ith. We also define the matrices $\boldsymbol{G}$ to be the estimate of $\boldsymbol{\Sigma}$, and $\boldsymbol{S}$ the sample covariance matrix. Furthermore, we identify the following partitions:[1]

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_{\backslash i,\backslash i} & g_{\backslash i,i} \\ g'_{\backslash i,i} & g_{i,i} \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{\backslash i,\backslash i} & s_{\backslash i,i} \\ s'_{\backslash i,i} & s_{i,i} \end{pmatrix}. \tag{A.1}$$

Banerjee, Ghaoui, and d'Aspremont, 2008 show that the solution for $w_{\backslash i,i}$ can be computed by solving the following box-constrained quadratic program:

$$g_{\backslash i,i} = \arg\min_{y} \left\{ y' \boldsymbol{G}^{-1}_{\backslash i,\backslash i} y : ||y - s_{\backslash i,i}||_\infty \leq \rho \right\}, \tag{A.2}$$

or in an equivalent way, by solving the dual problem

$$\min_{\beta^{(i)}} \left\{ \frac{1}{2} ||\boldsymbol{G}^{1/2}_{\backslash i,\backslash i} \beta^{(i)} - c||^2 + \rho ||\beta^{(i)}||_1 \right\}, \tag{A.3}$$

where $c = \boldsymbol{G}^{-1/2}_{\backslash i,\backslash i} s_{\backslash i,i}$ and $\hat{\beta}^{(i)} = \boldsymbol{G}^{-1}_{\backslash i,\backslash i} g_{\backslash i,i}$. (A.3) resembles a lasso least square problem (see Tibshirani, 1996). The algorithm estimates then the ith variable on the others using as input

[1] The dimension of $\boldsymbol{G}_{\backslash i,\backslash i}$, $g_{\backslash i,i}$ and $g_{i,i}$ are respectively $((n-1) \times (n-1))$, $((n-1) \times 1)$ and (1×1).

$\boldsymbol{G}_{\backslash i,\backslash i}$, where $\boldsymbol{G}_{\backslash i,\backslash i}$ is the current estimate of the upper left block. The algorithm then updates the corresponding row and column of $\boldsymbol{G}$ using $g_{\backslash i,i} = \boldsymbol{G}_{\backslash i,\backslash i}\hat{\beta}^{(i)}$ and cycles across the variables until convergence.
As noted by Friedman, Hastie, and Tibshirani, 2008, the resulting matrix remains positive definite and invertible if the procedure is initialized with a positive definite matrix.

Glasso algorithm

1. Start with $\boldsymbol{G} = \boldsymbol{S} + \rho\boldsymbol{I}$. The diagonal of $\boldsymbol{G}$ is unchanged in the next steps.

2. For each $i = 1, 2, \ldots, n, 1, 2, \ldots, n, \ldots$, solve the lasso problem (A.3), which takes as input $\boldsymbol{G}_{\backslash i,\backslash i}$ and $s_{\backslash i,i}$. This gives a $n-1$ vector solution $\hat{\beta}$. Fill in the corresponding row and column of $\boldsymbol{G}$ using $g_{\backslash i,i} = \boldsymbol{G}_{\backslash i,\backslash i}\hat{\beta}$.

3. Repeat until a convergence criterion is satisfied.

A.2 The *tlasso* Expectation-Maximization algorithm

Let $\boldsymbol{W} \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi}^{-1})$ and $\tau \sim \Gamma(v/2, v/2)$ be random variables distributed as a multivariate Gaussian and a gamma distribution, respectively. Then:

$$\mathbf{X} = \boldsymbol{\mu} + \frac{\boldsymbol{W}}{\sqrt{\tau}} \sim t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v), \tag{A.4}$$

where $\boldsymbol{\Psi}^{-1}$ is the dispersion matrix of the t-Student distribution, $\boldsymbol{\mu}$ is a $m \times 1$ vector, and v is the number of degrees of freedom in the multivariate t-Student distribution.
The EM algorithm treats τ as a hidden variable in the E-step, exploiting the fact that the conditional distribution of $\mathbf{X}$ given τ is $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Psi}^{-1}/\tau)$. In the M-step, the algorithm then maximizes the penalized log-likelihood of the latent Gaussian vector using the *glasso* procedure.
Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an n-sample drawn from $t_m(\boldsymbol{\mu}, \boldsymbol{\Psi}^{-1}, v)$. The EM algorithm relies on the following two steps:

- **E-step**

  - Given:

$$\mathbb{E}[\tau|\mathbf{X} = \mathbf{x}] = \frac{v + m}{v + (\delta_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Psi}))}, \tag{A.5}$$

where $\delta_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Psi}(\mathbf{x} - \boldsymbol{\mu})$, from the current estimates $\widehat{\boldsymbol{\mu}}^{(t)}$ and $\widehat{\boldsymbol{\Psi}}^{(t)}$, we compute $\widehat{\tau}^{(t+1)}$ for the $(t+1)$th iteration:

$$\widehat{\tau}_i^{(t+1)} = \frac{v + m}{v + (\delta_{\mathbf{X}_i}(\widehat{\boldsymbol{\mu}}^{(t)}, \widehat{\boldsymbol{\Psi}}^{(t)}))} \qquad i = 1, \ldots, n. \tag{A.6}$$

- **M-step**
  - Compute the estimates at iteration $t+1$:

$$\widehat{\boldsymbol{\mu}}^{(t+1)} = \frac{\sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)} \mathbf{X}_i}{\sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)}}, \tag{A.7}$$

$$\widehat{\boldsymbol{S}}^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}_i^{(t+1)} [\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}][\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(t+1)}]'. \tag{A.8}$$

  - Then, the parameter $\widehat{\boldsymbol{\Psi}}^{(t+1)}$ is obtained by solving the following optimization problem:

$$\widehat{\boldsymbol{\Psi}}^{(t+1)} = \arg\max_{\boldsymbol{\Psi}} \left(log|\boldsymbol{\Psi}| - tr(\boldsymbol{\Psi} \widehat{\boldsymbol{S}}^{(t+1)}) - \lambda ||\boldsymbol{\Psi}||_1 \right), \tag{A.9}$$

    which relates to Equation 1.9 for *glasso* and can be solved using the algorithm from Friedman, Hastie, and Tibshirani, 2008.

The E and M steps are sequentially iterated until a convergence criterion is met, that is, until the maximum term in absolute value of the matrix difference between $\widehat{\boldsymbol{\Psi}}$s in two consecutive iterations is smaller than a given threshold. While convergence to a stationary point is guaranteed in the penalized versions of EM (McLachlan and Krishnan, 2007), the algorithm is not guaranteed to find the global maximum since the penalized log-likelihood function to be maximized is not concave (Finegold and Drton, 2011).

We finally underline that the estimates of the covariance matrix obtained using *tlasso* are always positive-definite and invertible, as in each M-step the *glasso* is applied on a positive-definite matrix.

Appendix B

Appendix for Chapter 2

B.1 Alternative covariance estimation methods

Here, we briefly describe the benchmark covariance estimators we use in the comparative analysis in Chapter 2. Differently from *glasso* and *tlasso*, these approaches provide an estimate for the covariance matrix and not for the precision matrix. Hence, we compute the precision matrix for such methods to be plug-in into the minimum variance portfolio by inverting the covariance. In particular, we consider the *sample covariance* and the equally weighted methods (that are commonly regarded as naive approaches) and two state-of-art estimators: random matrix theory and Ledoit Wolf Shrinkage.

The equally weighted (EW) portfolio, a tough benchmark to beat (DeMiguel, Garlappi, and Uppal, 2009), can be interpreted as an extreme shrinkage estimator of the global minimum variance portfolio, obtained using the identity matrix as the estimate of the covariance matrix. Indeed, using (2.3), we obtain $\widehat{\mathbf{w}}_{EW} = \dfrac{\boldsymbol{I}\mathbf{1}}{\mathbf{1}'\boldsymbol{I}\mathbf{1}} = \frac{1}{n}\mathbf{1}$. By assuming zero correlations and equal variances, such approach is very conservative in terms of estimation error and it suitable in case of severe unpredictability of the parameters.

The second naive approach is the sample covariance estimator, defined as:

$$\boldsymbol{S} = \frac{1}{t-1}\sum_{\tau=1}^{t}(X_\tau - \bar{X})(X_\tau - \bar{X})', \tag{B.1}$$

where t is the length of the estimation period, X_i is the multivariate variate vector of assets' returns at time τ and $\bar{X}$ is the vector of the average return for the n assets. Such estimator, when

computed on datasets with a number of asset close to the length of the window size, is typically characterized by a larger eigenvalue dispersion compared to true covariance matrix, causing the matrix to be ill-conditioned (Meucci, 2009). Therefore, when computing the precision matrix by inverting the covariance matrix, estimates are typically not reliable and unstable on different samples as its ill-conditioning nature *amplifies* the effects of the estimation error in the covariance matrix.

The shrinkage methodology of Ledoit-Wolf (LW) is well-known to better control for the presence of estimation errors, especially for datasets with a large ratio of n/t, where n is the number of assets and t the length of the estimation window. The Ledoit-Wolf shrinkage estimator is defined to be a convex combination of the sample covariance matrix $\boldsymbol{S}$ and $\widehat{\boldsymbol{\Sigma}}_T$, a highly structured target estimator, such that $\widehat{\boldsymbol{\Sigma}}_{LW} = a\boldsymbol{S} + (1-a)\widehat{\boldsymbol{\Sigma}}_T$ with $a \in [0, 1]$. Following Ledoit and Wolf, 2004b, we consider as structured estimator $\widehat{\boldsymbol{\Sigma}}_T$ the constant correlation matrix, such that all the pairwise correlations are identical and equal to the average of all the sample pairwise correlations. As the target estimator is characterized by good conditioning, the resulting shrinkage estimator $\widehat{\boldsymbol{\Sigma}}_{LW}$ has a smaller eigenvalues dispersion than the sample covariance matrix. In fact, the sample covariance matrix is *shrunk* towards the structured estimator, with intensity depending on the value of the *shrinkage constant* a. Ledoit-Wolf estimation of a is based on the minimization of the expected distance between $\widehat{\boldsymbol{\Sigma}}_{LW}$ and $\boldsymbol{\Sigma}$. For further details, the reader is referred to Ledoit and Wolf, 2004b.[1]

The last approach we focus on is the so called random matrix theory (RMT) estimator $\widehat{\boldsymbol{\Sigma}}_{RMT}$, introduced by Laloux et al., 1999. The approach is based on the fact that, in the case of financial time series, the smallest eigenvalues of the correlation matrices are often dominated by noise. From the known distribution of the eigenvalues of a random matrix, it is possible then to filter out the part of spectrum that is likely associated with estimation error and maintain only the eigenvalues that carry useful information (Laloux et al., 1999). In particular, when assuming $i.i.d.$ returns, the eigenvalues of the sample correlation matrix are then distributed according to a Marcenko-Pastur (MP) distribution as a consequence of the estimation error. Therefore, we

[1]Interestingly, the Ledoit-Wolf shrinkage is closely related to portfolio optimization with L_2 penalization of weight estimates. Indeed, the optimization problem $\min_{\boldsymbol{w}\in C}(\boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{w} + a\boldsymbol{w}'\boldsymbol{w})$, with $C = \{\boldsymbol{w}|\boldsymbol{1}'\boldsymbol{w} = 1\}$ can be equivalently stated as $\min_{\boldsymbol{w}\in C}(\boldsymbol{w}'(\widehat{\boldsymbol{\Sigma}} + a\boldsymbol{I})\boldsymbol{w})$, which then is equivalent to solving the problem using the Ledoit-Wolf shrinkage estimator with $\widehat{\boldsymbol{\Sigma}}_T = \boldsymbol{I}$ (Bruder et al., 2013).

can compute the eigenvalues that correspond to noise based on the minimum and maximum eigenvalues of the theoretical distribution, such that:

$$\lambda_{\min\max} = \sigma^2\big(1 \pm \sqrt{n/t}\big)^2, \tag{B.2}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the theoretical smallest and largest eigenvalues in a $n \times n$ random covariance matrix estimated by a sample of t observations and σ^2 is the variance of the i.i.d. asset returns. Only the eigenvalues outside the interval $[\lambda_{\min}, \lambda_{\max}]$ are then assumed to bring useful information, while the others correspond to noise. Here, we estimate the covariance matrix then by *eigenvalue clipping*, a technique that consists in substituting the eigenvalues smaller than $\lambda_{\max}$ with their average:

$$\widehat{\boldsymbol{\Sigma}}_{RMT} = \boldsymbol{V}\boldsymbol{\Lambda}_{RMT}\boldsymbol{V}', \tag{B.3}$$

where $\boldsymbol{V}$ represents the eigenvectors of the sample covariance matrix and $\boldsymbol{\Lambda}_{RMT}$ is the diagonal matrix with the ordered eigenvalues, where the eigenvalues $\lambda \leq \lambda_{\max}$ are substituted by their average (Bouchaud and Potters, 2009). The RMT filtering has then the effect of averaging the lowest eigenvalues, improving the conditioning of the matrix and therefore reducing the sensitivity of the precision matrix to estimation errors.

For further details the reader is referred to Laloux et al., 1999, Bouchaud and Potters, 2009 and Bruder et al., 2013.

Appendix C

Appendix for Chapter 3

C.1 Robustness Checks

Here, we report some robustness checks on the model by testing different specifications for the input data. We control for two potential sources of misspecification in the model (ARMA-GARCH effects in the time series and exposures of the banks to common risk factors), and we investigate the estimated network structure computed by considering only distressed market periods, which are characterized by increasing CDS spreads. We report a set of network indicators computed on the estimated networks in Table C.1. For brevity, we present only the data for the *during-crisis* period.

First, we control for the presence of serial correlation and heteroskedasticity. The *tlasso* model requires the data to be approximately i.i.d., and the presence of serial correlation in CDS log-differences or in their variance might be a problem for the estimation. We consider two different multivariate models: first a Constant Conditional Correlation (CCC) GARCH(1,1) model (Bollerslev, 1990), and a Dynamic Conditional Correlation (DCC) GARCH(1,1) (Engle, 2002). For both models, we estimate an ARMA(1,1) process for the mean. Once the model is fitted, we estimate the *tlasso* model on the residuals, in a two-step procedure similar to Anufriev and Panchenko, 2015 and Billio et al., 2012. We report some of the indicators of Section 3.4.3 computed on the residuals of the GARCH models in Table C.1 (columns b,c). The results are rather similar to the ones computed directly on the log-differences of CDS spreads (column a), consistently with the fact that volatility clustering and serial correlation in the data were not particularly relevant.

As a second robustness check, we apply a factor model to the data to control for the exposure to conditioning variables. The factors we consider are 5-year sovereign CDS spreads of the 11

European countries to which the banks in our sample belong (i.e., Austria, Belgium, Denmark, France, Germany, Italy, Netherland, Portugal, Spain, Sweden and United Kingdom). In order to avoid collinearity issues, we perform a principal component analysis, retaining the first 8 orthogonal factors, which together explain more than 80% of the variability in the dataset. From Table C.1, (column d), we see that the value of the indicators is rather similar to those computed from original data. Notably, the modularity of the country partition is still positive, indicating that the clustering of banks along national borders is present even after controlling for the exposure to sovereign debt.

Finally, we test if the network structure is different in presence of market distress. We focus on the periods characterized by upward movements of the CDS spreads, that is, increases in credit risk. We use the following method to identify these periods of distress. We first construct an equally weighted porfolio of bank CDS spreads and then select the points in time at which the log-returns of the porfolio are positive, corresponding to an average increase of the CDS spreads. In this way, we obtain for the *during crisis* period 102 observations, which we use to compute the sparse partial correlation network with *tlasso*. Also in this case, the values of the network indicators presented in Table C.1 (column e) are similar to those obtained for the original specification, indicating that the partial correlation networks are stable also in periods of increasing spreads. This result is consistent with the fact that the distribution of the CDS spreads log-difference is symmetric, as implied by the multivariate t-Student assumption and suggested by the analysis of the data in Section 3.4.1.

TABLE C.1: Robustness check - Network indicators

	During-crisis period				
	Raw data (a)	CCC-GARCH (b)	DCC-GARCH (c)	Country factors (d)	Distressed market (e)
Density	49.25%	47.53%	47.74%	44.95%	43.44%
Avg strength	0.95	0.95	0.95	0.93	0.96
Largest connected component	100%	100%	100%	100%	100%
Positive edges %	87.77%	90.95%	90.99%	90.43%	84.65
Positive edges % (weighted)	96.59%	97.43%	97.43%	97.63%	94.91
Freeman Centralization	0.47	0.43	0.43	0.48	0.47
Coreness	0.11	0.12	0.12	0.13	0.11
Modularity (wrt country)	0.36	0.34	0.34	0.36	0.37
Clustering coefficient	2.70%	2.54%	2.54%	2.60%	2.82%
Assortativity (wrt centrality)	0.61	0.59	0.59	0.66	0.43

Network indicators computed on *tlasso networks* in the *during-crisis* period (01/01/2009 – 31/12/2012) with alternative model specifications. *Raw data (a)* refers to the network computed with the specifications presented in Section 3.4. *CCC-GARCH (b) and DCC-GARCH(c)* refer to the network computed on the residuals of Constant and Dynamic Conditional correlation GARCH model fitted on the log-differences of CDS spreads, respectively. *Country factors (d)* refers to the network computed on the residuals of a factor regression where the factors are Sovereign European CDS spreads for 11 European countries. *Distress market (e)* refers to the network computed in periods of increasing spreads.

C.2 List of banks and summary statistics

Bank Name	Short Name	Country	Area	*mean*	*sd*	*sk*	*ku*	*v (d.f.)*	*ks t (1%)*	*ks norm (1%)*	ECB	EBA
B.ca M. Paschi Di Siena S.p.A.	M.te Paschi	Italy	S	0.00	0.09	0.26	3.87	6.47	not rej.	not rej.	Yes	Yes
Banco Popolare - So.Co.	Banco Popolare	Italy	S	0.00	0.09	0.63	4.48	5.39	not rej.	not rej.	Yes	Yes
B.ca Pop. Di Milano - S.C.R.L.	BPM	Italy	S	0.00	0.08	0.68	6.4	2.65	not rej.	rej.	Yes	No
BNL - Gruppo BNP Paribas	BNL	Italy	S	0.00	0.09	0.08	3.79	7.60	not rej.	not rej.	Yes**	Yes**
Intesa Sanpaolo S.p.A.	Intesa S.P.	Italy	S	0.00	0.11	0.30	3.65	8.59	not rej.	not rej.	Yes	Yes
Mediobanca S.p.A.	Mediobanca	Italy	S	0.00	0.08	0.03	6.33	2.50	not rej.	rej.	Yes	No
UniCredit S.p.A.	Unicredit	Italy	S	0.00	0.10	0.45	4.45	6.76	not rej.	not rej.	Yes	Yes
Banco Santander S.A.	B.co Santander	Spain	S	0.00	0.10	0.09	4.14	7.60	not rej.	not rej.	Yes	Yes
B.co Bilbao Vizcaya Arg. S.A.	BBVA	Spain	S	0.00	0.09	0.09	3.90	8.20	not rej.	not rej.	Yes	Yes
Banco Comercial Português	Banco Com. Port.	Portugal	S	0.00	0.09	0.79	6.39	2.44	not rej.	rej.	Yes	No
BNP Paribas	BNP Paribas	France	C	0.00	0.10	0.04	3.95	6.38	not rej.	not rej.	Yes	Yes
Crédit Agricole S.A.	Cr. Agricole	France	C	0.00	0.09	0.35	3.90	8.71	not rej.	not rej.	Yes	Yes
Crédit Lyonnais	Credit Lyonnais	France	C	0.00	0.09	0.29	3.83	8.80	not rej.	not rej.	Yes***	Yes***
Société Générale S.A.	Soc. Gen.	France	C	0.00	0.09	0.37	4.04	6.89	not rej.	not rej.	Yes	Yes
KBC Bank	KBC Bank	Belgium	C	0.00	0.06	0.91	7.94	2.38	not rej.	rej.	Yes	Yes
Bayerische Landesbank	Bayerische LBank	Germany	C	0.00	0.06	0.28	6.89	1.75	rej.	rej.	Yes	Yes
commerzbank AG	Commerzbank	Germany	C	0.00	0.09	0.60	6.63	5.52	not rej.	not rej.	Yes	Yes
Deutsche Bank AG	DB	Germany	C	0.00	0.09	0.08	5.24	5.78	not rej.	not rej.	Yes	Yes
Erste Group Bank AG	Erste Group	Austria	C	0.00	0.07	0.28	8.37	2.41	not rej.	rej.	Yes	Yes
Co operatieve Rabobank U.A.	Rabobank	Netherl.	C	0.00	0.07	-0.11	4.39	4.25	not rej.	not rej.	Yes	Yes
ING Bank	ING Bank	Netherl.	C	0.00	0.08	0.05	3.89	6.08	not rej.	not rej.	Yes	Yes
Skand. Enskilda Banken	Skandinaviska	Sweden	O	0.00	0.05	0.36	6.47	2.15	not rej.	rej.	No*	Yes
Svenska Handelsbanken	Svenska Handb.	Sweden	O	0.00	0.06	0.81	6.48	2.88	not rej.	rej.	No*	Yes
Danske Bank	Danske Bank	Denmark	O	0.00	0.06	0.20	5.22	3.17	not rej.	not rej.	No*	Yes
Barclays Bank PLC	Barclays Bank	UK	O	0.00	0.10	0.30	4.82	7.21	not rej.	not rej.	No*	Yes
Bank of Scotland PLC	Bank of Scotland	UK	O	0.00	0.08	0.53	6.48	4.11	not rej.	not rej.	No*	No
HSBC Bank PLC	HSBC Bank	UK	O	0.00	0.09	0.11	4.35	6.59	not rej.	not rej.	No*	Yes
Lloyds Bank PLC	Lloyds Bank	UK	O	0.00	0.09	0.64	5.67	6.75	not rej.	not rej.	No*	Yes
Natwest Bank PLC	Natwest Bank	UK	O	0.00	0.08	0.47	5.49	4.86	not rej.	not rej.	No*	No
Royal Bank of Scotland PLC	RBS	UK	O	0.00	0.09	0.80	6.35	5.79	not rej.	not rej.	No*	Yes
Standard Chartered Bank	Std. Chartered	UK	O	0.00	0.09	0.14	4.08	5.72	not rej.	not rej.	No*	No

TABLE C.2: *Area* indicates the geographical location of the banks (*S* refers to Southern Europe, *C* to central Europe and *O* to countries outside Eurozone). Data reports mean, standard deviation (*sd*), skewness (*sk*) and kurtosis (*ku*) for the period 01/01/2009 – 30/06/2016. Degrees of freedom (v) refer to the univariate t distribution fitted on the data. *ks t (1%)* and *ks norm (1%)* refer to the Kolmogorov Smirnov test for goodness of fit of t and normal distribution with $\alpha = 0.01$ (the value refers to the rejection of the distribution considered). The last two columns refer to being under European Central Bank supervision, and of being in the group of financial institutions considered by the EBA EU-wide stress test exercise, respectively. *Not in Eurozone. **BNP Paribas Group. ***Credit Agricole Group.

Appendix D

Appendix for Chapter 4

D.1 Quantile functions for conditional distributions

We discuss here how to compute conditional quantile functions under multivariate t-Student distributions and we present the linear approximation.

Let $\mathbf{X} \sim t_p(\mu, \mathbf{\Sigma}, \nu)$ where $\mu, \mathbf{\Sigma}, \nu$ are the location, scale and number of degrees of freedom of the distribution, respectively. We consider the following partition:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim t_p \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{11} \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{22} \end{bmatrix}, \nu \right). \tag{D.1}$$

The conditional distribution can be obtained considering the following scale-mixture representation of a multivariate t-Student distribution:

$$\mathbf{X} = \mu + \mathbf{\Sigma}^{1/2} \boldsymbol{Z} \sqrt{q}, \tag{D.2}$$

where $\boldsymbol{Z}$ is a standard p-variate Gaussian, $q \sim \chi^2_\nu / \nu$ and $\boldsymbol{Z}$ is independent from q. Following Ding, 2016, it is possible to show that the conditional distribution of $\mathbf{X}_2 | \mathbf{X}_1$ is itself a t-Student distribution, although with a different number of degrees of freedom. For the case with $\mu = \mathbf{0}$ we have:

$$\mathbf{X}_2 | \mathbf{X}_1 \sim t_{p_2} \left(\mu_{2|1}, \frac{\nu + d(\mathbf{x}_1)}{\nu + p_1} \mathbf{\Sigma}_{22|1}, \nu + p_1 \right), \tag{D.3}$$

$$\mu_{2|1} = \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{X}_1, \tag{D.4}$$

$$\mathbf{\Sigma}_{22|1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}, \tag{D.5}$$

where p_1 and p_2 are the dimension of $\mathbf{X}_1$ and $\mathbf{X}_2$, and $d(\mathbf{X}_1) = \mathbf{X}_1'\mathbf{\Sigma}_{11}^{-1}\mathbf{X}_1$ is the squared Mahalanobis distance from $\mathbf{X}_1$ to $\mu_1 = 0$ with scale matrix $\mathbf{\Sigma}_{11}$. We can then compute the conditional quantiles as functions of $\mathbf{x}_1$ for the case in which $\mathbf{X}_2$ is univariate, obtaining

$$Q(X_2|X_1, \tau) = Q_{t,\nu+p_1}(\tau)\left(\frac{\nu + d(\mathbf{x}_1)}{\nu + p_1}\mathbf{\Sigma}_{22|1}\right)^{1/2} + \mu_{2|1}, \tag{D.6}$$

where $Q_{t,\nu+p_1}(\tau)$ is the quantile function of a t-Student distribution with $\nu + p_1$ degrees of freedom.

Remark. *Differently from the Gaussian case, the conditional quantile of $X_2|X_1$ is not an affine function in $\mathbf{X}_1$, as $d(\mathbf{X}_1)$ is a non-linear function of $\mathbf{X}_1$.*

In Chapter 4 and in the rest of this Appendix, we refer to the case with $\mu = \mathbf{0}$ and where $X_i = X_2$ is univariate.

D.1.1 Proof of Proposition 1

Proof. We obtain the best linear approximation by minimizing the squared distance between the quantile function and an affine function $g(x) = \alpha + \beta x$. To simplify the notation we consider $\mathbf{X}_1 = \mathbf{X}_{\setminus i}$ and $\mathbf{X}_2 = \mathbf{X}_i$. We also define for convenience the function

$$\check{g}(x) = \alpha + \check{\beta}x = g(x) - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x. \tag{D.7}$$

After defining the quantity $\sigma_{22|1} := \left(\frac{\nu + d(\mathbf{X}_1)}{\nu + p_1}\mathbf{\Sigma}_{22|1}\right)^{1/2}$, we express the distance between the true quantile and the function $g(x)$ as a function of $\mathbf{X}_1$ as

$$g(\mathbf{X}_1) - Q_\tau(X_2|X_1) = \check{g}(\mathbf{X}_1) + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{X}_1 - Q_t(\tau)\sigma_{22|1} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{X}_1 \tag{D.8}$$

$$= \alpha + \mathbf{x}_1\check{\beta} + Q_t(\tau)\sigma_{22|1}. \tag{D.9}$$

The optimal parameters $\alpha, \check{\beta}$ are then obtained by minimizing the expected squared distance over $\mathbf{X}_1$:

$$\min_{\alpha,\check{\beta}} \mathbb{E}_{X_1}\left[\left(\mathbf{X}_1\check{\beta}+\alpha-Q_t(\tau)\sigma_{22|1}\right)'\left(\mathbf{X}_1\check{\beta}+\alpha-Q_t(\tau)\sigma_{22|1}\right)\right]= \tag{D.10}$$

$$\min_{\alpha,\check{\beta}} \mathbb{E}_{X_1}\left[\check{\beta}'\mathbf{X}_1'\mathbf{X}_1\check{\beta}+\alpha^2+Q_t^2(\tau)\sigma_{22|1}^2+2\check{\beta}'\mathbf{X}_1'\alpha-2\check{\beta}'\mathbf{X}_1'Q_t(\tau)\sigma_{22|1}-2\alpha Q_t(\tau)\sigma_{22|1}\right]= \tag{D.11}$$

$$\min_{\alpha,\check{\beta}} \check{\beta}\mathrm{CoV}(\mathbf{X})\check{\beta}+\alpha^2+Q_t^2(\tau)\mathbb{E}[\sigma_{22|1}^2]+2\alpha\check{\beta}'\mathbb{E}[\mathbf{X}_1]-2Q_t(\tau)\check{\beta}'\mathbb{E}[\mathbf{X}_1'\sigma_{22|1}]-2\alpha Q_t(\tau)\mathbb{E}[\sigma_{22|1}]= \tag{D.12}$$

$$\min_{\alpha,\check{\beta}} \check{\beta}\mathrm{CoV}(\mathbf{X}_1)\check{\beta}+\alpha^2+Q_t^2(\tau)\mathbb{E}[\sigma_{22|1}^2]-2\alpha Q_t(\tau)\mathbb{E}[\sigma_{22|1}]. \tag{D.13}$$

The last equality is satisfied thanks to $\mathbb{E}[\mathbf{X}_1]=0$ by construction and $\mathbb{E}[\mathbf{X}_1'\sigma_{22|1}]=\int_\Omega \mathbf{x}_1'\sigma_{22|1}df_{\mathbf{x}_1}=0$, as $\mathbf{x}_1'\sigma_{22|1}f_{\mathbf{x}_1}$ is an odd function (note that $f_{\mathbf{x}_1}(\mathbf{X}_1)$ is an even function since it is the density function of an elliptical distribution).

The function is convex and quadratic, and it has a global minimum. We compute the first order condition by setting the partial derivative of the function h in D.13 w.r.t. α and $\check{\beta}$ to zero:

$$\frac{\partial h}{\partial \alpha}=2\alpha-2Q_t(\tau)\mathbb{E}[\sigma_{22|1}]=0 \tag{D.14}$$

$$\frac{\partial h}{\partial \beta}=2\mathrm{CoV}(\mathbf{X}_1)\check{\beta}=0 \tag{D.15}$$

and we find the optimal solution $\alpha=Q_t\mathbb{E}[\sigma_{22|1}]$ and $\check{\beta}=0$, from which $\beta=\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x$ using (D.7). □

D.1.2 Proof of Proposition 2

Proof. The difference $D(\mathbf{X}_1)$ between the τ-quantile function (D.6) and its linear approximation (4.31) is:

$$D(\mathbf{x}_1) = Q_{t,\nu+p-1}(\tau)\left(\frac{\nu + d(\mathbf{X}_1)}{\nu + p - 1}\mathbf{\Sigma}_{22|1}\right)^{1/2} - Q_{t,\nu+p-1}(\tau)\mathbb{E}_{X_1}\left[\left(\frac{\nu + d(\mathbf{X}_1)}{\nu + p - 1}\mathbf{\Sigma}_{22|1}\right)^{1/2}\right]. \tag{D.16}$$

Considering that $d(\mathbf{X}_1)$ is an even, positive, and convex function, and has minimum in $\mathbf{X}_1 = 0$, $\mathbf{\Sigma}_{22|1}$ is positive by construction, $(\cdot)^{1/2}$ is an increasing function, and the second term of (D.16) is a constant, it follows that the function is convex for $Q > 0$ (i.e. $\tau > 0.5$) and concave for $Q < 0$ (i.e. $\tau < 0.5$). Since the max is for $\mathbf{X}_1 = 0$, the difference function $D(\mathbf{X}_1)$ is negative for $\mathbf{X}_1$ with all negative values, and the thesis follows. □

Remark. *Assuming that the focus is on low quantiles, the ΔCoVaR computed using a linear function for the conditional quantile (as in the classical implementation of Adrian and Brunnermeier, 2016), it will underestimate the true ΔCoVaR if the data are jointly t-Student distributed.*

Remark. *The underestimation of ΔCoVaR computed using linear τ-quantiles is more severe for t-Student with lower degrees of freedom ν.*

D.2 List of banks considered in the empirical analysis

TABLE D.1: List of the banks studied in the empirical analysis.

Institution	Country
BANCO SANTANDER SA	Spain
BANCO BILBAO VIZCAYA	Spain
BANCO SABADELL	Spain
INTESA SANPAOLO SPA	Italy
UNICREDIT SPA	Italy
BANCA MONTE PASCHI	Italy
BANCO BPM SPA	Italy
UNIONE DI BAN	Italy
AIB GROUP PLC	Ireland
BANK OF IRELAND	Ireland
BANK OF GREECE SA	Greece
CREDIT SUISSE GROUP	Switzerland
UBS GROUP AG	Switzerland
DANSKE BANK A/S	Denmark
JYSKE BANK A/S	Denmark
NORDEA BANK	Denmark
DNB ASA	Norway
SKANDINAVISKA ENSK	Sweden
SV. HANDELSBANKEN AB	Sweden
SWEDBANK AB	Sweden
HSBC HOLDINGS PLC	United Kingdom
LLOYDS BANKING GROUP	United Kingdom
BARCLAYS PLC	United Kingdom
ROYAL BANK	United Kingdom
STANDARD CHARTERED	United Kingdom
KOMERCNI BANKA, A.S.	Czech Republic
BNP PARIBAS SA	France
CREDIT AGRICOLE SA	France
STE. GENL. DE FRANCE	France
NATIXIS	France
ING GROEP N.V.	Netherlands
DEUTSCHE BANK AG	Germany
COMMERZBANK AG	Germany
KBC GROUP NV	Belgium
ERSTE GROUP BANK AG	Austria
RAIFFEISEN BANK	Austria

D.3 Optimal communities identified in the networks

TABLE D.2: Communities in 2007–12. The number and order of the communities is selected automatically by the algorithm. We matched the most similar ones across networks to simplify the comparison.

Community	Banks (ΔCoVaR)	Banks (median QGM)
1	Credit Suisse Group (CH), Bank Of Greece Sa (GR), UBS Group Ag (CH), ING Groep N.V. (NL), Deutsche Bank Ag (D), Commerzbank Ag (D), Erste Group Bank Ag (A), Raiffeisen Bank (A)	Bank Of Greece Sa (GR), ING Groep N.V. (NL), Erste Group Bank Ag (A), Raiffeisen Bank (A), KBC Group Nv (B)
2	AIB Group Plc (IRL), Bank Of Ireland (IRL), Hsbc Holdings Plc (GB), Lloyds Banking Group (GB), Barclays Plc (GB), Royal Bank (GB), Standard Chartered (GB), KBC Group Nv (B)	AIB Group Plc (IRL), Bank Of Ireland (IRL), Hsbc Holdings Plc (GB), Lloyds Banking Group (GB), Barclays Plc (GB), Royal Bank (GB), Standard Chartered (GB)
3	Danske Bank A/S (S), Jyske Bank A/S (S), Nordea Bank (S), DNB Asa, Skandinaviska Ensk, Sv. Handelsbanken Ab, Swedbank Ab, Komercni Banka, A.S. (CZ)	Danske Bank A/S (S), Jyske Bank A/S (S), Nordea Bank (S), DNB Asa, Skandinaviska Ensk, Sv. Handelsbanken Ab, Swedbank Ab, Komercni Banka, A.S. (CZ)
4	BNP Paribas Sa (F), Credit Agricole Sa (F), Ste. Genl. De France (F), Natixis (F)	BNP Paribas Sa (F), Credit Agricole Sa (F), Ste. Genl. De France (F), Natixis (F)
5	Intesa Sanpaolo Spa (I), Unicredit Spa (I), Banca Monte Paschi (I), Banco BPM Spa (I), Unione Di Ban (I)	Intesa Sanpaolo Spa (I), Unicredit Spa (I), Banca Monte Paschi (I), Banco BPM Spa (I), Unione Di Ban (I)
6	Banco Santander Sa (E), Banco Bilbao Vizcaya (E), Banco Sabadell (E)	Banco Santander Sa (E), Banco Bilbao Vizcaya (E), Banco Sabadell (E)
7		Credit Suisse Group (CH), UBS Group Ag (CH), Deutsche Bank Ag (D), Commerzbank Ag (D)

TABLE D.3: Communities in 2013-18. The number and order of the communities is selected automatically by the algorithm. We matched the most similar ones across networks to simplify the comparison.

Community	Banks (ΔCoVaR)	Banks (median QGM)
1	Bank Of Greece Sa, Credit Suisse Group (CH), UBS Group Ag (CH), Hsbc Holdings Plc (GB), Lloyds Banking Group (GB), Barclays Plc (GB), Royal Bank (GB), Standard Chartered (GB), Deutsche Bank Ag (D), Commerzbank Ag (D)	Hsbc Holdings Plc (GB), Lloyds Banking Group (GB), Barclays Plc (GB), Royal Bank (GB), Standard Chartered (GB)
2	AIB Group Plc (IRL), Danske Bank A/S (S), Jyske Bank A/S (S), Nordea Bank (S), DNB Asa, Skandinaviska Ensk, Sv. Handelsbanken Ab, Swedbank Ab, Komercni Banka, A.S. (CZ)	Danske Bank A/S (S), Jyske Bank A/S (S), Nordea Bank (S), DNB Asa, Skandinaviska Ensk, Sv. Handelsbanken Ab, Swedbank Ab, Komercni Banka, A.S. (CZ)
3	Bank Of Ireland (IRL), BNP Paribas Sa (F), Credit Agricole Sa (F), Ste. Genl. De France (F), Natixis (F), ING Groep N.V. (NL), KBC Group Nv (B), Erste Group Bank Ag (A), Raiffeisen Bank (A)	Bank Of Ireland (IRL), BNP Paribas Sa (F), Credit Agricole Sa (F), Ste. Genl. De France (F), Natixis (F), ING Groep N.V. (NL), KBC Group Nv (B), Erste Group Bank Ag (A), Raiffeisen Bank (A)
4	Intesa Sanpaolo Spa (I), Unicredit Spa (I), Banca Monte Paschi (I), Banco BPM Spa (I), Unione Di Ban (I)	Intesa Sanpaolo Spa (I), Unicredit Spa (I), Banca Monte Paschi (I), Banco BPM Spa (I), Unione Di Ban (I)
5	Banco Santander Sa (E), Banco Bilbao Vizcaya (E), Banco Sabadell (E)	Banco Santander Sa (E), Banco Bilbao Vizcaya (E), Banco Sabadell (E), AIB Group Plc (IRL)
6		Bank Of Greece Sa, Credit Suisse Group (CH), UBS Group Ag (CH), Deutsche Bank Ag (D), Commerzbank Ag (D)

Bibliography

Abbassi, Puriya et al. (2017). "Credit risk interconnectedness: What does the market really know?" In: *Journal of Financial Stability* 29, pp. 1–12.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2015). "Systemic risk and stability in financial networks". In: *The American Economic Review* 105.2, pp. 564–608.

Acharya, Viral, Robert Engle, and Matthew Richardson (2012). "Capital shortfall: A new approach to ranking and regulating systemic risks". In: *American Economic Review* 102.3, pp. 59–64.

Adrian, Tobias and Markus K Brunnermeier (2016). "CoVaR". In: *American Economic Review* 106.7, pp. 1705–41.

Albert, Réka and Albert-László Barabási (2002). "Statistical mechanics of complex networks". In: *Rev. Mod. Phys.* 74 (1), pp. 47–97.

Aldasoro, Iñaki and Iván Alves (2018). "Multiplex interbank networks and systemic importance An application to European data". In: *Journal of Financial Stability* 35.C, pp. 17–37. DOI: 10.1016/j.jfs.2016.12.008.

Ali, Alnur, J Zico Kolter, and Ryan J Tibshirani (2016). "The multiple quantile graphical model". In: *Advances in Neural Information Processing Systems*, pp. 3747–3755.

Allen, Franklin and Douglas Gale (2000). "Financial contagion". In: *Journal of Political Economy* 108.1, pp. 1–33.

Alter, Adrian, Ben R Craig, and Peter Raupach (2015). "Centrality-based capital allocations". In: *International Journal of Central Banking* 11.3, pp. 329–377.

Amini, Hamed, Rama Cont, and Andreea Minca (2016). "Resilience to contagion in financial networks". In: *Mathematical Finance* 26.2, pp. 329–365.

Anand, Kartik, Ben Craig, and Goetz Von Peter (2015). "Filling in the blanks: Network structure and interbank contagion". In: *Quantitative Finance* 15.4, pp. 625–636.

Anufriev, Mikhail and Valentyn Panchenko (2015). "Connecting the dots: Econometric methods for uncovering networks with an application to the Australian financial institutions". In: *Journal of Banking & Finance* 61, pp. 241–255.

Anufriev, Mikhail et al. (2016). "A model of network formation for the overnight interbank market". In: *Available on SSRN*.

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9.3, pp. 203–228.

Baba, Kunihiro, Ritei Shibata, and Masaaki Sibuya (2004). "Partial correlation and conditional correlation as measures of conditional independence". In: *Australian & New Zealand Journal of Statistics* 46.4, pp. 657–664.

Baitinger, Eduard and Jochen Papenbrock (2017). "Interconnectedness Risk and Active Portfolio Management: The Information-Theoretic Perspective". In:

Balla, Eliana, Ibrahim Ergen, and Marco Migueis (2014). "Tail dependence and indicators of systemic risk for large US depositories". In: *Journal of Financial Stability* 15, pp. 195–209.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont (2008). "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data". In: *Journal of Machine Learning Research* 9, pp. 485–516.

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *science* 286.5439, pp. 509–512.

Barrat, Alain et al. (2004). "The architecture of complex weighted networks". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.11, pp. 3747–3752.

Barucca, Paolo and Fabrizio Lillo (2018). "The organization of the interbank network and how ECB unconventional measures affected the e-MID overnight market". In: *Computational Management Science* 15.1, pp. 33–53.

Basel III, BCBS (2013). "The Liquidity Coverage Ratio and liquidity risk monitoring tools". In: *Bank for International Settlements*.

Battiston, Stefano et al. (2012a). "Debtrank: Too central to fail? Financial networks, the FED and systemic risk". In: *Scientific Reports* 2.

Battiston, Stefano et al. (2012b). "Default cascades: When does risk diversification increase stability?" In: *Journal of Financial Stability* 8.3, pp. 138–149.

BCBS, June (2011). "Basel III: A global regulatory framework for more resilient banks and banking systems". In: *Bank for International Settlements*.

Bekiros, Stelios et al. (2017). "Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets". In: *European Journal of Operational Research* 256.3, pp. 945–961.

Belloni, Alexandre, Mingli Chen, and Victor Chernozhukov (2016). "Quantile Graphical Models: Prediction and Conditional Independence with Applications to Financial Risk Management". In: *arXiv preprint arXiv:1607.00286*.

Belloni, Alexandre, Victor Chernozhukov, et al. (2011). "ℓ1-penalized quantile regression in high-dimensional sparse models". In: *The Annals of Statistics* 39.1, pp. 82–130.

Benoit, Sylvain et al. (2017). "Where the risks lie: A survey on systemic risk". In: *Review of Finance* 21.1, pp. 109–152.

Billio, Monica et al. (2012). "Econometric measures of connectedness and systemic risk in the finance and insurance sectors". In: *Journal of Financial Economics* 104.3, pp. 535–559.

Black, Fischer and Robert Litterman (1992). "Global portfolio optimization". In: *Financial Analysts Journal* 48.5, pp. 28–43.

Bollerslev, Tim (1990). "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model". In: *The Review of Economics and Statistics*, pp. 498–505.

Bonaccolto, Giovanni, Massimiliano Caporin, and Sandra Paterlini (2018). "Conditional Quantile-Located VaR". In: *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer, pp. 167–171.

Bonacich, Phillip (1987). "Power and centrality: A family of measures". In: *American Journal of sociology* 92.5, pp. 1170–1182.

Bondell, Howard D, Brian J Reich, and Huixia Wang (2010). "Noncrossing quantile regression curve estimation". In: *Biometrika* 97.4, pp. 825–838.

Borgatti, Stephen P and Martin G Everett (2000). "Models of core/periphery structures". In: *Social Networks* 21.4, pp. 375–395.

Bouchaud, Jean-Philippe and Marc Potters (2009). "Financial applications of random matrix theory: a short review". In: *arXiv preprint arXiv:0910.1205*.

Brodie, J. et al. (2009). "Sparse and Stable Markowitz Portfolios". In: *Proceedings of the National Academy of Science* 106(30), pp. 12267–12272.

Brownlees, Christian and Robert F Engle (2016). "SRISK: A conditional capital shortfall measure of systemic risk". In: *The Review of Financial Studies* 30.1, pp. 48–79.

Brownlees, Christian T and Geert Mesters (2017). "Detecting granular time series in large panels". In: *Working Paper, available on SSRN 3040636*.

Brownlees, Christian T, Eulalia Nualart, and Yucheng Sun (2018). "Realized Networks". In: *Journal of Applied Econometrics* available online.

Bruder, Benjamin et al. (2013). "Regularization of Portfolio Allocation." In: Working Paper, available on SSRN.

Brusco, Sandro and Fabio Castiglionesi (2007). "Liquidity coinsurance, moral hazard, and financial contagion". In: *The Journal of Finance* 62.5, pp. 2275–2302.

Caccioli, Fabio, Thomas A Catanach, and J Doyne Farmer (2012). "Heterogeneity, correlations and financial contagion". In: *Advances in Complex Systems* 15.2, p. 1250058.

Caccioli, Fabio et al. (2014). "Stability analysis of financial contagion due to overlapping portfolios". In: *Journal of Banking & Finance* 46, pp. 233–245.

Centola, Damon, Víctor M Eguíluz, and Michael W Macy (2007). "Cascade dynamics of complex propagation". In: *Physica A: Statistical Mechanics and its Applications* 374.1, pp. 449–456.

Centola, Damon and Michael Macy (2007). "Complex contagions and the weakness of long ties". In: *American journal of Sociology* 113.3, pp. 702–734.

Chinazzi, Matteo and Giorgio Fagiolo (2015). "Systemic risk, contagion, and financial networks: A survey". In: *Institute of Economics, Scuola Superiore SantâĂŹAnna, Laboratory of Economics and Management (LEM) Working Paper Series* 2013/08.

Choueifaty, Yves and Yves Coignard (2008). "Toward maximum diversification". In: *Journal of Portfolio Management* 35.1, p. 40.

Chun, Hyonho et al. (2016). "Graphical models via joint quantile regression with component selection". In: *Journal of Multivariate Analysis* 152, pp. 162–171.

Cifuentes, Rodrigo, Gianluigi Ferrucci, and Hyun Song Shin (2005). "Liquidity risk and contagion". In: *Journal of the European Economic Association* 3.2-3, pp. 556–566.

Cont, R. (2001). "Empirical properties of asset returns: Stylized facts and statistical issues". In: *Quantitative Finance* 1, pp. 223–236.

Cont, Rama and Andreea Minca (2016). "Credit default swaps and systemic risk". In: *Annals of Operations Research* 247.2, pp. 523–547.

Craig, Ben and Martín Saldías (2016). *Spatial Dependence and Data-Driven Networks of International Banks*. International Monetary Fund, WP/16/184.

Craig, Ben and Goetz Von Peter (2014). "Interbank tiering and money center banks". In: *Journal of Financial Intermediation* 23.3, pp. 322–347.

Dash, Rajashree and Pradipta Kishore Dash (2016). "A hybrid stock trading framework integrating technical analysis with machine learning techniques". In: *The Journal of Finance and Data Science* 2.1, pp. 42–57.

De Miguel, V. and F. J. Nogales (2009). "Portfolio selection with robust estimation". In: *Operations Research* 57, pp. 560–577.

De Miguel, V. et al. (2009). "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norm". In: *Management Science* 55, pp. 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *Review of Financial Studies* 22.5, pp. 1915–1953.

Dempster, Arthur P (1972). "Covariance selection". In: *Biometrics* 28.1, pp. 157–175.

Diebold, Francis X and Kamil Yılmaz (2014). "On the network topology of variance decompositions: Measuring the connectedness of financial firms". In: *Journal of Econometrics* 182.1, pp. 119–134.

Ding, Peng (2016). "On the conditional distribution of the multivariate t distribution". In: *The American Statistician* 70.3, pp. 293–295.

Eisenberg, Larry and Thomas H Noe (2001). "Systemic risk in financial systems". In: *Management Science* 47.2, pp. 236–249.

Elsinger, Helmut et al. (2013). "Network models and systemic risk assessment". In: *Handbook on Systemic Risk* 1, pp. 287–305.

Engle, RF and CT Brownlees (2010). "Volatility, correlation and tails for systemic risk measurement". Manuscript, Stern School of Business, New York University.

Engle, Robert (2002). "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

Engle, Robert F and Raul Susmel (1993). "Common volatility in international equity markets". In: *Journal of Business & Economic Statistics* 11.2, pp. 167–176.

Fagiolo, Giorgio (2007). "Clustering in complex directed networks". In: *Physical Review E* 76.2, p. 026107.

Fan, Jianqing and Runze Li (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties". In: *Journal of the American Statistical Association* 96.456, pp. 1348–1360.

Fan, Jianqing, Lingzhou Xue, and Hui Zou (2014). "Strong oracle optimality of folded concave penalized estimation". In: *Annals of Statistics* 42.3, p. 819.

Fan, Jianqing, Jingjin Zhang, and Ke Yu (2012). "Vast portfolio selection with gross-exposure constraints". In: *Journal of the American Statistical Association* 107.498, pp. 592–606.

Fernandes, Guilherme Barreto and Rinaldo Artes (2016). "Spatial dependence in credit risk and its improvement in credit scoring". In: *European Journal of Operational Research* 249.2, pp. 517–524.

Finegold, Michael and Mathias Drton (2011). "Robust graphical modeling of gene networks using classical and alternative t-distributions". In: *The Annals of Applied Statistics* 5.2A, pp. 1057–1080.

Finger, Karl, Daniel Fricke, and Thomas Lux (2013). "Network analysis of the e-MID overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes". In: *Computational Management Science* 10.2-3, pp. 187–211.

Finger, Karl and Thomas Lux (2017). "Network formation in the interbank money market: An application of the actor-oriented model". In: *Social Networks* 48, pp. 237–249.

Fortunato, Santo (2010). "Community detection in graphs". In: *Physics Reports* 486.3, pp. 75–174.

Foygel, Rina and Mathias Drton (2010). "Extended Bayesian information criteria for Gaussian graphical models". In: *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*. ISBN: 9781617823800.

Freeman, Linton C (1978). "Centrality in social networks conceptual clarification". In: *Social Networks* 1.3, pp. 215–239.

Freixas, Xavier, Luc Laeven, and José-Luis Peydró (2015). *Systemic risk, crises, and macroprudential regulation*. MIT Press.

Fricke, Daniel and Thomas Lux (2015a). "Core–periphery structure in the overnight money market: evidence from the e-MID trading platform". In: *Computational Economics* 45.3, pp. 359–395.

– (2015b). "On the distribution of links in the interbank network: Evidence from the e-MID overnight money market". In: *Empirical Economics* 49.4, pp. 1463–1495.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2014). "glasso: Graphical lasso-estimation of Gaussian graphical models". In: *R package*.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2008). "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3, pp. 432–441.

FSB (2013). "Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement". In: *Financial Stability Board, BIS*. URL: `http://www.bis.org/publ/bcbs255.pdf`.

Gai, Prasanna, Andrew Haldane, and Sujit Kapadia (2011). "Complexity, concentration and contagion". In: *Journal of Monetary Economics* 58.5, pp. 453–470.

Gai, Prasanna and Sujit Kapadia (2010). "Contagion in financial networks". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 466. 2120. The Royal Society, pp. 2401–2423.

Galstyan, Aram and Paul Cohen (2007). "Cascading dynamics in modular networks". In: *Physical Review E* 75.3, p. 036109.

Giacometti, Rosella, Gabriele Torri, and Sandra Paterlini (2020). "Tail Risks in Large Portfolio Selection: Penalized Quantile and Expectile Minimum Deviation Models". In: *Available at SSRN 3587466*.

Girardi, Giulio and A Tolga Ergün (2013). "Systemic risk measurement: Multivariate GARCH estimation of CoVaR". In: *Journal of Banking & Finance* 37.8, pp. 3169–3180.

Girvan, Michelle and Mark EJ Newman (2002). "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 7821–7826.

Goh, K-I, Byungnam Kahng, and D Kim (2001). "Universal behavior of load distribution in scale-free networks". In: *Physical Review Letters* 87.27, p. 278701.

Goto, Shingo and Yan Xu (2015). "Improving mean variance optimization through sparse hedging restrictions". In: *Journal of Financial and Quantitative Analysis* 50.06, pp. 1415–1441.

Granovetter, Mark S (1977). "The strength of weak ties". In: *Social networks*. Elsevier, pp. 347–367.

Greppi, Alessandro et al. (2018). "Bayesian Networks for Financial Market Signals Detection". In: *Classification,(Big) Data Analysis and Statistical Learning*. Springer, pp. 219–226.

Haldane, Andrew G et al. (2009). "Rethinking the financial network". In: *Speech delivered at the Financial Student Association, Amsterdam, April* 28.

Hanley, James A and Barbara J McNeil (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology* 143.1, pp. 29–36.

Hautsch, Nikolaus, Julia Schaumburg, and Melanie Schienle (2014). "Financial network systemic risk contributions". In: *Review of Finance* 19.2, pp. 685–738.

Højsgaard, Søren, David Edwards, and Steffen Lauritzen (2012). *Graphical models with R*. Springer Science & Business Media.

Huang, Wei and Chunguang Li (2007). "Epidemic spreading in scale-free networks with community structure". In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.01, P01014.

Huang, Wei-Qiang and Stan Uryasev (2018). "The CoCVaR approach: systemic risk contribution measurement". In: *Journal of Risk* 20.4, pp. 75–93.

Hübsch, Arnd and Ursula Walther (2017). "The impact of network inhomogeneities on contagion and system stability". In: *Annals of Operations Research* 254.1-2, pp. 61–87.

Hurd, Tom R (2016). *Contagion! Systemic Risk in Financial Networks*. Springer.

Hüttner, Amelie, Jan-Frederik Mai, and Stefano Mineo (2016). "Portfolio selection based on graphs: Does it align with Markowitz-optimal portfolios?" In: *Dependence Modeling* 6.1, pp. 63–87.

Idier, Julien, Gildas Lamé, and Jean-Stéphane Mésonnier (2014). "How useful is the Marginal Expected Shortfall for the measurement of systemic exposure? A practical assessment". In: *Journal of Banking & Finance* 47, pp. 134–146.

Iori, Giulia, Saqib Jafarey, and Francisco G Padilla (2006). "Systemic risk on the interbank market". In: *Journal of Economic Behavior & Organization* 61.4, pp. 525–542.

Iori, Giulia et al. (2015). "Networked relationships in the e-MID Interbank market: A trading model with memory". In: *Journal of Economic Dynamics and Control* 50, pp. 98–116.

Jenkins, Nicole Thorne, Michael D Kimbrough, and Juan Wang (2016). "The extent of informational efficiency in the credit default swap market: evidence from post-earnings announcement returns". In: *Review of Quantitative Finance and Accounting* 46.4, pp. 725–761.

Justel, Ana, Daniel Peña, and Rubén Zamar (1997). "A multivariate Kolmogorov-Smirnov test of goodness of fit". In: *Statistics & Probability Letters* 35.3, pp. 251–259.

Kaltwasser, Pablo Rovira and Alessandro Spelta (2018). "Identifying systemically important financial institutions: a network approach". In: *Computational Management Science*, pp. 1–31.

Koenker, Roger (2005). *Quantile regression, volume 38 of Econometric Society Monographs*.

Koenker, Roger and Kevin F Hallock (2001). "Quantile regression". In: *Journal of Economic Perspectives* 15.4, pp. 143–156.

Kolm, P. N., R. Tütüncü, and F.J Fabozzi (2014). "60 years following Harry Markowitz's contribution to portfolio theory and operations research". In: *European Journal of Operational Research* 234 (2), pp. 343–582.

Kotz, Samuel and Saralees Nadarajah (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.

Kremer, Philipp J., Andreea Talmaciu, and Sandra Paterlini (2017). "Risk Minimization in Multi-Factor Portfolios: What is the Best Strategy?" In: *Annals of Operations Research*, pp. 1–37.

Laloux, Laurent et al. (1999). "Noise dressing of financial correlation matrices". In: *Physical Review Letters* 83.7, pp. 1467–1469.

Laloux, Laurent et al. (2000). "Random matrix theory and financial correlations". In: *International Journal of Theoretical and Applied Finance* 3.03, pp. 391–397.

Lam, Clifford and Jianqing Fan (2009). "Sparsistency and rates of convergence in large covariance matrix estimation". In: *Annals of Statistics* 37.6B, p. 4254.

Lange, Kenneth L, Roderick JA Little, and Jeremy MG Taylor (1989). "Robust statistical modeling using the t distribution". In: *Journal of the American Statistical Association* 84.408, pp. 881–896.

Lauritzen, Steffen L (1996). *Graphical models*. Clarendon Press.

Ledoit, Olivier and Michael Wolf (2004a). "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of Multivariate Analysis* 88.2, pp. 365–411.

– (2004b). "Honey, I shrunk the sample covariance matrix". In: *The Journal of Portfolio Management* 30.4, pp. 110–119.

– (2011). "Robust performances hypothesis testing with the variance". In: *Wilmott* 2011.55, pp. 86–89.

Lee, Seung Hwan (2013). "Systemic liquidity shortages and interbank network structures". In: *Journal of Financial Stability* 9.1, pp. 1–12.

Lelyveld, Iman van and Daan in 't Veld (2014). "Finding the core: Network structure in interbank markets". In: *Journal of Banking & Finance* 49, pp. 27–40.

Loh, Po-Ling and Martin J Wainwright (2013). "Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima". In: *Advances in Neural Information Processing Systems*, pp. 476–484.

Mainik, Georg and Eric Schaanning (2014). "On dependence consistency of CoVaR and some other systemic risk measures". In: *Statistics & Risk Modeling* 31.1, pp. 49–77.

Mantegna, Rosario N (1999). "Hierarchical structure in financial markets". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 11.1, pp. 193–197.

Markowitz, Harry (1952). "Portfolio selection". In: *The Journal of Finance* 7.1, pp. 77–91.

McAssey, Michael P (2013). "An empirical goodness-of-fit test for multivariate distributions". In: *Journal of Applied Statistics* 40.5, pp. 1120–1131.

McLachlan, Geoffrey and Thriyambakam Krishnan (2007). *The EM algorithm and extensions*. John Wiley & Sons.

Meucci, Attilio (2009). *Risk and asset allocation*. Springer Science & Business Media.

Michaud, Richard O (1989). "The Markowitz optimization enigma: Is optimized optimal?" In: *ICFA Continuing Education Series* 1989.4, pp. 43–54.

Mistrulli, Paolo Emilio (2011). "Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns". In: *Journal of Banking & Finance* 35.5, pp. 1114–1127.

Mønsted, Bjarke et al. (2017). "Evidence of complex contagion of information in social media: An experiment using Twitter bots". In: *PloS one* 12.9, e0184148.

Murphy, Kevin P (2012). *Machine learning: a probabilistic perspective*. MIT press.

Newman, Mark EJ (2002). "Spread of epidemic disease on networks". In: *Physical review E* 66.1, p. 016128.

– (2004). "Fast algorithm for detecting community structure in networks". In: *Physical review E* 69.6, p. 066133.

– (2010). *Networks: an introduction*. Oxford university press.

Newman, Mark EJ and Juyong Park (2003). "Why social networks are different from other types of networks". In: *Physical Review E* 68.3, p. 036122.

O'Sullivan, David JP et al. (2015). "Mathematical modeling of complex contagion on clustered networks". In: *Frontiers in Physics* 3, p. 71.

Page, Lawrence et al. (1999). *The PageRank citation ranking: Bringing order to the web.* Tech. rep. Stanford InfoLab.

Paltalidis, Nikos et al. (2015). "Transmission channels of systemic risk and contagion in the European financial network". In: *Journal of Banking & Finance* 61, S36–S52.

Pastor-Satorras, Romualdo et al. (2015). "Epidemic processes in complex networks". In: *Reviews of Modern Physics* 87.3, p. 925.

Patel, Jigar et al. (2015). "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques". In: *Expert Systems with Applications* 42.1, pp. 259–268.

Pecora, Nicolò, Pablo Rovira Kaltwasser, and Alessandro Spelta (2016). "Discovering SIFIs in interbank communities". In: *PloS one* 11.12, e0167781.

Peralta, Gustavo and Abalfazl Zareei (2016). "A network approach to portfolio selection". In: *Journal of Empirical Finance* 38, pp. 157–180.

Poon, Ser-Huang, Michael Rockinger, and Jonathan Tawn (2003). "Extreme value dependence in financial markets: Diagnostics, models, and financial implications". In: *The Review of Financial Studies* 17.2, pp. 581–610.

Puliga, Michelangelo, Guido Caldarelli, and Stefano Battiston (2014). "Credit default swaps networks and systemic risk". In: *Scientific Reports* 4.

Reid, Fergal and Neil Hurley (2011). "Diffusion in networks with overlapping community structure". In: *arXiv preprint arXiv:1105.5849*.

Rigobon, Roberto (2016). "Contagion, spillover and interdependence". In: *European Central Bank, Working Paper Series* No 1975 / November 2016.

Rockafellar, R Tyrrell and Stan Uryasev (2013). "The fundamental risk quadrangle in risk management, optimization and statistical estimation". In: *Surveys in Operations Research and Management Science* 18.1-2, pp. 33–53.

Roncalli, T. (2014). *Introduction to risk parity and budgeting*. Financial Mathematics Series. Chapman & Hall / CRC.

Rothman, Adam J et al. (2008). "Sparse permutation invariant covariance estimation". In: *Electronic Journal of Statistics* 2, pp. 494–515.

Salathé, Marcel and James H Jones (2010). "Dynamics and control of diseases in networks with community structure". In: *PLoS Computational Biology* 6.4, e1000736.

Schwarcz, Steven L (2008). "Systemic risk". In: *Geo. LJ* 97, p. 193.

Stegehuis, Clara, Remco van der Hofstad, and Johan SH van Leeuwaarden (2016). "Epidemic spreading on complex networks with community structures". In: *Scientific Reports* 6.

Stevens, Guy VG (1998). "On the inverse of the covariance matrix in portfolio analysis". In: *The Journal of Finance* 53.5, pp. 1821–1827.

Temizsoy, Asena, Giulia Iori, and Gabriel Montes-Rojas (2015). “The role of bank relationships in the interbank market”. In: *Journal of Economic Dynamics and Control* 59, pp. 118–141.

Tibshirani, Robert (1996). “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Tong, Yung Liang (2012). *The multivariate normal distribution*. Springer Science & Business Media.

Torri, Gabriele (2017). “Systemic Risk and Community Structure in the European Banking System”. In: *International Conference Mathematical Methods in Economics (MME)*. University of Hradec Králové, pp. 807–812.

Torri, Gabriele and Rosella Giacometti (2017). “Financial Contagion in International Banking Networks: A Simulation Study”. In: *Financial Management of Firms and Financial Institutions*. VŠB-TU Ostrava, pp. 857–865.

Torri, Gabriele, Rosella Giacometti, and Sandra Paterlini (2018). “Robust and sparse banking network estimation”. In: *European Journal of Operational Research* 270.1, pp. 51–65.

– (2019). “Sparse precision matrices for minimum variance portfolios”. In: *Computational Management Science* 16.3, pp. 375–400.

Tumminello, Michele et al. (2005). “A tool for filtering information in complex systems”. In: *Proceedings of the National Academy of Sciences* 102.30, pp. 10421–10426.

Wagner, Wolf (2010). “Diversification at financial institutions and systemic crises”. In: *Journal of Financial Intermediation* 19.3, pp. 373–386.

Watts, Duncan J and Steven H Strogatz (1998). “Collective dynamics of âĂŸsmall-worldâĂŹ networks”. In: *Nature* 393.6684, pp. 440–442.

Weng, Lilian, Filippo Menczer, and Yong-Yeol Ahn (2013). “Virality prediction and community structure in social networks”. In: *Scientific Reports* 3.

Witten, Daniela M, Jerome H Friedman, and Noah Simon (2011). “New insights and faster computations for the graphical lasso”. In: *Journal of Computational and Graphical Statistics* 20.4, pp. 892–900.

Won, Joong-Ho et al. (2013). “Condition-number-regularized covariance estimation”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.3, pp. 427–450.

Wu, Yichao and Yufeng Liu (2009). “Variable selection in quantile regression”. In: *Statistica Sinica*, pp. 801–817.

Yuan, Ming and Yi Lin (2007). "Model selection and estimation in the Gaussian graphical model". In: *Biometrika* 94.1, pp. 19–35.

Zhang, Cun-Hui et al. (2010). "Nearly unbiased variable selection under minimax concave penalty". In: *The Annals of Statistics* 38.2, pp. 894–942.

Zou, Hui (2006). "The adaptive lasso and its oracle properties". In: *Journal of the American Statistical Association* 101.476, pp. 1418–1429.

Zou, Hui and Runze Li (2008). "One-step sparse estimates in nonconcave penalized likelihood models". In: *Annals of Statistics* 36.4, p. 1509.