ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis - MSc Quantitative Finance

Identifying the underlying graphical structure of systemic risks in the banking sector

Author:
Andrei Iliev
546909

Supervisor:
C., Zhou

Second assessor:
J.A., Oorschot

Abstract

Recent financial crises have made clear the crucial importance of taking into account the complex cross-border and domestic interlinkages across institutions. We follow the approach of Engelke and Hitz (2020) who adapt the notion of conditional independence to extremal data of threshold exceedences and show that this allows the construction of sparse graphical models from variogram matrices which characterise the family of Hüsler-Reiss distributions. We apply extremal graphical models to a dataset comprising 15 years of stock returns on 32 of the largest publicly traded banks across 16 countries in the European Economic Area. Additionally, we apply the key concepts of closeness centrality and betweenness centrality from network theory to identify the set of banks that have systemic importance with respect to the underlying financial network in times of market-wide crises. Our results suggest that extreme events observed in financial data tend to be geographically isolated and are connected to the rest of the network through a small set of systemically important banks.

August 1, 2021

The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Contents

1 Introduction

Abnormal temperature levels, extreme flooding, and financial crises are few examples of events which typically lie in the tail of the distribution and have very small probabilities of occurrence. Such events, however, can have devastating impacts on human life. The accurate statistical modelling of such rare events, thus, is of significant importance.

Extreme value theory (EVT) provides a theoretical framework for constructing models that extrapolate beyond the empirical distribution of a particular variable of interest and allow for statistical inference about events that have never yet occurred. Univariate extremes have been extensively studied and the results for block-maxima and peaks-over-threshold methods have found various applications in many different fields (cf. Beirlant et al. (2004)). In many situations, however, the subject of interest is not just one extremal event but a chain of contemporaneous rare events that lead to a catastrophic situation. For example, extreme flooding caused by abnormal rainfall may be observed in several locations simultaneously. In financial networks, for instance, the risk of systemic instability in the banking sector and collapse of the entire financial system depends strongly on the structural relationships among many financial institutions. For such complex systems, the severity of an extreme episode depends on the strength of the multivariate extremal dependence between univariate extremes.

Multivariate extreme value theory, thus, models the dependence between different components of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. Max-stable distributions and multivariate Pareto distributions arise as the limits of normalised maxima and threshold exceedances of the random vector $\boldsymbol{X}$, respectively. In practice, the complexity of these models is considerable even in moderate dimensions, and for current statistical models tractable statistical inference is limited to the 2- and 3-dimensional case. A recent domain of research in extreme value statistics is centered around the modelling of sparse structures in extreme events through machine learning techniques and graphical models (cf. Engelke and Ivanovs (2021)).

Recent financial crises have made it clear that financial institutions should not be taken in isolation. It is of crucial importance for the stability of the financial system to take into account the complex cross-border and domestic interlinkages across institutions.

In this paper, we focus on identifying the underlying network for extreme events that drive systemic risks in the banking sector using extremal graphical models. We achieve that by following the approach of Engelke and Hitz (2020) who adapt the notion of conditional independence to extremal data of threshold exceedances and show that this allows the construction of possibly sparse graphical models from variogram matrices which characterise the family of Hüsler-Reiss distributions.

This paper is organized as follows. Section 2 provides an overview of the relevant literature on extreme value theory and graphical models. Section 3 gives information on the data sets used throughout the paper and presents summary statistics illustrating the "stylized facts" of real-life financial time-series that influence the modelling techniques used in the Methodology section. In Section 4, we introduce the theoretical concepts and modelling techniques required to construct sparse block graphs for extreme events in the banking system and to identify systemically important banks. In Section 5, we apply our model construction methodology to a data set comprising 32 banks and discuss the optimal graphical structure and the implications this has for systemic importance. Section 6 draws the conclusion.

2 Literature Review

Classical univariate EVT investigates the question how to make consistent statistical inference about the distribution of extreme values in a population of interest. The two most popular methods for modelling extreme values are the block maxima (BM) and peaks-over-threshold (POT) methods. In the block-maxima approach, pioneered in applications dealing with hydrology and climatology such as river heights and sea levels, data is split into several non-overlapping, equally-sized time periods, or blocks (see the plot on the left side of Figure 1 for an illustration). The normalized maxima within each block are assumed to be independent and identically distributed, and to converge to a non-degenerate limiting distribution. In analogy to the central limit theorem, Fisher and Tippett (1928) , showed that the distribution of the limit must belong to the class of generalized extreme value (GEV) distributions. From practical considerations, the block maxima method could be considered wasteful of important information, since it discards

all non-maximal data points. The peaks-over-threshold approach has been the preferred workhorse in practice since it models all large observations that exceed a designated high threshold and therefore uses the whole extremal dataset (right plot in Figure 1). Asymptotically, the exceedances have been suggested to occur according to a Poisson process and the excesses (the sizes by which the threshold is exceeded) have been shown to converge to a generalized Pareto distribution (cf., Pickands (1975)). Another important result demonstrated by Pickands (1975) and Davison and Smith (1990), is that a necessary condition for the existence of limit results for threshold exceedances is the convergence of the corresponding block maxima to a GEV distribution. This close link between limit results for threshold exceedances and limit results for block maxima is often exploited in practical applications. We refer to Embrechts et al. (1997) and Davison and Huser (2015) for key results and detailed overview of univariate extreme value statistics.

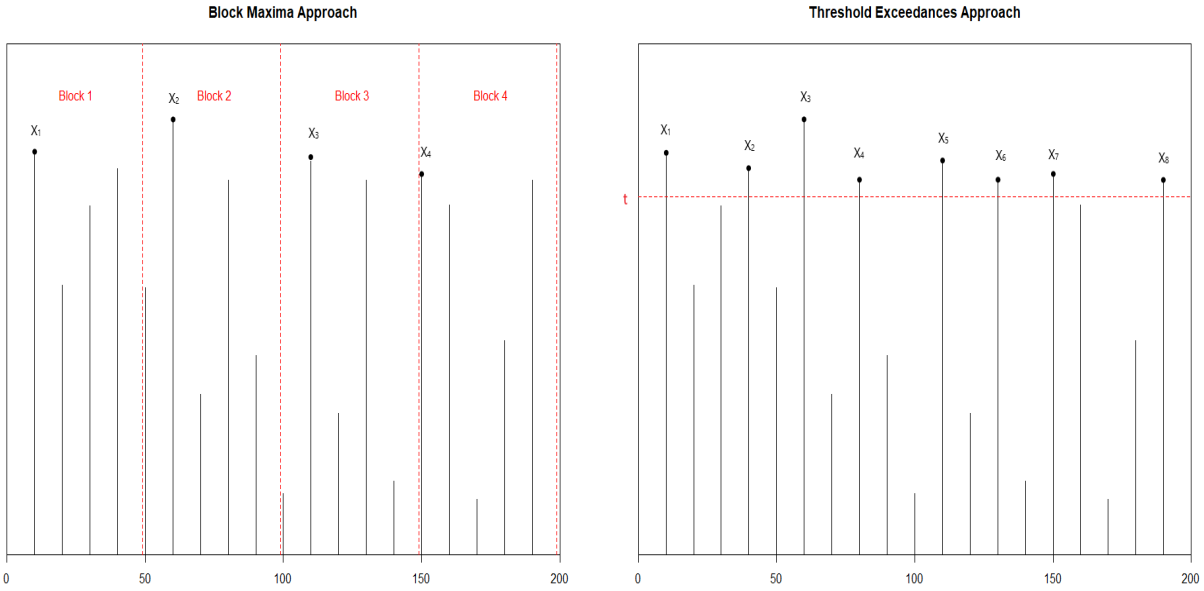


Figure 1: The Block Maxima (left) and Peaks-Over-Threshold (right) Approach

For applications in practice, such as estimating the tail risk of a portfolio of assets or the systemic risk in the banking sector, we are predominantly interested in multiple events that drive an extreme episode. Multivariate extreme value theory studies the joint distribution of extremes of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. In this setting, the problem of modelling the tail behaviour of a random vector $\boldsymbol{X}$ is typically split into modelling of the marginal tail of each component of $\boldsymbol{X}$ and modelling of the extremal dependence structure. Similar to the univariate setting, two different

but related directions have been explored - the component-wise maxima approach and the peaks-over-threshold approach. In the former approach, the asymptotic behaviour of normalised component-wise maxima of i.i.d. random vectors is represented in terms of max-stable distributions and various probabilistic properties and statistical aspect have been established (see de Haan and Resnick (1977), Smith et al. (1990) and Tawn (1990)). The latter approach proceeds by choosing a suitable high threshold for each component of $\boldsymbol{X}$ and considers an event to be extreme when at least one component exceeds its threshold. Thus, one models the whole tail of a given extreme episode and, in theory, might be able to capture the variation and extremal dependence of the components in greater detail compared to the maxima approach. Multivariate generalized Pareto (GP) distributions are the only stable distributions that arise as the limit of normalized threshold exceedances (Rootzén and Tajvidi (2006)). However, for both multivariate maxima and multivariate threshold exceedances the construction of tractable models and robust statistical inference in high dimensions is challenging and most current applications are only limited to moderate dimensions.

These limitations have influenced a relatively new body of research focusing on the adaptation of classical methods for dimension reduction and detection of sparse models to multivariate extreme data. Advances in this field link approaches from machine learning (such as clustering and principal component analysis) and graphical modelling to multivariate EVT (see Engelke and Ivanovs (2021) for a review of recent advances). Graphical models in the classical, non-extreme setting rely on the notion of conditional independence for dimension reduction through the factorization of continuous high-dimensional distributions to lower dimensions. In a recent work, Engelke and Hitz (2020), introduce a general theory of extremal conditional independence for muiltivariate Pareto distributions that allows the construction of possibly sparse graphical structures. In particular, they show that for Hüsler-Reiss distributions the underlying tree structure can be learned and expanded to block-graphs by likelihood-based methods and statistical inference can be carried out on lower-dimensional densities.

With this paper, we aim to shed light on the complex relationships occurring between financial institutions in times of financial crises. To achieve that, we focus on the identification of the underlying network for extreme events that drives systemic risks in the banking industry through the application of extremal graphical models to a dataset

comprising the largest financial institutions in the European Economic Area (EEA). In addition, we apply key concepts from network theory, known as centrality measures, to identify the set of banks that have systemic importance with respect to the underlying financial network in times of market-wide crises. Our methodology bridges together financial time-series modelling with graphical modelling for threshold exceedances which can allow policymakers, academic economists and practitioners in the banking sector to better understand the mechanisms that drive financial distress.

3 Data

For our analysis, we have selected some of the largest publicly traded banks in the European Economic Area, sorted by their market capitalisation, together with financial institutions from the countries most heavily hit by the 2008 financial crisis, namely Portugal, Greece, and Ireland. Our dataset was obtained from Datastream and consists of nearly 15 years of data measured in terms of daily stock prices taken at closing times and adjusted for splits, dividends, and other capital actions. The time period considered is from April 25, 2005 to December 31, 2020, with holidays and market closures treated as missing values. This yields a dataset comprising $n = 4094$ observations on $d = 32$ financial institutions (cf. Table 1)

A first step in our analysis is to transform our data into daily return series

$$S_t^{(1)}, \ldots, S_t^{(d)}, \qquad \text{with } t \in \{1, \ldots, n\}$$

by taking the logarithmic difference of the price as

$$S_t^d = 100 * \log \frac{P_t^d}{P_{t-1}^d}, \tag{1}$$

where P_t^d denotes the closing price on day t of asset d.

ID	Bank	Symbol	Country	Market Cap	Rank
1	Ereste Group Bank AG	EBS	Austria	11.03	25
2	Raiffeisen Bank Intl AG	RBI	Austria	6.34	28
3	KBC Group NV	KBC	Belgium	17.86	21
4	Danske Bank A/S	DANSKE	Denmark	17.67	22
5	Nordea Bank Abp	NDA-Fl	Finland	33.10	11
6	BNP Paribas SA	BNP	France	59.47	3
7	Crédit Agricole SA	ACA	France	26.98	17
8	Societe Generale SA	GLE	France	28.78	16
9	Commerzbank AG	CBK	Germany	10.05	26
10	Deutsche Bank AG	DBK	Germany	28.85	15
11	National Bank of Greece	ETE	Greece	5.20	30
12	AIB Group PLC	AIBG	Ireland	21.18	18
13	Bank of Ireland PLC	BIRG	Ireland	5.76	29
14	Banco BPM SpA	BPM	Italy	3.54	31
15	Intesa Sanpaolo SpA	ISP	Italy	36.09	10
16	UniCredit SpA	UCG	Italy	30.465	14
17	ING Groep NV	INGA	Netherlands	37.402	9
18	DNB ASA	DNB	Norway	18.45	19
19	Banco Comercial Portugues SA	BCP	Portugal	3.26	32
20	Banco de Sabadell SA	SAB	Spain	6.78	27
21	Banco Santander SA	SAN	Spain	71.03	2
22	BBVA SA	BBVA	Spain	40.94	6
23	Skandinaviska Enskilda Banken AB	SEB-A	Sweden	16.2	24
24	Svenska Handelsbanken AB	SHB-A	Sweden	18.09	20
25	Swedbank AB	SWED-A	Sweden	16.71	23
26	Credit Suisse Group AG	CSGN	Switzerland	32.38	13
27	UBS Group AG	UBSG	Switzerland	48.28	4
28	Barclays PLC	BARC	UK	38.87	8
29	HSBC Holdings PLC	HSBC	UK	137.81	1
30	Lloyds Banking Group PLC	LLOY	UK	47.34	5
31	Natwest Group PLC	NWG	UK	39.43	7
32	Standard Chartered PLC	STAN	UK	32.77	12

Table 1: List of the European financial institutions used in our analysis with index number, ticker symbol, country, mean market capitalization (in EUR) and rank with respect to the mean market capitalization in the period from April 25,2005 to December 31, 2020.

Financial time-series data is known to exhibit some particular empirical properties which make their modelling a non-trivial task. Some of these "stylized facts" include non-iid return series, strong serial correlation of absolute and squared returns and non-normality (McNeil et al. (2015)). The graphs of the daily price series and daily log returns for a subset of the original 32 financial series are presented in Figure 2 and are generally consistent with previous research. The graphs on the left side show that extreme returns in one asset often coincide with extreme returns in other series, as is the case during the 2008 financial crisis and the consequent European sovereign debt crisis in the early 2010s. Visual inspections of the return series plot shown on the right reveals that extreme returns tend to be followed by other extreme returns. This phenomenon is known as volatility clustering and is an indication that our series exhibit non-stationarity in volatility.

From the cross-correlograms in Figure 3 it is evident that the squared returns exhibit significant negative serial dependence and tend to be extremely persistent. Additionally, the Ljung-Box test of randomness performed on the raw return data and the squared returns, respectively, shows strong evidence against the null hypothesis of independent and identically distributed financial returns, Table 2.

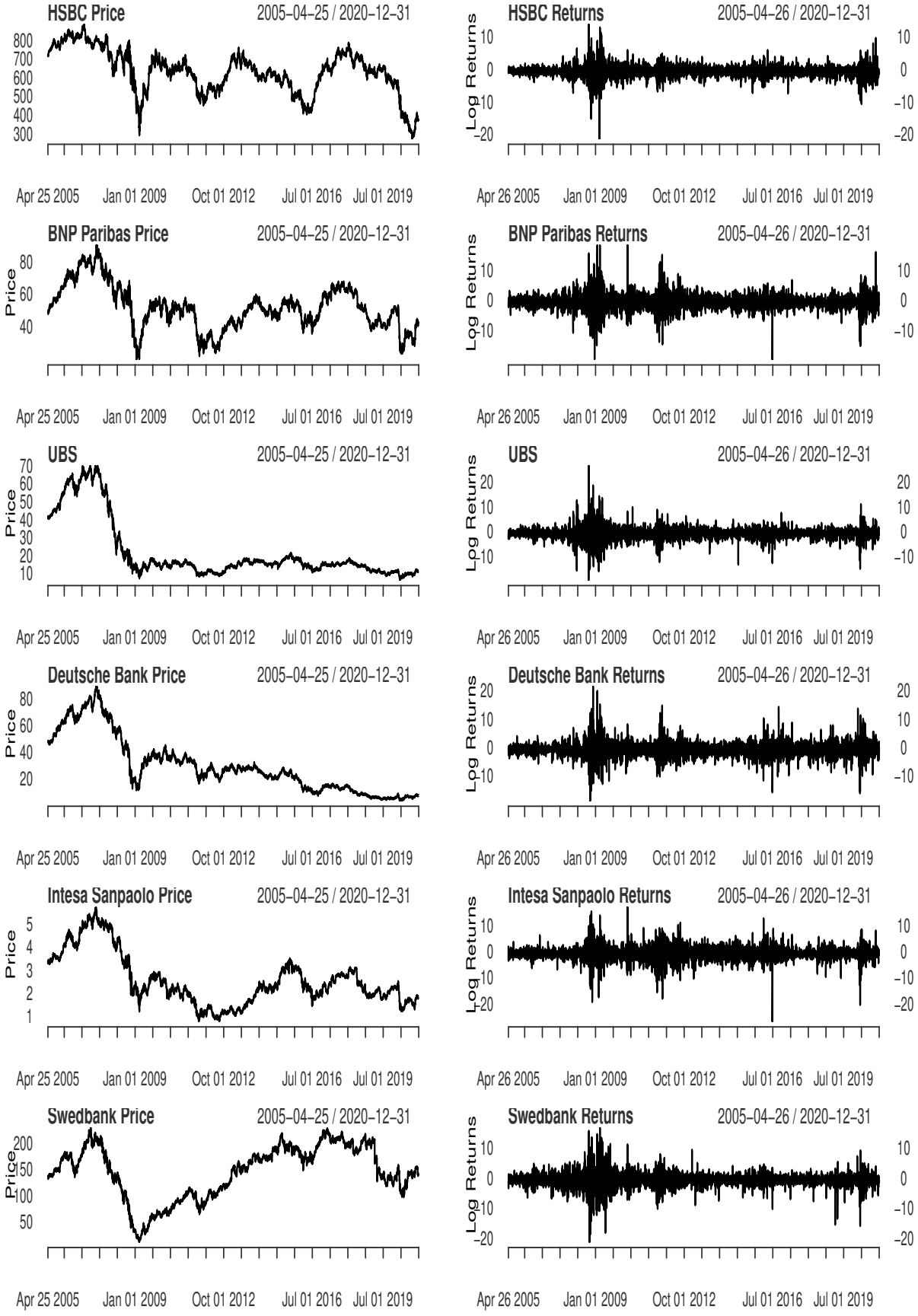


Figure 2: Time-series plot of stock price (left) and returns (right) associated with a subset of the original 32 banks. The data span the period from April 25, 2005 to December 31, 2020.

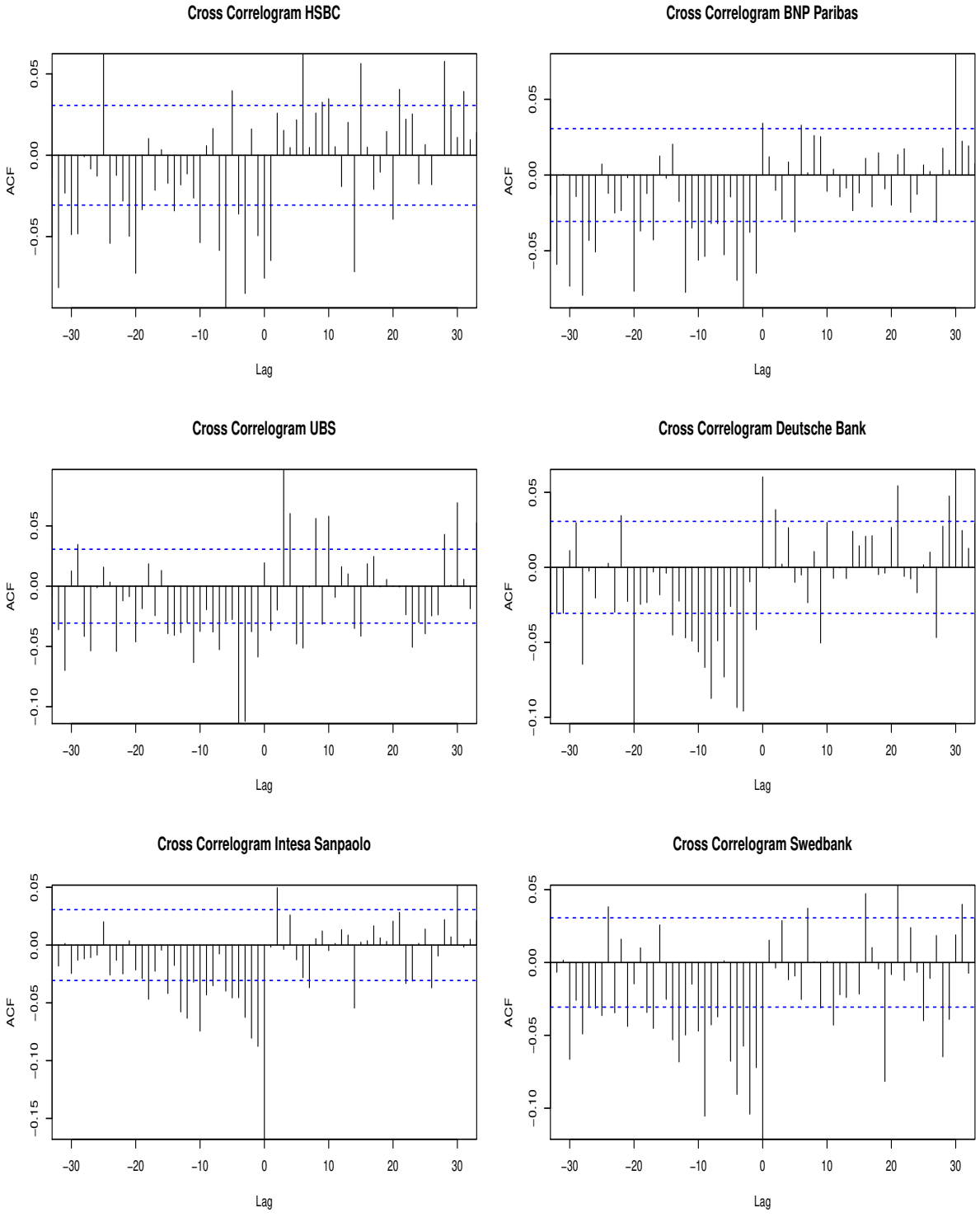


Figure 3: Sample cross correlogram plots of the return series and squared returns associated with a subset of the original 32 banks. The data span the period from April 25, 2005 to December 31,2020.

Distribution of financial stock returns are often observed to diverge from the normal distribution since they tend to exhibit heavier tails and skewness compared to the Gaussian distribution. The Q-Q plot is a frequently used visual tool for comparing the relationship between quantiles of the empirical distribution of the series and theoretical quantiles of a reference distribution (McNeil et al. (2015)). In Figure 4 we show the Q-Q plots of the four return series considered earlier against the normal distribution. If the empirical distributions are linearly related to the Gaussian distribution, we would expect the points in the plots to lie on a line. In our case, however, the inverted S-shape we observe suggests that the normal distribution provides a poor fit to the daily log returns. In addition, we applied a formal test of normality, known as the Jarque-Bera test. For all 32 bank stocks the null hypothesis of normality is rejected (see Table 2).

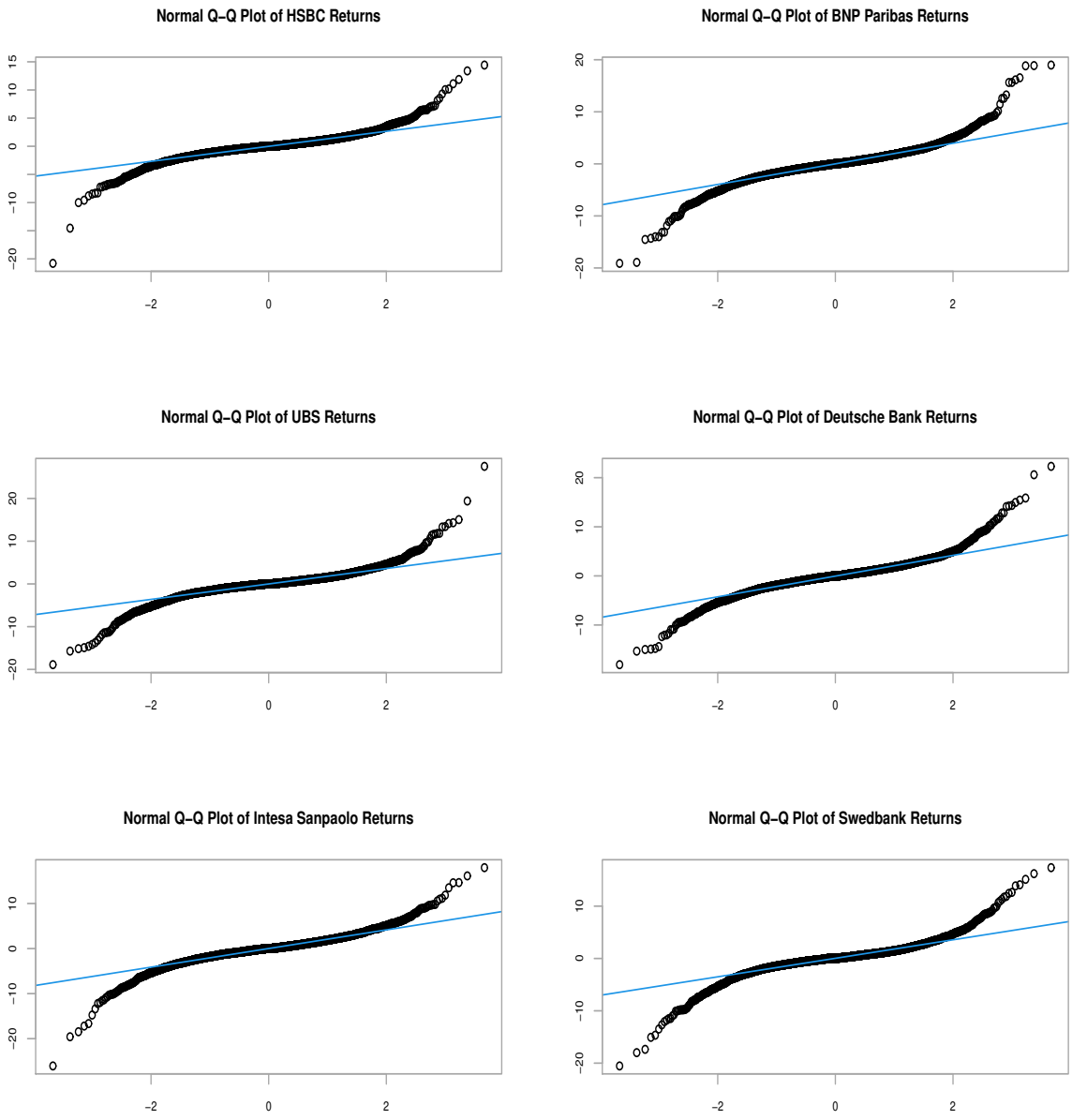


Figure 4: Q-Q plots of daily returns for a subset of the original 32 banks against a normal reference distribution. The data span the period from April 25,2005 to December 31,2020.

Symbol	Country	LBraw	LBsq	p-value (JB)
EBS	Austria	0.00	0.00	0.00
RBI	Austria	0.07	0.00	0.00
KBC	Belgium	0.00	0.00	0.00
DANSKE	Denmark	0.00	0.00	0.00
NDA-Fl	Finland	0.23	0.00	0.00
BNP	France	0.36	0.00	0.00
ACA	France	0.01	0.00	0.00
GLE	France	0.00	0.00	0.00
CBK	Germany	0.01	0.00	0.00
DBK	Germany	0.03	0.00	0.00
ETE	Greece	0.00	0.00	0.00
AIBG	Ireland	0.00	0.00	0.00
BIRG	Ireland	0.00	0.00	0.00
BPM	Italy	0.00	0.00	0.00
ISP	Italy	0.99	0.00	0.00
UCG	Italy	0.14	0.00	0.00
INGA	Netherlands	0.01	0.00	0.00
DNB	Norway	0.13	0.00	0.00
BCP	Portugal	0.00	0.00	0.00
SAB	Spain	0.00	0.00	0.00
SAN	Spain	0.10	0.00	0.00
BBVA	Spain	0.00	0.00	0.00
SEB-A	Sweden	0.05	0.00	0.00
SHB-A	Sweden	0.00	0.00	0.00
SWED-A	Sweden	0.59	0.00	0.00
CSGN	Switzerland	0.00	0.00	0.00
UBSG	Switzerland	0.00	0.00	0.00
BARC	UK	0.00	0.00	0.00
HSBC	UK	0.05	0.00	0.00
LLOY	UK	0.00	0.00	0.00
NWG	UK	0.00	0.01	0.00
STAN	UK	0.30	0.00	0.00

Table 2: Tests of randomness for returns applied to the set of 32 banks for the period from April 25,2005 to December 31,2020. The columns LBraw and LBsq give $p-$values for Ljung-Box tests applied to the raw and squared returns, respectively. The last column shows p-values for Jarque-Bera tests of normality.

4 Methodology

4.1 Financial Time-series Modelling

As we showed in Section 3, bank asset returns exhibit strong temporal dependence and volatility clustering, and their distributions are typically leptokurtic and heavy-tailed. This precludes the direct application of statistical models within EVT that require independent and identically distributed data. In order to obtain approximately iid returns we filter the data set. We follow an approach similar to Hilal et al. (2014) and McNeil and Frey (2000).

As a first step, we convert the log returns from Section 3 to log losses, $\boldsymbol{L}_t = (L_t^{(1)}, \ldots, L_t^{(d)}) = -(S_t^{(1)}, \ldots, S_t^{(d)})$, for $t = 1, \ldots, n$, so that large losses are positioned on the upper tail of the loss distribution. To remove the temporal dependence and volatility clustering of the dataset we consider the ARMA-GARCH family which have proved to be parsimonious but effective tools in practice. In particular, we use ARMA processes to model the dynamics of the conditional mean and GARCH processes for the volatility of the univariate series of log losses (see Bollerslev (1986) for more information on GARCH models). We use the Akaike's information criterion ($AIC = 2 \times \textit{log-likelihood} - 2 \times \textit{number of parameters}$) to select a model which gives a good balance between parsimony and goodness-of-fit. The AIC score suggests that ARMA(1,0)-GARCH(1,1) is the most appropriate specification for a large proportion of the univariate series (see Table 4 in the Appendix). Specifically, we model the dynamics of the log-losses of any given bank, $L_t^{(j)}$ for $j \in \{1, \ldots, d\}$ and $t \in \{1, \ldots, n\}$ as follows

$$L_t^{(j)} = \mu_t^{(j)} + \sigma_t^{(j)} \cdot Z_t^{(j)}, \tag{2}$$

$$\mu_t^{(j)} = \phi \cdot L_{t-1}^{(j)}, \tag{3}$$

$$\left(\sigma_t^{(j)}\right)^2 = \alpha_0 + \alpha_1 \cdot \left(\varepsilon_{t-1}^{(j)}\right)^2 + \beta \cdot \left(\sigma_{t-1}^{(j)}\right)^2, \tag{4}$$

$$\varepsilon_{t-1}^{(j)} = L_t^{(j)} - \mu_t^{(j)}, \tag{5}$$

$$Z^{(j)} \sim WN(0, 1). \tag{6}$$

where we impose the constraint $|\phi| < 1$ for stationarity of the conditional mean $\mu_t^{(j)}$. Furthermore, we assume $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta \geq 0$, and $\alpha_1 + \beta < 1$ to ensure stationarity

and positivity of the conditional variance $\sigma_{t-1}^{(j)}$.

The distribution of the innovations $Z^{(j)}$ is often assumed to be standard normal and the GARCH model parameters are estimated through maximum likelihood. However, it has been shown that the assumption of normality tends to underestimate the heaviness of the tails of the innovations (McNeil and Frey (2000)). Another common approach is to assume that the innovations come from a Student's t-distribution scaled to have unit variance. In this paper, we adopt a third alternative which makes minimal assumptions about the distribution of the innovations. Essentially, we assume that our model specification fits the dynamics of the data but the innovations are misspecified as standard normal and estimate the model parameters through quasi-maximum likelihood (QML) (McNeil et al. (2015)).

The corresponding standardized filtered losses are calculated from the residuals as

$$X_t^{(j)} = \frac{L_t^{(j)} - \hat{\mu}_t^{(j)}}{\hat{\sigma}_t^{(j)}}. \tag{7}$$

If our model is correctly specified $X_t^{(j)}$ should be approximately i.i.d. We can now move to modelling the tail behaviour of the standardized residuals.

4.2 Multivariate Extreme Value Theory for Financial Returns

During times of financial distress, strong interlinkages between banks can lead to spread of losses across institutions and amplify the risk of market-wide crashes. In order to focus exclusively on modelling the extremal dependence structure we estimate the marginal distributions $\widehat{H}^{(j)}$ of $X^{(j)}$ non-parametrically to normalize our data to standard Pareto distributions. This is tantamount to the transformation $1/[1 - \widehat{H}^{(j)}(X^{(j)})]$. In the remainder of Section 4.2, we assume that the random vector of filtered financial losses $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})$, has already been normalized to standard Pareto marginals $F^{(j)}$, $j \in \{1, \ldots, d\}$.

4.2.1 Modelling of the extremal dependence

Under the assumption of multivariate regular variation for $\boldsymbol{X}$ the convergence of the properly normalized tail probabilities is equal to the stable tail dependence function (cf., Resnick (2008))

$$\Lambda(\boldsymbol{x}) = \lim_{u\to\infty} u[1 - \mathbb{P}(\boldsymbol{X} \leq u\,\boldsymbol{x})], \qquad \boldsymbol{x} \in \mathcal{E}, \tag{8}$$

where Λ denotes the exponent measure defined on the cone $\mathcal{E} = [0,\infty)^d \backslash \{\boldsymbol{0}\}$. Accordingly, the high threshold exceedances of $\boldsymbol{X}$ converge to a multivariate Pareto distribution (cf., Rootzén and Tajvidi (2006) and Rootzén et al. (2018))

$$\mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}) = \lim_{u\to\infty} \mathbb{P}(\mathbf{X}/u \leq \boldsymbol{x} \mid \|\boldsymbol{X}\|_\infty > u) = \frac{\Lambda(\boldsymbol{x} \wedge \boldsymbol{1}) - \Lambda(\boldsymbol{z})}{\Lambda(\boldsymbol{1})}, \qquad \boldsymbol{x} \in \mathcal{E}, \tag{9}$$

and concentrate on $\mathcal{L} = \{\boldsymbol{x} \in \mathcal{E} : \|\boldsymbol{x}\|_\infty > 1\}$. A necessary condition for the threshold exceedances to asymptotically have a multivariate GP distribution is the maxima of the normalized random vector $\boldsymbol{X}$ to be asymptotically extreme value distributed (cf. Rootzén and Tajvidi (2006), that is

$$\lim_{n\to\infty} \mathbb{P}\left\{ \max_{t=1,\dots,n} X_t^{(1)} \leq nz_1, ..., \max_{t=1,\dots,n} X_t^{(d)} \leq nz_d \right\} = \mathbb{P}(\boldsymbol{Z} \leq \boldsymbol{z}), \qquad \text{for any } \boldsymbol{z} = (z_1, ..., z_d), \tag{10}$$

where $\boldsymbol{Z}$ is the max-stable distribution arising as the limit of the standardized random vector $\boldsymbol{X}$. Since $\boldsymbol{X}$ has standard Pareto marginals, the marginal components of $\boldsymbol{Z}$ follow a standard Fréchet distribution, $\mathbb{P}(Z_i \leq z) = \exp(-1/z), z \geq 0$, and its distribution is given by

$$\mathbb{P}(\boldsymbol{Z} \leq \boldsymbol{z}) = \exp\left[-\boldsymbol{\Lambda}(\boldsymbol{z})\right], \qquad \boldsymbol{z} \in \mathcal{E}. \tag{11}$$

The exponent measure Λ possesses a number of convenient properties. First, it encodes all available information on the extremal dependence of the normalised random vector $\boldsymbol{X}$. Second, as Engelke and Hitz (2020) show, a direct consequence of the max-stability and the standard Fréchet marginals of $\boldsymbol{Z}$ is that the Radon-Nikodym derivative of the exponent measure Λ, denoted by λ is homogeneous of order $-(d+1)$ and has normalised marginals, and, therefore, is a valid density which defines a max-stable distribution. Essentially, under the assumption that the density $f_{\boldsymbol{Y}}(\boldsymbol{y})$ of $\boldsymbol{Y}$ is positive and continuous it

is proportional to λ, and thus homogeneous of order $-(d+1)$ with

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\lambda(\boldsymbol{y})}{\Lambda(\mathbf{1})}, \qquad \boldsymbol{y} \in \mathcal{L}, \tag{12}$$

since the d-variate extremal coefficient $\Lambda(\mathbf{1}) \in [1, d]$ is a known constant.

4.2.2 Hüsler-Reiss Pareto Distributions

From a practical perspective a valid statistical model in the class of multivariate generalized Pareto distributions needs to find the balance between flexibility and tractability. Some of the classical multivariate statistical models which have found applications in different domains such as hydrology, climate science, and finance include the Dirichlet mixture model (Boldi and Davison (2007), the d-dimensional extremal logistic distribution, and the class of Hüsler-Reiss distributions of Hüsler and Reiss (1989). In this paper we focus on the latter due to its high degree of flexibility and, as we later show, the suitable properties it has which enable it to learn extremal graphical structures in financial applications.

Consider the realization of a Poisson point process $\{U_l\}_{l\in\mathbb{N}}$ on $\mathbb{R}^+$ and independent copies $\{\boldsymbol{W}_l\}_{l\in\mathbb{N}}$ of a centered d-dimensional Gaussian distribution $\boldsymbol{W}$ with covariance matrix denoted as Σ. We define the max-stable Hüsler-Reiss distribution as

$$\boldsymbol{Z} = \max_{l\in\mathbb{N}} U_l \, \exp\{\boldsymbol{W}_l - \mathrm{diag}(\Sigma)/2\}, \tag{13}$$

where its distribution is parameterized by the symmetric and conditionally negative definite variogram matrix Γ associated with $\boldsymbol{W}$

$$\Gamma = \mathbf{1}\mathrm{diag}(\Sigma)^T + \mathrm{diag}(\Sigma)\mathbf{1}^T - 2\Sigma, \tag{14}$$

see Kabluchko et al. (2009). The Hüsler-Reiss multivariate Pareto distribution $\boldsymbol{Y}$ associated with $\boldsymbol{Z}$ is similarly parameterized by the variogram matrix $\Gamma = (\Gamma_{ij})_{1\leq i,j\leq d}$.

For any $k \in \{1, \ldots, d\}$, Engelke et al. (2015) show that the density of its exponent

measure can be written as

$$\lambda(\boldsymbol{y}) = y_k^{-2} \prod_{i \neq k} y_i^{-1} \phi_{d-1}(\tilde{\boldsymbol{y}}_{\setminus k}; \Sigma^{(k)}), \qquad \boldsymbol{y} \in \mathcal{E}, \tag{15}$$

$$\tilde{\boldsymbol{y}}_{\setminus k} = \{\log(y_i/y_k) + \Gamma_{ik}/2\}, \tag{16}$$

where $\phi_l(\cdot; \Sigma)$ denotes the density of the zero-mean d-variate normal distribution with covariance matrix Σ and

$$\Sigma^{(k)} = \frac{1}{2}\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\}_{i,j \neq k} \in \mathbb{R}^{(d-1)\times(d-1)}. \tag{17}$$

The superscript in $\Sigma^{(k)}$ denotes the strictly positive definite covariance matrix Σ with the k-th component omitted. A convenient property of Hüsler-Reiss distributions is their stability under taking marginals, in the sense that if a random vector $\boldsymbol{Z}_V$, $V = \{1, \ldots, d\}$ is Hüsler-Reiss distributed, then for $I \subset V$ and $k \in I$ the lower dimensional marginals are also Hüsler-Reiss distributed with a parameter matrix induced by the respective entries of Λ

$$\lambda_I(\boldsymbol{y}_I) = \int_{[0,\infty)^{d-|I|}} \lambda_I(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}_{\setminus I} = y_k^{-2} \prod_{i \in I \setminus \{k\}} y_i^{-1} \phi_{|I|-1}\big\{\tilde{\boldsymbol{y}}_{I\setminus\{(k)\}}; \Sigma_I^k\big\}, \tag{18}$$

where we have kept the notation from the previous section.

4.2.3 Threshold Selection

An important question that arises in practical applications of threshold exceedances models, both in the univariate and multivariate case, is the following: "How can we select the threshold u such that the limiting distribution of the normalized exceedances is well-approximated by a generalized Pareto distribution?". The issue at hand is twofold. On the one hand, the univariate densities of $X^{(j)} - u \leq x \,|\, X^{(j)} > u$ should be adequately approximated by univariate GP distributions. This implies the apparent trade-off between bias and variation. If the threshold is set too high and this results in too few exceedances being captured, the uncertainty in our estimates is likely to be large and would limit our ability to draw statistically justified conclusions. Conversely, choosing a threshold that is too low might result in exceedances that do not follow a GP distribution, unless the raw data is GP distributed. On the other hand, the dependence structure

$\mathbf{X}/u \leq \boldsymbol{x} \,|\, \|\boldsymbol{X}\|_\infty > u$ should also be well-approximated by a multivariate Pareto distribution.

Threshold selection in the univariate setting has been well-studied and various approaches, most using graphical diagnostics, have been developed such as Zipf plots, threshold stability plots, and mean excess plots (see Scarrott and MacDonald (2012) for a recent review). Threshold selection for multivariate GP models, however, has just recently started to attract more attention and research is scarce. A method, recently introduced by Kiriliouk et al. (2019) and exploited by Engelke and Volgushev (2020) in their definition of non-parametric minimum spanning trees for threshold exceedances, uses the coefficient of upper tail dependence. For any non-empty set $S \in V = \{1, \ldots, d\}$, it is defined as

$$\chi_S = \lim_{u \to 1} \chi_S(q) = \lim_{u \to 1} \frac{\mathbb{P}\left[\bigcap_{i \in S} X^{(i)} > \frac{1}{1-u}\right]}{1-u}, \quad u \in (0,1), \tag{19}$$

whenever the limit exists, and where $\boldsymbol{X} = \{X^{(1)}, \ldots, X^{(d)}\}$ is the random vector of filtered financial losses normalized to standard Pareto marginals and u is the quantile of the marginal Pareto distribution. The empirical estimator for the tail dependence coefficient for a subset S is given by

$$\widehat{\chi}_S(u) = \frac{\sum_{t=1}^{n} \mathbb{1}\{X^{(i)} > \frac{1}{1-u}, \ldots, X^{(j)} > \frac{1}{1-u}\}}{n(1-u)}, \quad u \in (0,1), \quad i, j \in S \tag{20}$$

Consequently, a threshold can be determined by inspecting the empirical estimates of $\chi_S(u)$ for different randomly chosen subsets S against a range of threshold values close to one. It is important to note, however, that the empirical estimator of extremal dependence has a serious drawback. For high thresholds u and for high-dimensional subsets S, the estimates will be unreliable and will suffer from the curse of dimensionality, i.e. $\widehat{\chi}_S(u) \approx 0$, for u close to 1.

4.3 Graphical Extremal Models for Threshold Exceedances

A major limitation of extreme value models in high dimensions and the Hüsler-Reiss model class, in particular, is the rapid increase in the complexity of the possible dependence structures due to the high number of free parameters that need to be estimated, e.g., the

Hüsler-Reiss model requires the estimation of $(d-1)d/2$ parameters.

Graphical models are classical tools allowing the construction of possibly sparse and parsimonious models in high dimensions, and have the advantage of being interpretable in terms of the underlying graph. Naturally, the probabilistic graphical model random vector $\boldsymbol{Y}$ taking values in the Cartesian product $\mathcal{Y} = \times \mathcal{Y}_i$ with $\mathcal{Y}_i \subset \mathbb{R}$, is represented by an undirected graph $G = (V, E)$ with a set of nodes $V = \{1, \ldots, d\}$ and a set of edges $E \subset V \times V$. It is defined by the set of conditional independence constraints of $\boldsymbol{Y}$, where for disjoint subsets $A, B, C \subset V$, $\boldsymbol{Y}_A$ is said to be conditionally independent of $\boldsymbol{Y}_C$ given $\boldsymbol{Y}_B$, $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_C \,|\, \boldsymbol{Y}_B$, if B separates A from C. If, in addition, $\boldsymbol{Y}$ has a positive and continuous Lebesgue density $f_{\boldsymbol{Y}}$ on $\mathcal{Y}$, it follows from the Hammersley-Clifford theorem that the density factorizes for as

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_{C \in \mathcal{C}} \varphi_C(\boldsymbol{y}_C), \qquad \boldsymbol{y} \in \mathcal{Y} \tag{21}$$

where $\mathcal{C}$ denotes the set of all cliques and φ_C are suitably defined functions on $\times_{i \in C} \mathcal{Y}_i$. For decomposable graphs, that is, graphs in which each minimal separator is a clique, the factorization can be expressed in terms of marginal densities

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\prod_{C \in \mathcal{C}} f_C(\boldsymbol{y}_C)}{\prod_{D \in \mathcal{D}} f_D(\boldsymbol{y}_D)}, \qquad \boldsymbol{y} \in \mathcal{Y}, \tag{22}$$

and $\mathcal{D}$ is a multiset containing all separator sets. We refer the interested reader to Lauritzen (1996) for basic notions and in-depth treatment of undirected graphs.

In the remainder of Section 4 we return to the standardized setting of Section 4.2.1 where the threshold exceedances of $\boldsymbol{X}$ converge to a multivariate Pareto distribution. The notion of conditional independence is not directly applicable to threshold exceedances, since the support $\mathcal{L} = \{\boldsymbol{x} \in \mathcal{E} : \|\boldsymbol{x}\|_\infty > 1\}$ of a multivariate Pareto distribution $\boldsymbol{X}$ is not a product space. Engelke and Hitz (2020) show, nonetheless, that there is an alternative definition of extremal conditional independence for $\boldsymbol{X}$. For $k \in V$, they introduce the random vector $\boldsymbol{X}^k$ as $\boldsymbol{X} \,|\, X_k > 1$, which clearly is supported on the product space $\mathcal{L}^k = \{\boldsymbol{x} \in \mathcal{L} : x_k > 1\}$. Thus, for non-empty disjoint subsets $A, B, C \subset V$ we say

that, given $\boldsymbol{X}_B$, $\boldsymbol{Y}_A$ is conditionally independent of $\boldsymbol{X}_C$ if

$$\forall k \in \{1, \dots, d\}: \quad \boldsymbol{X}_A^k \perp\!\!\!\perp \boldsymbol{X}_C^k \,|\, \boldsymbol{X}_B^k, \tag{23}$$

and denote it by $\boldsymbol{X}_A \perp_e \boldsymbol{X}_C \,|\, \boldsymbol{X}_B$. Furthermore, for a decomposable graph $G = (V, E)$ there exists a natural extension of the Hammersley-Clifford theorem which allows the factorization of the density $f_{\boldsymbol{X}}$ of $\boldsymbol{X}$, given that it is positive and continuous on $\mathcal{L}$, as

$$f_{\boldsymbol{X}}(\boldsymbol{y}) = \frac{\lambda(\boldsymbol{y})}{\Lambda(\mathbf{1})} = \frac{1}{\Lambda(\mathbf{1})} \frac{\prod_{C \in \mathcal{C}} \lambda_C(\boldsymbol{y}_C)}{\prod_{D \in \mathcal{D}} \lambda_D(\boldsymbol{y}_D)}, \qquad \boldsymbol{y} \in \mathcal{L}, \tag{24}$$

where $\mathcal{C}$ and $\mathcal{D}$ are the sets of cliques and separators, and the factors λ_I are the marginals of the exponent measure density corresponding to $\boldsymbol{X}_I$ for any $I \subset V$.

The theory of conditional extremal independence and extreme graphical models developed by Engelke and Hitz (2020) is somewhat generic, in the sense that it allows us to use different parametric families or even non-parametric methods. Here we focus on the former and note that the stability of Hüsler-Reiss distributions under taking marginals makes them natural candidates for the further parameterization of graphical models and will allow us to define possibly sparse and tractable graphical structures in higher dimensions.

4.3.1 Construction of Hüsler-Reiss Block Graphs

Engelke and Hitz (2020) show that the variogram matrix Γ associated with the class of Hüsler-Reiss distributions plays a similar role as the the precision matrix for Gaussian distributions - it allows us to identify the conditional independencies between components of Γ. There are, however, some important differences.

The construction of Hüsler-Reiss models defined in Section 4.2.2, implies that there might exist multiple, possibly degenerate Gaussian distributions $\boldsymbol{W}$ with variogram Γ which lead to the same max-stable Hüsler-Reiss distribution. Thus, we cannot simply transform the covariance matrix Σ to the conditional independence structure of the Hüsler-Reiss Pareto distribution to which it corresponds. There exists, however, a valid set of covariance matrices $\tilde{\Sigma}^{(k)} \in \mathbb{R}^{d \times d}$, $k \in V$, which coincide with $\Sigma^{(k)}$ for $i, j \neq k$ and zeros in

place of the k-th row and column

$$\tilde{\Sigma}^{(k)} = \frac{1}{2}\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\}_{i,j\in V} \in \mathbb{R}^{d\times d}. \tag{25}$$

Consequently, for a random vector $\boldsymbol{X}^k$, $k \in V$, defined as $\boldsymbol{X}|X_k > 1$, it can be seen that

$$\boldsymbol{X}^k \overset{d}{=} P\exp\big\{\boldsymbol{W}^k - \mathrm{diag}(\tilde{\Sigma}^{(k)})/2\big\}, \tag{26}$$

where P is a standard Pareto distributed random variable, independent of the centered normal distribution $\boldsymbol{W}^k$ with covariance matrix $\tilde{\Sigma}^{(k)}$. Combined with the definition of conditional independence in (23), this suggests that the precision matrix of $\Sigma^{(k)}$ (keeping notation from Section 4.2.2) denoted by $\Omega^{(k)} = (\Sigma^{(k)})^{-1}$ contains all relevant information on the conditional independence structure of $\boldsymbol{X}$. Thus, for a Hüsler-Reiss Pareto distribution $\boldsymbol{X}$ with a set of nodes $V = \{1, \dots, d\}$

$$X_i \perp_e X_j \mid \boldsymbol{X}_{V\setminus\{i,j\}} \quad \Longleftrightarrow \quad \begin{cases} \Omega^{(k)}_{ij} = 0 & if\ i,j \neq k, \\ \sum_{l\neq k} \Omega^{(k)}_{lj} = 0 & if\ i = k, j \neq k, \\ \sum_{l\neq k} \Omega^{(k)}_{il} = 0 & if\ j = k, i \neq k. \end{cases} \tag{27}$$

where $i, j \in V$ with $i \neq j$, and $k \in V$. We say that $\Omega^{(k)}$ contains the graphical structure of $\boldsymbol{X}$.

The nature of Γ of a Hüsler-Reiss distribution suggests that unless conditional independences are explicitly imposed the resulting graph will be complete with $d(d-1)/2$ edges. An important feature of graphical models for multivariate Perto distributions is that they allow us to restrict model complexity and for certain types of graphs, to estimate models parameters on lower-dimensional marginals. Throughout the rest of this paper we consider block graphs - decomposable connected graphs $G = (V, E)$ with clique set $\mathcal{C}$ and singleton separator set $\mathcal{D}$. Let $\boldsymbol{X} = (X^{(i)})_{i\in V}$ be a random vector which follows a multivariate Pareto distribution with density $f_{\boldsymbol{X}}(\boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{L}$ as before. Assuming $f_{\boldsymbol{X}}(\boldsymbol{y})$ is an extremal graphical model in the sense of Equation (23) with respect to the block graph G, Equation (24) implies a natural construction principle for multivariate Hüsler-Reiss Pareto distributions. First, assume all cliques $C \in \mathcal{C}$ are Hüsler-Reiss distributed with valid exponent measure density $\lambda_C(\boldsymbol{y}_C, \Gamma^{(C)})$ and parameterized by a $|C| \times |C|$-dimensional variogram

$\Gamma^{(C)}$. Since all separator sets are assumed to be singleton and $\lambda_D(\boldsymbol{y}_D) = y_D^{-2}$ is a valid homogeneous density, the consistency constraint $\lambda_D(\boldsymbol{y}_D) = \int_{[0,\infty)^{[C\setminus D]}} \lambda_C(\boldsymbol{y}_C, \Gamma^{(C)}) \mathrm{d}\boldsymbol{y}_{C\setminus D}$ is trivially satisfied. Thus, a valid d-variate Hüsler-Reiss Pareto distribution that factorizes according to a known graph G is defined by the product of the lower-dimensional exponent measure densities

$$f_{\boldsymbol{X}}(\boldsymbol{y}) = \frac{1}{\Lambda(\mathbf{1})} \prod_{C\in\mathcal{C}} \frac{\lambda_C(\boldsymbol{y}_C; \Gamma^{(C)})}{\prod_{j\in\mathcal{D}} y_j^{-2}} \prod_{i\in V} y_i^{-2}, \qquad \boldsymbol{y} \in \mathcal{L}. \tag{28}$$

Engelke and Hitz (2020) show that the d-variate Hüsler-Reiss distribution that factorizes with respect to the block graph G can be uniquely determined as a solution to the problem:

$$\begin{aligned} &find\ a\ feasible\ variogram\ matrix\ \Gamma, \\ &subject\ to \begin{cases} \Gamma_{ij} = \Gamma_{ij}^{(C)}, & for\ i,j\ \in C\ and\ all\ C \in \mathcal{C}, \\ \Omega_{ij}^{(k)} = 0, & for\ all\ k \in V, i, j \neq k\ and\ (i,j) \notin E, \end{cases} \end{aligned} \tag{29}$$

where $\Gamma^{(C)}$ denotes the variogram matrix of a random vector $\boldsymbol{X}$ with $|C|$-variate Hüsler-Reiss distributions on each clique $C \in \mathcal{C}$ and $\Omega^{(k)},\ k \in V$ is the precision matrix of $\Sigma^{(k)}$ as before. Importantly, through (28) and (29) we can construct high-dimensional Hüsler-Reiss distributions out if many low-dimensional ones with number of paramters $\frac{1}{2}\sum_{C\in\mathcal{C}} |C|(|C|-1)$ possibly much smaller than the full d-variate Hüsler-Reiss model.

4.4 Estimation of Hüsler-Reiss Block Grpahs

Let $\boldsymbol{X} = (X^{(1)}, \cdots, X^{(d)})$ be the d-dimensional standardized vector of filtered financial losses with standard Pareto marginal distributions. Under this standardized setting, the threshold exceedances of $\boldsymbol{X}$ are multivariate Pareto distributed as in Equation (9). Assume for now that the underlying block graph $G = (V, E)$ with node set $V = \{1, \ldots, d\}$ and edge set $E \subset V \times V$ is known and fixed. Furthermore, assume $\boldsymbol{X}$ follows a Hüsler-Reiss Pareto distribution with variogram matrix Γ and factorizes with respect to the the block graph G as in Equation (28). The likelihood of $\boldsymbol{X}$ is proportional as a function of

Γ to

$$f_{\boldsymbol{X}}(\boldsymbol{y}, \Gamma) \propto \frac{1}{Q_\Gamma} \prod_{C \in \mathcal{C}} \frac{\lambda_C(\boldsymbol{y}_C; \Gamma^{(C)})}{\Lambda_C(\mathbf{1}; \Gamma^{(C)})}, \qquad Q_\Gamma = \frac{\Lambda(\mathbf{1}; \Gamma)}{\prod_{C \in \mathcal{C}} \Lambda_C(\mathbf{1}; \Gamma^(C)} \tag{30}$$

where Q_Γ is a normalising constant which depends on all parameters through $\Lambda(\mathbf{1}; \Gamma)$. Engelke and Hitz (2020) show that for graphs that factorize according to (28), Q_Γ contains limited information on Γ and instead of using the full joint likelihood to estimate the model parameters we can use the separate likelihoods of each clique $C \in \mathcal{C}$

$$f_C(\boldsymbol{y_C}; \Gamma^{(C)}) = \frac{\lambda_C(\boldsymbol{y}_C; \Gamma^{(C)})}{\Lambda_C(\mathbf{1}; \Gamma^{(C)})}, \qquad \boldsymbol{y_C} \in \mathcal{L}_C. \tag{31}$$

In practice, some observations of $\boldsymbol{X_C}$ might not be high enough, i.e. they fall below the specified threshold u, to justify the use of the multivariate Pareto distribution $\boldsymbol{X}$. For such components of $\boldsymbol{X}$ we apply censoring from below to avoid the scenario where small values have an excessively strong effect on the fit. Censored likelihoods allow for a more robust estimation of extreme value distributions (see Smith et al. (1997) and Ledford and Tawn (1997) for more information). For a clique $C \in \mathcal{C}$ and a data point $\boldsymbol{X}_C$ with $\|\boldsymbol{X}_C\|_\infty > u$ let H be the set of indices $h \in C$ such that $x^h < u$, and respectively, $y^h < 1$. The censored likelihood for a clique containing such points can be calculated as

$$f_C^{cens}(\boldsymbol{y}_C; \Gamma^{(C)}) = \int_{[0,1]^J} f_C(\boldsymbol{y}_C; \Gamma^{(C)}) \mathrm{d}\boldsymbol{y}_J, \qquad \boldsymbol{y}_C \in \mathcal{L}_C, \tag{32}$$

and thus we only use the information that such components Y^h are smaller than 1, but not the exact magnitude. Consequently, the censored log-likelihood for each clique C is given by

$$L^{cens}(\Gamma^{(C)}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \sum_{\boldsymbol{y}_t \in \mathcal{L}_C} \log\{f_C^{cens}(\boldsymbol{y}_t^C; \Gamma^{(C)})\} \tag{33}$$

where each y_t^C, $t = \{1, \ldots, d\}$ has its own censoring set $H^{(t)} \subset C$.

In the previous section and up to this point we analyse Hüsler-Reiss Pareto distribution that factorise according to a known and fixed block graph G. In our application to bank losses, however, the graphical structure is unknown and the conditional independence structure should be learned from data. Among the family of connected graphs trees are a natural starting point due to their simplicity and flexibility. A spanning tree $G^{(T)} = (V, E^{(T)})$ is a special case of a block graph with a unique path between any two pairs of

nodes. Therefore, a tree has no cycles, $|E^{(T)}| = |V| - 1$ and all cliques are of size two with singleton separator sets. Due to their relatively simple structure trees are among the most parsimonious graph models and suitable baseline models that allow the construction of more complex sparse graphs. The cardinality of the set of all possible spanning trees is d^{d-2}, however for the class of minimum spanning trees there exist greedy algorithms that allow us to determine the tree structure efficiently (see, Kruskal (1956) or Prim (1957). The minimum spanning tree $G^{(MST)} = (V, E^{(MST)})$ is the tree that minimizes the sum of weights on its edges $(i, j) \in E$ according to

$$G^{(MST)} = \underset{G=(V,E)}{\arg\min} \sum_{(ij)\in E} w_{ij} \tag{34}$$

where the weights w_{ij} denotes the "distance" between nodes i and j. Using an algorithm due to Chow and Liu (1968) we can search for the conditional independence tree that maximizes the likelihood of the graph model. Assume all pairs of nodes follow the same class of Pareto distributions, namely, bivariate Hüsler-Reiss distributions. Within this parametric family, the maximal log-likelihood for any given tree is the sum over the maximized clique log-likelihoods as in Equation (33). The minimum spanning tree problem in Equation (34) thus entails the identification of the tree that maximizes the log-likelihood over all trees and all distributions within the Hüsler-Reiss family with weights defined as

$$w_{ij} = -L^{cens}(\widehat{\Gamma}^{(C)}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) - 2 \sum_{y_t^{(i)}>1} \log y_t^{(i)} - 2 \sum_{y_t^{(j)}>1} \log y_t^{(j)}, \tag{35}$$

where we include the censored marginal densities y_i^2 and y_j^2 as parameters for optimization (see, Engelke and Hitz (2020)).

We can anticipate that the complexity of the financial system is of higher order than the bivariate relationships we are imposing. Thus a minimum spanning tree might not be the true graphical structure to explain the complex dependencies between extreme observations of negative asset returns. We extend the spanning tree constructed in (34) with weights (35) by adding additional edges restricting the graphical structure to block graphs with clique size of at most three. Suppose $G^{(1)} = (V, E_1)$ is the estimated Hüsler-Reiss minimum spanning tree with edge set denoted by E_1. We apply greedy forward selection and in each step we add additional edges $\{i, j\}$, $i, j \in V$ so that $G^{(k+1)} =$

(V, E_{k+1}), with $E_{k+1} = E_k \cup \{i, j\}$, for $k = 1, 2, \ldots,$ is still in the class of block graphs with cliques of size at most three. The clique likelihoods of each graph estimated in this way yield a unique estimate $\widehat{\Gamma}$ of the variogram matrix through Equation (29). We compare different Hüsler-Reiss Graph models by using the Akaike information criterion $AIC = -2L^{cens}(\widehat{\Gamma}^{(C)}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) + 2p$, where p is the number of free parameters in the respective model (Engelke and Hitz (2020)). At each iteration we add the edge $\{i, j\}$ that leads to the biggest decrease in the value of the Akaike information criterion. We continue this procedure until no more edges can be added. The resulting optimal graphical representation for extreme financial events is the one that gives the minimum value of the AIC.

4.5 Measuring the Systemic Importance of Financial Institutions

In times of financial distress, a topic that receives increased public attention is the potential bailing out of particularly large banks by regulators. In such scenario, authorities are inclined to bailout large financial institutions under the argument that the insolvency of such institutions poses a significant risk to the financial system, and to the stability of the economy in general (Zhou (2010)). This suggests that the systemic importance of a bank is strongly related to its size. In order to test the presence of a correlation between size and systemic importance in our graphical model for extremes in the following section we propose two measures on the importance of banks during financial crises.

From network analysis perspective, there are a number of established centrality metrics which measure the importance of a particular node with respect to various summary statistics derived from the characteristics of the underlying network. Naturally, centrality measures have been initially applied to problems in social network theory related to a person's importance, power, or influence (see, Marsden (2005) for an overview). In addition, they have found wide application in many other fields of network theory, such as telecommunication networks and river networks. We pay special attention to closeness centrality and betweenness centrality which in our estimated extremal graphical model can aid us in identifying systemically important financial institutions.

Albeit both centrality measures are well-defined for unweighted graphs, we investigate the implications of the dependence structure captured by the Hüsler-Reiss graphical model estimated using the methodology of the previous sections. Essentially, the stability of the Hüsler-Reiss distribution under taking marginals allows us to weigh each edge $(i, j) \in E$ by the "distance" w_{ij}, given by the estimated bivariate tail correlation $2 - 2\Phi(\sqrt{\hat{\Gamma}_{ij}}/2)$. Consequently, the path between two non-adjacent vertices l and m is given by the sum of edge weights between the two vertices.

Closeness centrality sums the shortest paths of a vertex to every other vertex (Bavelas (1950)), i.e., it indicates how close a vertex is to all other vertices of the network. The closeness of a vertex $j \in V$ is given by

$$CC(j) = \frac{1}{\sum_{i \in V \setminus \{j\}} d_j(i)}, \tag{36}$$

where $d_j(i)$ denotes the shortest-path distance from j to i. Betweenness centrality, on the other hand, measures the fraction of shortest paths that pass through a vertex (see Freeman (1977)). It is given by

$$BC(j) = \sum_{lk \in V \setminus \{j\},\, l \neq k} \frac{\sigma_{lk}(j)}{\sigma_{lk}}, \tag{37}$$

where $\sigma_{lk}(j)$ summarizes the number of shortest paths from vertex l to vertex k passing through j and σ_{lk} denotes the total number of shortest paths from vertex l to k. Essentially, betweenness centrality is a very different metric from closeness centrality: it helps us identify financial institutions that play a "bridge spanning" role in a graph. In other words, a bank with high betweenness centrality indicates that more banks depend on this specific bank to make connections with other institutions. Therefore, we regard such a bank as possessing higher systemic importance during market wide financial crises.

5 Empirical Application

In this section, we determine the optimal block graph that captures the extremal dependence between banks, evaluate the robustness of the estimated graphical structure and discuss the evidence on the systemic importance of banks during financial crises by means of centrality measures. As detailed in Section 3, our dataset consists of daily returns, which we convert to log-losses, on 32 publicly traded European banks over the period from April 25, 2005 to December 31, 2020. The final dataset comprises of $n = 4094$ daily observations per series. Due to the strong temporal dependence and volatility clustering observed in financial time-series data we first filter the univariate series through the use of an ARMA-GARCH process.

Based on the model selection strategy detailed in Section 4.1 we consider the log-losses $\boldsymbol{L}_t = (L_t^{(1)}, \ldots, L_t^{(d)})$ to be realisations from ARMA(1,0)-GARCH(1,1) processes. As evidenced by Figure 5, the filtered losses $X_t^{(j)} = (L_t^{(j)} - \hat{\mu}_t^{(j)})/\hat{\sigma}_t^{(j)}$ obtained from this model show no visual signs of serial correlation in their squared values, in contrast with the results on the raw data in Section 3. We conclude that while there is strong evidence that the raw data is not independent and identically distributed, the standardized residuals can be treated as approximately iid. In addition, the Quantile-quantile plots in Figure 6 of the filtered losses against the normal distribution confirm the assumption of non-Gaussianity in the innovation processes and the presence of fat tails in the distribution of the filtered log-losses.

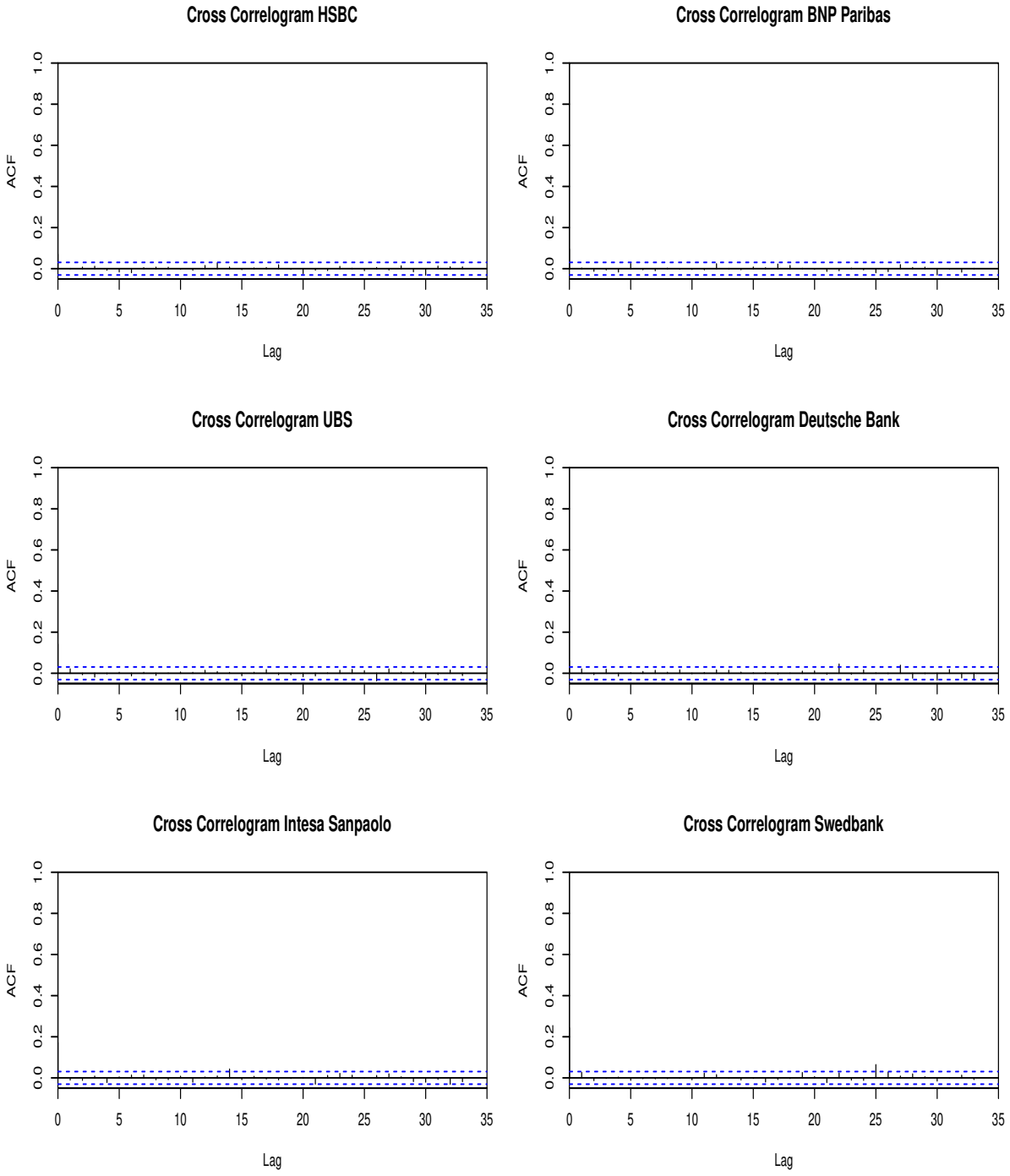


Figure 5: The sample cross correlogram plots of the squared values of the filtered log-losses associated with a subset of the original 32 banks. The data span the period from April 25,2005 to December 31,2020.

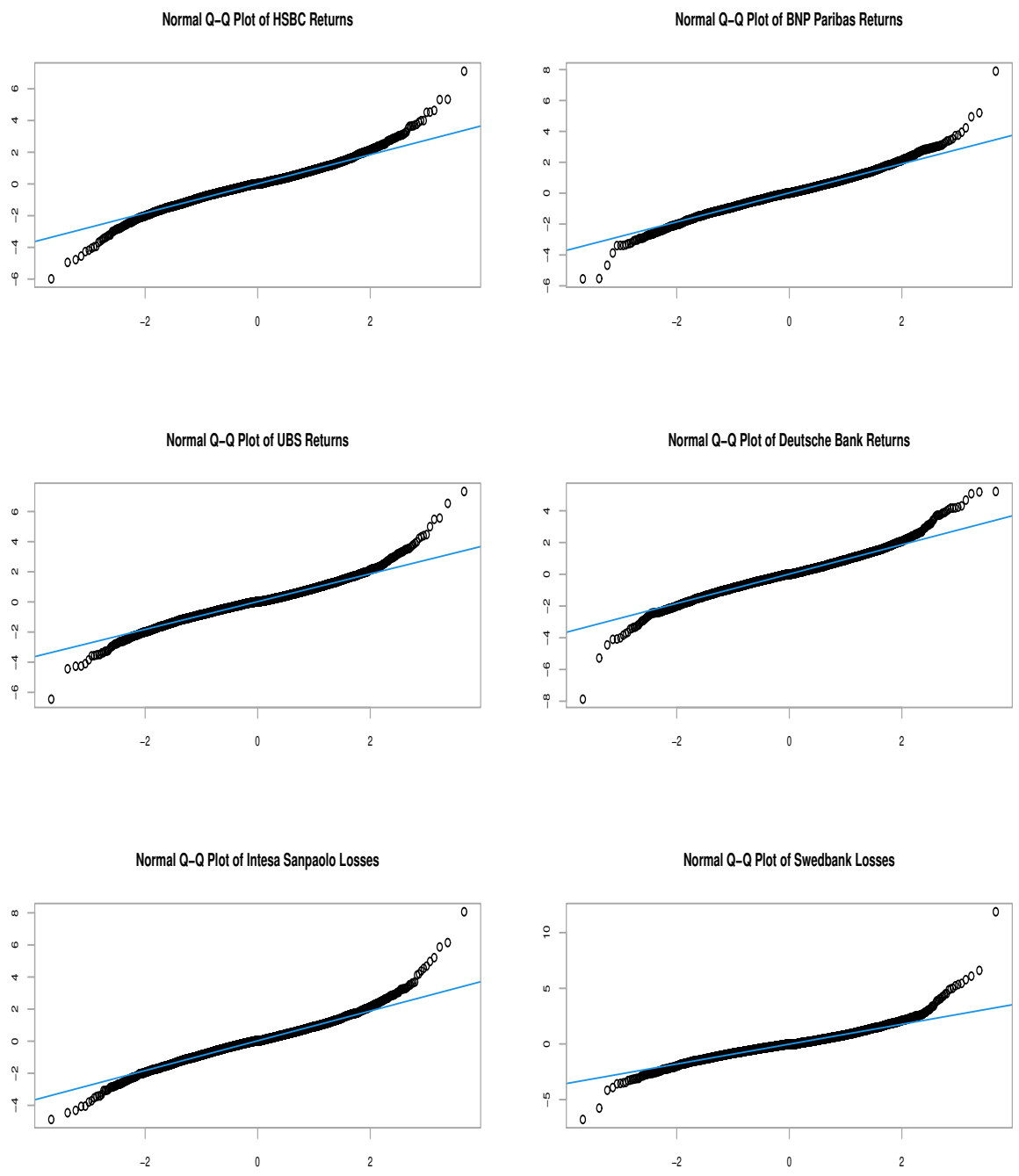


Figure 6: Q-Q plots of the filtered log-losses associated with a subset of the original 32 banks against a normal reference distribution. The data span the period from April 25,2005 to December 31,2020.

5.1 Estimation of the Optimal Graph Structure

We proceed to model the tail dependence of the standardized filtered losses $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})$, $j \in \{1, \ldots, d\}$, using an extremal graphical model. For our multivariate dataset, we first conduct exploratory data analysis to check whether the assumption of extremal tail dependence is satisfied by inspecting the plots of the empirical estimates of the tail correlation coefficients $\hat{\chi}(q)$ for values of q close to 1. Figure 12 in the Appendix displays plots of the empirical tail correlations, $\hat{\chi}_S(q)$ for arbitrary chosen non-empty subsets of banks, S. A common observation across all subsets is that the tail correlation coefficient decreases as q and $|S|$ increase. This is a common artifact found in practice which could be attributed to instability of the estimates for high thresholds and the curse of dimensionality. Nevertheless, $\hat{\chi}_S(q)$ seems to converge to a positive value and there appears to be a stable dependence in the tail between the filtered log-losses from around $q = 0.9$.

Based on the preceding analysis we estimate the extremal dependence in bank returns by the Hüsler-Reiss minimum spanning tree for a threshold q equal to the 0.9 quantile of the marginal Pareto distribution. In our estimations, we use the methods implemented in the R package **graphicalExtremes** (Engelke et al. (2019)). This yields 2411 observations with at least one exceedance. Motivated by the estimation approach detailed in Section 4.2, we estimate the extremal tree structure $\hat{G}^{(MST)} = (V, \hat{E}^{(MST)})$ with weights for each pair of nodes based on the clique log-likelihoods of bivariate Hüsler-Reiss distribution as in Equation 35. The estimated tree is shown in Figure 7. Unsurprisingly, the graph suggests that extreme events observed in financial data tend to be geographically isolated. For instance, extreme observations in stock prices for Scandinavian banks (i.e. banks with ID $\in \{5, 18, 23, 24, 25\}$) are connected to the rest of the graph through the Dutch *ING Groep NV* (ID = 17) and similarly South European banks (i.e. ID $\in \{14, 15, 16, 20, 22\}$) are conditionally independent of other European banks given the stock price value of the largest bank in Spain, and second largest in our dataset, *Banco Santander SA* (ID = 21). A striking feature of our minimum spanning tree is the prominent role played by the largest French bank, *BNP Paribas SA*, as it seems to connect the different geographic and economic regions in Europe with each other.

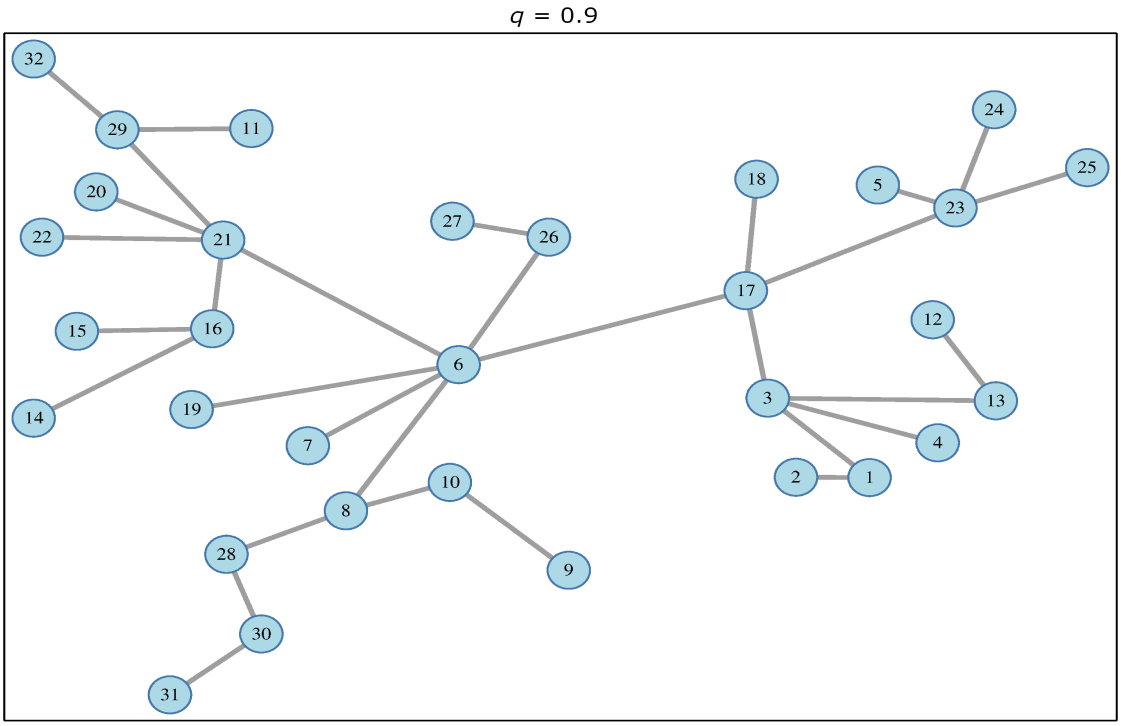


Figure 7: The tree induced by bank losses for our set of 32 banks for quantile $q = 0.9$. Indices can be found in Table 3.

The Hüsler-Reiss minimum spanning tree, however, might not be the true graphical model to capture the complex relationships occurring between banks. Therefore, we extend this model by adding additional edges iteratively while restricting the model to the class of block graphs with cliques of size two and three as described in Section 4.4. Essentially, this entails the maximization of the censored likelihood of bivariate Hüsler-Reiss densities. This results in a sequence of graph models, $\hat{G}^1 \dots \hat{G}^M$. We select the optimal graph representation, $\hat{G}^*$ for extreme events in bank losses as the one minimizing the Akaike information criterion. In particular, we choose the model with 44 edges and an AIC of 112020.0 as indicated by Figure 8. The 13 additional edges of the best graphical model bring a significant improvement over the simpler minimum spanning tree with 31 edges but an AIC of 114758.8. Compared to the minimum spanning tree from Figure 7, the optimal Hüsler-Reiss block graph further accentuates the strong domestic and regional interlinkages between financial institutions suggesting that extreme events in stock prices of geographically and economically "close" banks are inherently linked.

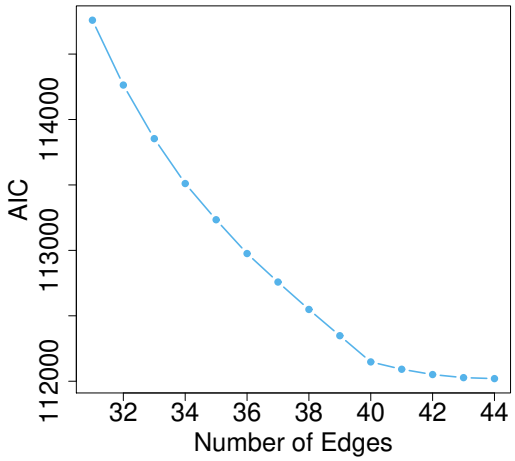


Figure 8: AIC values as a function of the number of edges, starting from the minimum spanning tree.

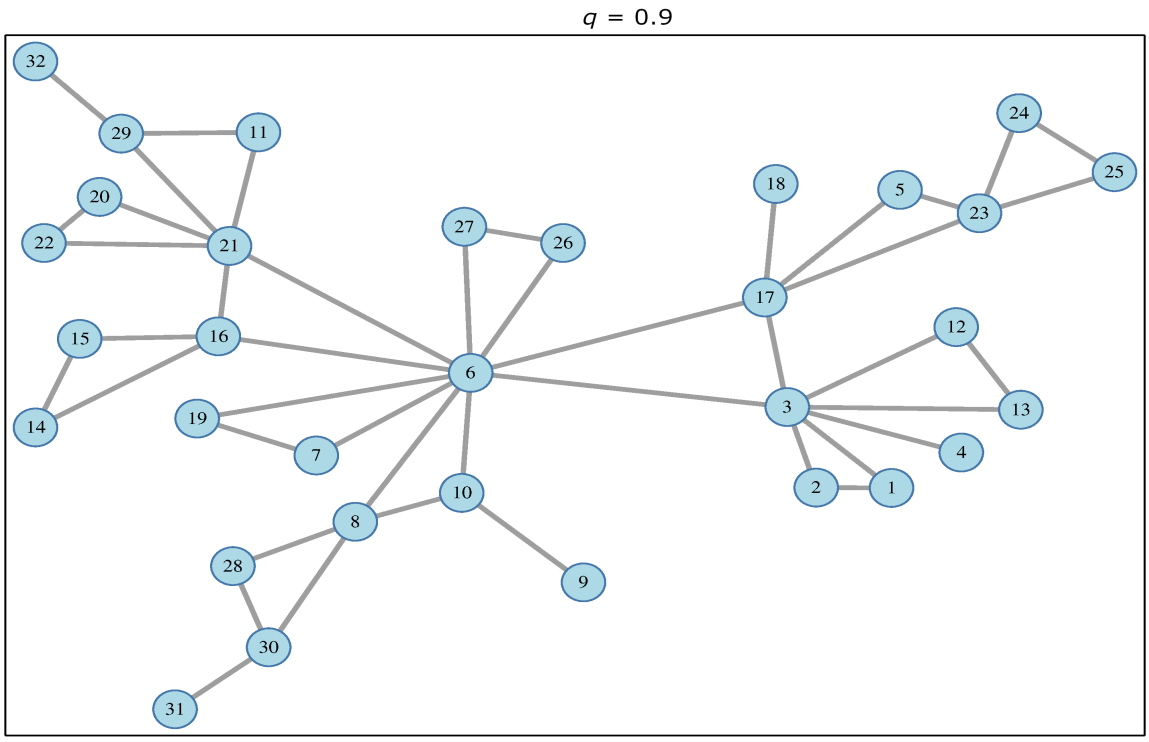
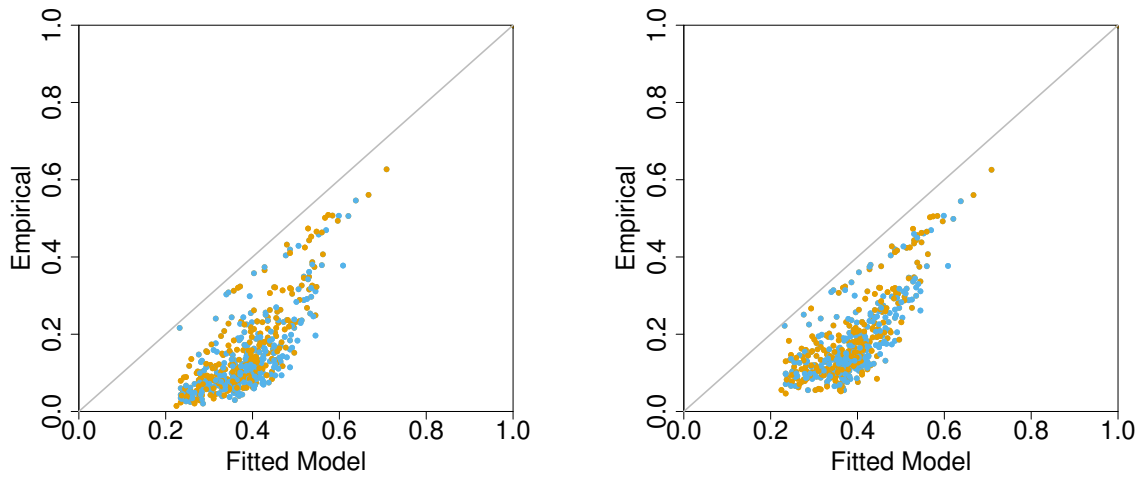


Figure 9: The estimated optimal block graph induced by bank losses for our set of 32 banks for quantile $q = 0.9$. Indices can be found in Table 3.

For any two pair of banks, $i, j \in V$, we calculate the bivariate tail correlation implied by the Hüsler-Reiss graph model as $2-2\Phi(\sqrt{\hat{\Gamma}_{ij}}/2)$, where the bivariate parameter estimates $\hat{\Gamma}_{ij}$ are taken from the estimated variogram matrix $\hat{\Gamma}$ and $\Phi(\cdot)$ is the cumulative density function of the standard Gaussian. In Figure 10 we compare empirical estimates of the

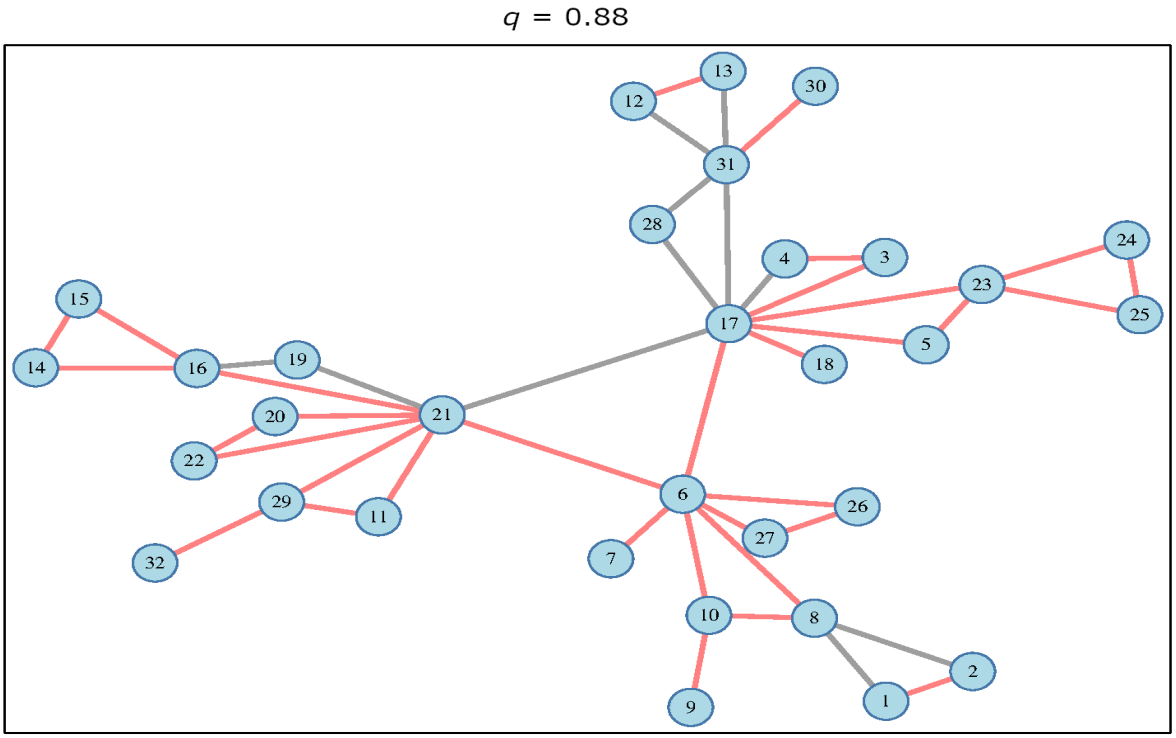
extremal correlation coefficients with those implied by the Hüsler-Reiss minimum spanning tree (left panel) and the optimal block graph (right panel). As expected, the minimum spanning tree does a poor job at capturing the bivariate extremal dependence between banks. Extending the Hüsler-Reiss tree model to block graphs with clique size of at most three brings a noticeable improvement in terms of this bivariate summary statistic. Interestingly, however, our estimated optimal model seems to be moderately biased, too.

To further assess the robustness of the optimal graph structure with respect to our chosen threshold, we estimate the graphs for the 88% and 92%-quantiles as shown in Figure 11. Overall, we see some variation of the optimal graph structure compared to the 90%-quantile graph but the majority of the 44 edges remain stable with 33 identical edges in the 88%-quantile graph and 28 in the 92%-quantile Hüsler-Reiss graph, respectively. A key difference that is clearly visible is the less prominent role played by the French bank *BNP Paribas SA* and the lack of a direct link between *ING Groep NV* and *Banco Santander SA*.

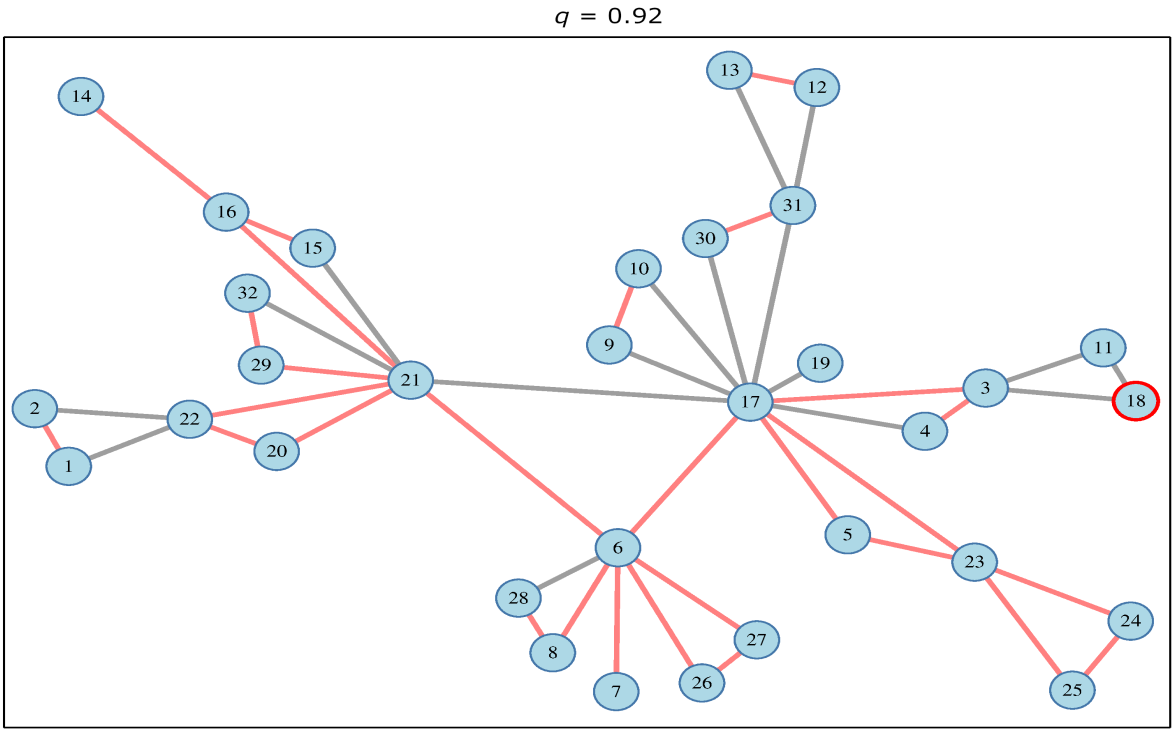


(a) (b)

Figure 10: Empirically estimated extremal dependence coefficients χ against those implied by the fitter minimum spanning tree (left panel) and HR graph mode with the minimum AIC value (right panel).



(a)



(b)

Figure 11: Estimated optimal graphical structures for quantiles $q = 0.88$ and $q = 0.92$. Edges in red represent edges which are identical to the graph with $q = 0.9$. Indices can be found in Table 3.

5.2 Systemic Importance of Financial Institutions

The economic crisis in the European Union during the early 2000s caused by the introduction of the euro in 1999, and the financial crisis in 2008, have shown that modern financial markets are strongly interdependent and interconnected. A key paradigm receiving increased attention in mass media, prudential regulation, and academic research alike, is the investigation on the role played by systemically important financial institutions during recessions. According to this paradigm large shocks to a small number of banks exhibiting a high degree of interconnectedness pose a significant risk to the whole financial system. Furthermore, the collapse of behemoths like Lehman Brothers in the US has brought the argument of "too big to fail", often used in support for bailing out of large companies, under serious scrutiny.

We continue the analysis of the previous section by studying the topology of the network implied by the optimal Hüsler-Reiss graphical model $\hat{G}^*$, from Section 5.1. As argued in Section 4.5, we identify the systemically important banks in our dataset of 32 financial institutions by calculating two classic centrality measures, closeness centrality and betweenness centrality, for each bank in the graph structure. The third column in Table 3 gives the closeness centrality rank order of each bank (raw values can be found in Table 5 in the Appendix). Unsurprisingly, the "root" of the optimal Hüsler-Reiss graph from the previous section, *BNP Paribas SA* , together with three other banks directly connected to it, *ING Groep NV, KBC Group NV, Banco Santander SA*, emerge as the financial institutions with the highest capacity to propagate information to other banks in the network in times of financial crises. An unanticipated addition to this list is the Portuguese bank, *Banco Comercial Portugues SA*, which despite being the smallest bank in terms of market capitalization, has a closeness centrality value which puts it among the most influential banks. This might have a convenient intuitive interpretation. One of the effects of the 2008 recessions on the Portuguese economy and financial system was the country's inability to repay or refinance its government debt. This led to a significant bailout package (EUR 78 billion and EUR 12 billion for bank recapitalization) from the International Monetary Fund and the EU, partly issued through other big EU banks, such as *BNP Paribas, Goldman Sachs International*, and *The Royal Bank of Scotland*, and financed through the European stock markets which led to an increased exposure to

other financial institutions. (cf. Sérgio and de Sousa (2016))

The last column in Table 3 gives the betweenness centrality rank order of each bank in our dataset (again, the raw values can be found in Table 5 in the Appendix). Banks with a rank value equal to 11 indicate that there are no shortest paths passing through that node. The financial institutions with the highest values of betweenness centrality (*BNP Paribas SA, ING Groep NV, KBC Group NV, Banco Santander SA*) is very similar to those which exhbit the highest values of closeness centrality, albeit in a slightly different order. This suggests that the financial institutions with the smallest "distance" to other banks in the network also play a "bridge-spanning" role between other pair of nodes. Thus, these four banks not only have the ability to quickly transmit contagion to all other banks during market-wide financial crises, but can also have an important influence on other instiutions as they can limit or amplify the contagion effects that pass through it. Furthermore, it is difficult to establish a clear linear relationship between size and systemic importance and there seem to be other characteristics which play an important role in determining systemic importance.

Bank ID	Bank	$CC(j)$	$BC(j)$
1	Ereste Group Bank AG	17	11
2	Raiffeisen Bank International AG	15	11
3	KBC Group NV	2	2
4	Danske Bank A/S	14	11
5	Nordea Bank Abp	20	11
6	BNP Paribas SA	1	1
7	Crédit Agricole SA	13	11
8	Societe Generale SA	8	5
9	Commerzbank AG	25	11
10	Deutsche Bank AG	7	8
11	National Bank of Greece	16	11
12	AIB Group PLC	9	11
13	Bank of Ireland Group PLC	12	11
14	Banco BPM SpA	24	11
15	Intesa Sanpaolo SpA	27	11
16	UniCredit SpA	6	6
17	ING Groep NV	3	4
18	DNB ASA	19	11
19	Banco Comercial Portugues SA	5	11
20	Banco de Sabadell SA	22	11
21	Banco Santander SA	4	3
22	BBVA SA	28	11
23	Skandinaviska Enskilda Banken AB	18	6
24	Svenska Handelsbanken AB	29	11
25	Swedbank AB	31	11
26	Credit Suisse Group AG	11	11
27	UBS Group AG	10	11
28	Barclays PLC	26	11
29	HSBC Holdings PLC	21	8
30	Lloyds Banking Group PLC	23	8
31	Natwest Group PLC	32	11
32	Standard Chartered PLC	30	11

Table 3: Rank order of the closeness centrality (CC) and betweenness centrality (BC) values for each bank (indices can be found in Table 3).

6 Conclusion

This paper investigates the complex cross-border and domestic interlinkages accross financial institutions in times of financial recessions. We use a novel dataset comprising 15 years of stock returns on 32 of the largest publicly traded banks across 16 countries in the European Economic Area. We show that by extending the influential approach of Engelke and Hitz (2020), who introduce the notion of extremal conditional independence for multivariate Pareto distributions,l, we are able to construct sparse graphical models that allow us to interpret and measure the systemic importance of financial institutions.

We first remove the temporal dependence and volatility clustering commonly observed in financial time-series data by modelling the data as realisations from ARMA-GARCH processes. We normalize our data to standard Pareto distributions and focus exclusively on modelling the extremal dependence structure in our data. Second, the flexibility and the stability of the Hüsler-Reiss distribution under taking marginals, coupled with the notion of conditional independence for extreme events allows us to learn from our data the underlying graphical structure for extreme events that drive systemic risks in the banking sector. Lastly, by analysing the topology of the underlying optimal graph we identify systemically important banks.

Our results show that financial institutions under the optimal Hüsler-Reiss block graph exhibit strong domestic and regional interlinkages suggesting that extreme events in the log-losses of geographically and economically "close" banks are inherently linked. In addition, the two network centrality measures that we propose, closeness centrality and betweenness centrality, suggest that there is a small subset of banks which play a "bridge-spanning" role and have a significant influence over other banks in the network.

Nevertheless, our research has one major limitation. Even-though the optimal Hüsler-Reiss block graph brings a significant improvement over the simpler minimum spanning tree, the bivariate tail correlation summary statistic suggests that our more complex model structure might be too simplistic to fully capture the complex relationships that occur between financial institutions during financial crises. A possible solution, which could remedy this drawback and allow the estimation of denser Hüsler-Reiss graphs is the estimation of graphs based on Lasso regularization.

References

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. New York: Wiley.

Boldi, M. O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(2):217–229.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467.

Davison, A. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2(1):203–235.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40(4):317–337.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Engelke, S., Hitz, S. A., and Gnecco, N. (2019). graphicalextremes: Statistical methodology for graphical extreme value models. R package version 0.1.0.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8(1):241–270.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265.

Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. In *Paper presented at the 11th International Conference on Extreme Value Analysis*.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.

Hilal, S., Poon, S.-H., and Tawn, J. (2014). Portfolio risk assessment using multivariate extreme value methods. *Extremes*, 17(4):531–556.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics*, 61(1):123–135.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50.

Lauritzen, S. (1996). *Graphical Models*. Oxford, UK: Oxford University Press.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):475–499.

Marsden, P. V. (2005). Network analysis. In Kempf-Leonard, K., editor, *Encyclopedia of Social Measurement*, pages 819–825. Elsevier, New York.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3):271–300. Special issue on Risk Management.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401.

Resnick, S. (2008). *Extreme Values, Regular Variation and Point Processes*. New York: Springer.

Rootzén, H., Segers, J., and L. Wadsworth, J. (2018). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized pareto distributions. *Bernoulli*, 12(5):917–930.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat Statistical Journal*, 10:33–60.

Sérgio, A. and de Sousa, A. R. (2016). *The Impact of the Financial Crisis on Portuguese Banks: The Problem of Portuguese Sovereign Debt*, pages 9–20. Palgrave Macmillan UK, London.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Smith, R. L., Tawn, J. A., and Yuen, H. K. (1990). Statistics of multivariate extremes. *International Statistical Review / Revue Internationale de Statistique*, 58(1):47–58.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Zhou, C. (2010). Are banks too big to fail? measuring systemic importance of financial institutions. *International Journal of Central Banking*, 23.

A Supplementary Tables and Figures

Symbol	with intercept			w/o intercept		
	(1,0)-(1,1)	(1,1)-(1,1)	(1,1)-(2,1)	(1,0)-(1,1)	(1,1)-(1,1)	(1,1)-(2,1)
EBS	**4.4413**	4.4416	4.4422	4.4418	4.4421	4.4427
RBI	4.6963	4.6968	4.6974	**4.6962**	4.6966	4.6972
KBC	**4.3751**	4.3758	4.3766	4.3755	4.3760	4.3767
DANSKE	**3.9260**	3.9262	3.9270	3.9264	3.9267	3.9274
NDA-Fl	**3.9440**	3.9444	3.9450	3.9442	3.9449	3.9455
ACA	**4.3820**	4.3824	4.3831	4.3821	4.3825	4.3832
BNP	4.1893	4.1898	4.1904	**4.1891**	4.1896	4.1902
GLE	4.4319	4.4319	4.4326	**4.4316**	**4.4316**	4.4323
CBK	4.6602	4.6607	4.6613	**4.6598**	4.6602	4.6609
DBK	4.3408	4.3413	4.3417	**4.3404**	4.3408	4.3412
ETE	5.4903	5.4905	5.4914	**5.4901**	5.4903	5.4912
AIBG	5.3506	5.3503	5.3513	5.3504	**5.3501**	5.3511
BIRG	5.0217	5.0198	5.0208	5.0214	**5.0195**	5.0205
BPM	4.7855	4.7860	4.7867	**4.7851**	4.7855	4.7863
ISP	4.2878	**4.2876**	4.2884	4.2881	4.2879	4.2887
UCG	4.6017	4.6019	4.6026	**4.6015**	4.6018	4.6025
INGA	**4.2673**	4.2678	4.2685	4.2674	4.2679	4.2686
DNB	3.9806	**3.9796**	3.9802	3.9816	3.9807	3.9814
BCP	4.6322	4.6327	4.6336	**4.6320**	4.6325	4.6334
BBVA	4.0630	4.0634	4.0631	**4.0629**	4.0632	**4.0629**
SAB	4.2129	4.2133	4.2140	**4.2125**	4.2128	4.2135
SAN	4.0969	**4.0964**	4.0971	4.0975	4.0972	4.0980
SEB-A	**3.9652**	**3.9652**	3.9659	3.9667	3.9669	3.9677
SHB-A	3.6872	3.6860	3.6867	3.6868	**3.6857**	3.6864
SWED-A	3.9980	**3.9976**	3.9983	**3.9976**	3.9981	3.9988
CSGN	4.1201	4.1206	4.1213	**4.1198**	4.1203	4.1210
UBSG	4.0207	4.0207	4.0215	**4.0204**	**4.0204**	4.0212
BARC	4.3633	4.3632	4.3640	**4.3628**	**4.3628**	4.3635
HSBC	3.4244	3.4243	3.4251	3.4239	**3.4238**	3.4247
LLOY	4.1503	4.1488	4.1497	4.1500	**4.1485**	4.1493
NWG	4.4752	**4.4741**	4.4750	4.4748	4.4736	4.4745
STAN	4.1430	4.1423	4.1430	4.1428	**4.1422**	4.1428

Table 4: AIC values for different specifications of an ARMA(u,v)-GARCH(p,q) process with and without a constant for the ARMA(u,v) process (for each row, boldfaced numbers indicate the specification for which the minimum AIC values is reached). For most of the univariate series ARMA(1,0)-GARCH(1,1) with long term mean equal to zero provides close to optimal fit.

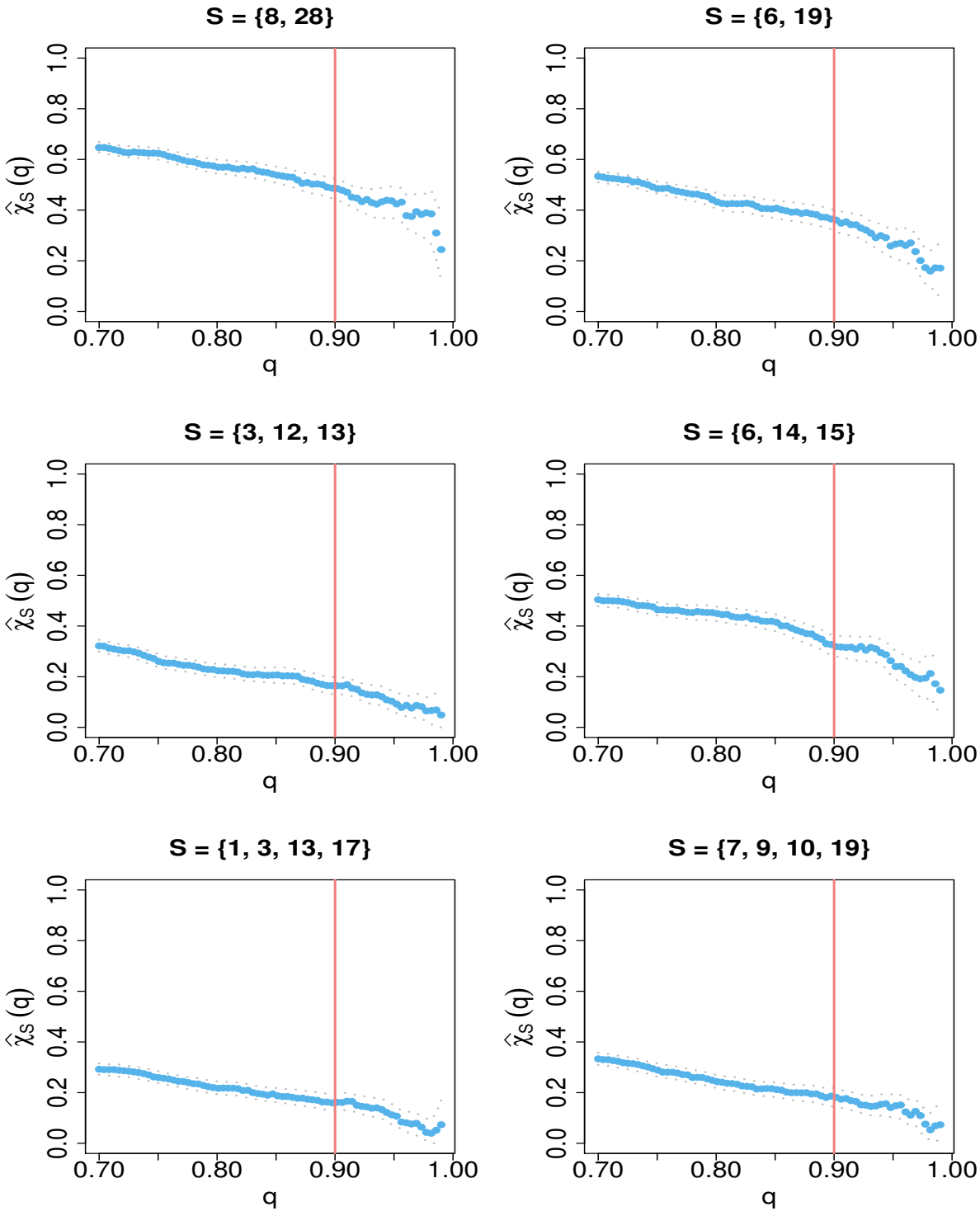


Figure 12: Estimated tail correlation coefficients $\widehat{\chi}_S(q)$ for randomly chosen subsets S of banks (indices can be found in Table 3) against a quantile q for time series data from 2005 to 2020. The vertical line corresponds to our chosen threshold.

Bank ID	Bank	$CC(j)$	$BC(j)$
1	Ereste Group Bank AG	0.7565	0
2	Raiffeisen Bank International AG	0.7794	0
3	KBC Group NV	1.0319	0.2968
4	Danske Bank A/S	0.7799	0
5	Nordea Bank Abp	0.7340	0
6	BNP Paribas SA	1.2565	0.7441
7	Crédit Agricole SA	0.7876	0
8	Societe Generale SA	0.8377	0.1806
9	Commerzbank AG	0.6407	0
10	Deutsche Bank AG	0.9037	0.0645
11	National Bank of Greece	0.7662	0
12	AIB Group PLC	0.8307	0
13	Bank of Ireland Group PLC	0.7934	0
14	Banco BPM SpA	0.6484	0
15	Intesa Sanpaolo SpA	0.6279	0
16	UniCredit SpA	0.9221	0.1247
17	ING Groep NV	0.9747	0.2882
18	DNB ASA	0.7485	0
19	Banco Comercial Portugues SA	0.9285	0
20	Banco de Sabadell SA	0.6666	0
21	Banco Santander SA	0.9318	0.2925
22	BBVA SA	0.6031	0
23	Skandinaviska Enskilda Banken AB	0.7544	0.1247
24	Svenska Handelsbanken AB	0.5686	0
25	Swedbank AB	0.5560	0
26	Credit Suisse Group AG	0.8242	0
27	UBS Group AG	0.8301	0
28	Barclays PLC	0.6346	0
29	HSBC Holdings PLC	0.7217	0.0645
30	Lloyds Banking Group PLC	0.6559	0.0645
31	Natwest Group PLC	0.5119	0
32	Standard Chartered PLC	0.5629	0

Table 5: Estimated raw values for closeness centrality (CC), betweenness centrality (BC) for each bank (indices can be found in Table 3).

B HR Tree Model

A Hüsler-Reiss graphical tree is defined as follows

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\Lambda(\mathbf{1})} \prod_{\{i,j\}\in E^{(T)}} \frac{\lambda_{ij}(y_i, y_j)}{y_i^{-2} y_j^{-2}} \prod_{i\in V} y_i^{-2}$$
$$= \frac{1}{\Lambda(\mathbf{1})} \prod_{\{i,j\}\in E^{(T)}} \frac{1}{y_i^{-2} y_j^{-2}} \frac{y_i^{-2} y_j^{-1}}{\sqrt{2\pi\Gamma_{ij}}} \exp\Big[- \frac{\{\log(y_j/y_i) + \Gamma_{ij}/2\}^2}{2\Gamma_{ij}} \Big] \prod_{i\in V} y_i^{-2}, \qquad (38)$$

where λ_{ij} denote the bivariate marginal distributions of the exponent measure Λ of $\boldsymbol{Y}$, which follow a bivariate Hüsler-Reiss distribution.

C Code

```
##################################################################
# Illustrate Block-Maxima vs Threshold Exceedances Approach
##################################################################
x1 <- c(0.2, 0.54)
x2 <- c(0.2, 0.7)

x_M <- c(x1, x2)
plot(x= c(c(seq(0,1,0.5)), c(seq(0,1,0.5)), y=x_M))
arrows(0,0.5,1,1)

plot(x= c(0.1,0.1), y= c(0,1), type="l")

x_V = seq(10,200,10)
x_M = matrix(x_V, byrow = FALSE, nrow = 20)
y_V1 = rep(0,20)
y_V2 = runif(20, 0.1, 0.9)
```

```
y_M = matrix(y_V, byrow = FALSE, nrow = 20)

plot(x=x_M, y=y_M, type="l")
plot(x=c(20,20), y=c(0,0.6),)

library(ggplot2)
geom_vline(xintercept=3)

plot.new()
par(mfrow = c(1,2))
plot(x= x_V, y=y_V1, xlim= c(0,200), ylim=c(0,1), type="n", frame.plot = TR
segments(x0 = x_V, y0 = y_V1, x1 = x_V, y1 = y_V2)
segments(x0 = c(49, 99, 149, 199), y0 = c(0,0,0,0), x1= c(49, 99, 149, 199)
text(x=seq(25,200, 50) , y=c(0.9, 0.9, 0.9, 0.9) , labels= c("Block 1", "Bl

plot(x= x_V, y=y_V1, xlim= c(0,200), ylim=c(0,1), type="n", frame.plot = TR
segments(x0 = x_V, y0 = y_V1, x1 = x_V, y1 = y_V2)
segments(x0 = c(0), y0 = c(0.7), x1= c(200), y1 = c(0.7), col="red", lty =


par('ann = FALSE', no.readonly = FALSE)

##########################################################################
##########################################################################
##########################################################################

############################
# Analysis of stock returns
############################

library(readxl)
library(dplyr)
```

```
library(xts)
library(PerformanceAnalytics)
library(MASS)

#LOAD DATA AND CONVRT TO XTS
fd <- file.path("C:", "Users", "andre", "Desktop", "Thesis", "Data", "Data.
data <- read_excel(fd, sheet = "Europe(P)")
# data_list <- lapply(excel_sheets(fd), read_excel, path = fd)
data_xts <- xts(data[,-1], order.by = as.POSIXct(data$DATE))
data_xts <- data_xts[1:4094]

#LIST OF BANKS SORTED BY COUNTRY AND NAME
colnames_sorted <- c("29", "6", "7", "21", "8", "28", "10", "30", "15", "17

# COMPUTE LOG RETURNS
data_returns <- diff(log(data_xts), lag = 1) * 100
data_returns <- data_returns[2:nrow(data_returns)]
data_losses <- - data_returns

# PLOT PRICES AND RETURNS
plot.new()
plot(x = data_xts[,"HSBC"], xlab = "Time", ylab = "Price",
     main = "Price")
plot.new()
par(mfrow = c(6,2))
plot(x = data_xts[,"HSBC"], grid.col = NA, xlab = "Time", ylab = "", main =
plot(x = data_returns[,"HSBC"], grid.col = NA, xlab = "Time", ylab = "Log_R

plot(x = data_xts[,"BNP_PARIBAS"], grid.col = NA, xlab = "Time", ylab = "P
plot(x = data_returns[,"BNP_PARIBAS"], grid.col = NA, xlab = "Time", ylab =

plot(x = data_xts[,"UBS"], grid.col = NA, xlab = "Time", ylab = "Price", ma
```

```
plot(x = data_returns[,"UBS"], grid.col = NA, xlab = "Time", ylab = "Log R

plot(x = data_xts[,"DEUTSCHE_BANK"], grid.col = NA, xlab = "Time", ylab = "
plot(x = data_returns[,"DEUTSCHE_BANK"], grid.col = NA, xlab = "Time", ylab

plot(x = data_xts[,"INTESA_SANPAOLO"], grid.col = NA, xlab = "Time", ylab =
plot(x = data_returns[,"INTESA_SANPAOLO"], grid.col = NA, xlab = "Time", yl

plot(x = data_xts[,"SWEDBANK"], grid.col = NA, xlab = "Time", ylab = "Price
plot(x = data_returns[,"SWEDBANK"], grid.col = NA, xlab = "Time", ylab = "I




# lines(x = basket[,"GDX.Close"], col = "goldenrod")
# lines(x = basket[,"DBO.Close"], col = "darkblue")
# lines(x = basket[,"VWO.Close"], col = "darkviolet")
# legend(x = 'topleft', legend = c("SPY", "QQQ", "GDX", "DBO", "VWO"),
#        lty = 1, col = myColors)

# QQ PLOTS
par(mfrow = c(3,2))
chart.QQPlot(data_returns[,"HSBC"], main="Normal Q-Q Plot of HSBC Returns",
chart.QQPlot(data_returns[,"BNP_PARIBAS"], main="Normal Q-Q Plot of BNP Pa
chart.QQPlot(data_returns[,"UBS"], main="Normal Q-Q Plot of UBS Returns", l
chart.QQPlot(data_returns[,"DEUTSCHE_BANK"], main="Normal Q-Q Plot of Deuts
chart.QQPlot(data_returns[,"INTESA_SANPAOLO"], main="Normal Q-Q Plot of Int
chart.QQPlot(data_returns[,"SWEDBANK"], main="Normal Q-Q Plot of Swedbank I

# CORRELOGRAMS
plot.new()
chart.ACF(data_returns[,"HSBC"], main = "ACF")
```

```
plot.new()
par(mfrow = c(3,2))
returns_HSBC <- data_returns[,"HSBC"]
returns_HSBC <- ts(returns_HSBC)
ccf(returns_HSBC[, 1, drop = TRUE], returns_HSBC[, 1, drop = TRUE]^2, main

returns_BNP <- data_returns[,"BNP_PARIBAS"]
returns_BNP <- ts(returns_BNP)
ccf(returns_BNP[, 1, drop = TRUE], returns_BNP[, 1, drop = TRUE]^2, main =

returns_UBS <- data_returns[,"UBS"]
returns_UBS <- ts(returns_UBS)
ccf(returns_UBS[, 1, drop = TRUE], returns_UBS[, 1, drop = TRUE]^2, main =

returns_DB <- data_returns[,"DEUTSCHE_BANK"]
returns_DB <- ts(returns_DB)
ccf(returns_DB[, 1, drop = TRUE], returns_DB[, 1, drop = TRUE]^2, main = "C

returns_ISP <- data_returns[,"INTESA_SANPAOLO"]
returns_ISP <- ts(returns_ISP)
ccf(returns_ISP[, 1, drop = TRUE], returns_ISP[, 1, drop = TRUE]^2, main =

returns_SB <- data_returns[,"SWEDBANK"]
returns_SB <- ts(returns_SB)
ccf(returns_SB[, 1, drop = TRUE], returns_SB[, 1, drop = TRUE]^2, main = "C



# hsbc_return <- data_hsbc[-1]

#Ljung-Box tests
names <- dimnames(data_losses)
LB_test <- lapply(1:ncol(data_returns), function(i) Box.test(data_returns[,
```

```
LB_statistic <- sapply(LB_test, "[[", "statistic")
LB_pvalue <- sapply(LB_test, "[[", "p.value")


returns_squared <- data_returns^2
LB_test_sq <- lapply(1:ncol(returns_squared), function(i) Box.test(returns_
LB_sq_stat <- sapply(LB_test_sq, "[[", "statistic")
LB_sq_p <- sapply(LB_test_sq, "[[", "p.value")

names <- array(names)

LB <- c(unlist(names, use.names = FALSE), c(LB_pvalue))
LB <- matrix(LB, ncol=2, byrow = FALSE)

LB_squared <- c(unlist(names, use.names = FALSE), c(LB_sq_p))
LB_squared <- matrix(LB_squared, ncol=2, byrow = FALSE)

# JB TEST
library(tseries)
JB_test <- lapply(1:ncol(returns_squared), function(i) jarque.bera.test(ret
JB_pvalue <- sapply(JB_test, "[[", "p.value")
JB<- c(unlist(names, use.names = FALSE), c(JB_pvalue))
JB <- matrix(JB, ncol = 2, byrow = FALSE)


# ARMA-GARCH
library(fGarch)

AIC1011 <- vector("numeric", ncol(data_returns))
AIC1111 <- vector("numeric", ncol(data_returns))
AIC1121 <- vector("numeric", ncol(data_returns))
```

```
losses_filtered <- numeric(0)
for (j in 1:ncol(data_losses)) {
  arma_var <- "arma(1,0)"
  garch_var <- "garch(1,1)"
  univariate_data <- data_losses[,j]
  f <- as.formula(paste("~", paste(arma_var, "+", garch_var)))
  garch <- garchFit( formula = f, data=univariate_data, include.mean = TRUE
                     cond.dist = c("QMLE"), trace = FALSE)
  # if (class(garch) == "try-error") next
  AIC <- garch@fit[["ics"]][1]
  AIC1011[j] <- AIC
  a <- garch@residuals/garch@sigma.t
  losses_filtered <- append(losses_filtered, a)
}

AIC <- c(unlist(names, use.names = FALSE), AIC1011, AIC1111, AIC1121)
AIC <- matrix(AIC, ncol = 4, byrow = FALSE)



losses_filtered <- matrix(losses_filtered, ncol = 32, byrow = FALSE)
losses_filtered <- losses_filtered[-1,]

# QQ PLOTS - FILTERED LOSSES
par(mfrow = c(3,2))
chart.QQPlot(losses_filtered[,"HSBC"], main="Normal Q-Q Plot of HSBC Return
chart.QQPlot(losses_filtered[,"BNP_PARIBAS"], main="Normal Q-Q Plot of BNP_
chart.QQPlot(losses_filtered[,"UBS"], main="Normal Q-Q Plot of UBS Returns"
chart.QQPlot(losses_filtered[,"DEUTSCHE_BANK"], main="Normal Q-Q Plot of D
chart.QQPlot(losses_filtered[,"INTESA_SANPAOLO"], main="Normal Q-Q Plot of
chart.QQPlot(losses_filtered[,"SWEDBANK"], main="Normal Q-Q Plot of Swedban
```

```
# CORRELOGRAMS - FILTERED LOSSES
par(mfrow = c(3,2))
ccf(losses_filtered[, "HSBC", drop = TRUE], losses_filtered[, "HSBC", drop
    ylim = range (-0.05,1), yaxs = "i", xaxs = "i")

ccf(losses_filtered[, "BNP_PARIBAS", drop = TRUE], losses_filtered[, "BNP_P
    ylim = range (-0.05,1),yaxs = "i", xaxs = "i")

ccf(losses_filtered[, "UBS", drop = TRUE], losses_filtered[, "UBS", drop =
    ylim = range (-0.05,1),yaxs = "i", xaxs = "i")

ccf(losses_filtered[, "DEUTSCHE_BANK", drop = TRUE], losses_filtered[, "DEU
    ylim = range (-0.05,1),yaxs = "i", xaxs = "i")

ccf(losses_filtered[, "INTESA_SANPAOLO", drop = TRUE], losses_filtered[, "I
    ylim = range (-0.05,1),yaxs = "i", xaxs = "i")

ccf(losses_filtered[, "SWEDBANK", drop = TRUE], losses_filtered[, "SWEDBANK
    ylim = range (-0.05,1),yaxs = "i", xaxs = "i")



hsbc_return <- data_hsbc[-1]

# EMPIRICAL CORR COEFF FOR DIFFERENT SUBSETS
par(mfrow = c(3,2), xpd=FALSE)
Banks <- losses_filtered[,c(5,6)]
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {8, 28}")
Banks <- losses_filtered[,c(2,31)]
```

```
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {6, 19}")
Banks <- losses_filtered[,c(21,30, 29)]
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {3, 12, 13}")
Banks <- losses_filtered[,c(2, 27, 9)]
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {6, 14, 15}")
Banks <- losses_filtered[,c(10, 29, 21, 24)]
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {1, 3, 13, 17}")
Banks <- losses_filtered[,c(19, 7, 3, 31)]
chiPlot(data=Banks, ylabel=expression(widehat(chi)[S]~(q)), chimod=NULL, ns
        qmin = 0.7, qmax = 0.99, maintxt = "S = {7, 9, 10, 19}")


chiEmp(data = Banks, nq=70, qmin = 0.5, qmax = 0.99)

# CONSTRUCTION OF OPTIMAL GRAPHICAL MODEL
library(graphicalExtremes)
library(igraph)

losses_normalized <- data2mpareto(losses_filtered, p= 0.9)
losses_tree <- mst_HR(losses_normalized, p = NULL, cens = TRUE)
plot(losses_tree$tree)
candidate_edges <- select_edges(losses_tree$tree)
losses_graph <- fmpareto_graph_HR(losses_normalized, graph = losses_tree$tr
optimal_graph <- losses_graph$graph[[length(losses_graph$graph)]]
V(optimal_graph)$name <- colnames_sorted


losses_normalized_88 <- data2mpareto(losses_filtered, p= 0.88)
```

```
losses_tree_88 <- mst_HR(losses_normalized_88, p = NULL, cens = TRUE)
candidate_edges_88 <- select_edges(losses_tree_88$tree)
losses_graph_88 <- fmpareto_graph_HR(losses_normalized_88, graph = losses_t
optimal_graph88 <- losses_graph_88$graph[[length(losses_graph_88$graph)]]
V(optimal_graph88)$name <- colnames_sorted

losses_normalized_92 <- data2mpareto(losses_filtered , p= 0.92)
losses_tree_92 <- mst_HR(losses_normalized_92, p = NULL, cens = TRUE)
candidate_edges_92 <- select_edges(losses_tree_92$tree)
losses_graph_92 <- fmpareto_graph_HR(losses_normalized_92, graph = losses_t
optimal_graph92 <- losses_graph_92$graph[[length(losses_graph_92$graph)]]
V(optimal_graph92)$name <- colnames_sorted



# plot(losses_tree$tree, vertex.size = 15, vertex.label = colnames_sorted,
#       edge.width = 3, margin = c(0,0,0,0), frame = TRUE, main = expression
#
# plot(losses_graph$graph[[length(losses_graph$graph)]], vertex.size = 15,
#       edge.width = 3, margin = c(0,0,0,0), frame = TRUE, main = expression

graphintersection88 <- optimal_graph %s% optimal_graph88
common_edges_88 <- E(graphintersection88)
ends(graphintersection88 , common_edges_88)

graphintersection92 <- optimal_graph %s% optimal_graph92
common_edges_92 <- E(graphintersection92)
ends(graphintersection92 , common_edges_92)



id <- tkplot(optimal_graph, canvas.width = 800, canvas.height = 600, vertex
             edge.width = 3, margin = c(0.2,0.2,0.2,0.2))
tkcreate(tkplot.canvas(id), "line", 740, 60, 740, 540, width=1)
```

```
tkcreate(tkplot.canvas(id), "line", 60, 540, 740, 540, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 740, 60, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 60, 540, width=1)
canvas <- tkplot.canvas(id)
coords <- tk_coords(tkp.id = 24, norm = TRUE)

id <- tkplot(losses_tree$tree, canvas.width = 800, canvas.height = 600, ver
             edge.width = 3, margin = c(0.2,0.2,0.2,0.2))


id <- tkplot(optimal_graph88, canvas.width = 800, canvas.height = 600, vert
       edge.width = 3, margin = c(0.2,0.2,0.2,0.2))
tkcreate(tkplot.canvas(id), "line", 740, 60, 740, 540, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 540, 740, 540, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 740, 60, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 60, 540, width=1)
canvas <- tkplot.canvas(id)
# width <- as.numeric(tkcget(canvas, "-width"))
# tkcreate(tkplot.canvas(id), "text", width/2, 40, text="q = 0.88",
#          justify="center", font=tcltk::tkfont.create(family="sans", slant
"italic", size=20))
# tcltk::tkpostscript(canvas, file="C:/Users/andre/Desktop/Thesis/Plots/Op
coords88 <- tk_coords(id, norm = TRUE)

id <- tkplot(optimal_graph92, canvas.width = 800, canvas.height = 600, vert
             edge.width = 3, margin = c(0.2,0.2,0.2,0.2))
tkcreate(tkplot.canvas(id), "line", 740, 60, 740, 540, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 540, 740, 540, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 740, 60, width=1)
tkcreate(tkplot.canvas(id), "line", 60, 60, 60, 540, width=1)
canvas <- tkplot.canvas(id)
coords92 <- tk_coords(id, norm = TRUE)
```

```
AICvsEdges <- numeric(0)
for (i in 1:length(losses_graph$graph)) {
  count_edges <- ecount(losses_graph$graph[[i]])
  AIC_value <- losses_graph$AIC[i]
  a <- c(AIC_value, count_edges)
  AICvsEdges <- append(AICvsEdges, a)

}
AICvsEdges <- matrix(AICvsEdges, ncol = 2, byrow = TRUE)


plot.new()
par(mfrow = c(1,1), xpd=FALSE)
plot(AICvsEdges[,1] ~ AICvsEdges[,2], pch = 19, col = "#56B4E9", type = 'b'
plot(Gamma2chi(losses_tree$Gamma) ~ emp_chi_mat(losses_filtered, p = 0.9),
     xlab = "Fitted_Model", ylab = "Empirical", ylim = c(0,1), xlim = c(0,1
segments(x0 = 0, y0=0, x1=1, y1=1, col = "grey", lwd = 2)

plot(Gamma2chi(losses_graph_92$Gamma[[length(losses_graph$Gamma)]]) ~ emp_c
                    xlab = "Fitted_Model", ylab = "Empirical", ylim = c(0,1
segments(x0 = 0, y0=0, x1=1, y1=1, col = "grey", lwd = 2)


#CENTRALITY MEASURES
Gamma_optimal <- losses_graph$Gamma[[length(losses_graph$Gamma)]]
Chi_optimal <- Gamma2chi(Gamma_optimal)
adjacency_matrix_optimal <- as_adjacency_matrix(optimal_graph, type = "both
weights <- adjacency_matrix_optimal * Chi_optimal
weighted_graph <- graph_from_adjacency_matrix(weights, mode = "upper", wei
closeness_centr <- closeness(weighted_graph, normalized = TRUE)
betweenness_centr <- betweenness(weighted_graph, normalized = TRUE)
```

```
write.
```

```
#FUNCTIONS

chiPlot <- function(data, ylabel, chimod, nsim, nq = 35, qmin = 0.5, qmax =
  tmp <- matrix(,nrow=nsim,ncol=nq)
  n <- nrow(data)
  for(j in 1:nsim){
    nsample <- sample(1:n,size=n,replace=T)
    newdata <- data[nsample,]
    tmp[j,] <- chiEmp(newdata,nq=nq,qmin=qmin,qmax=qmax)[,2]
  }
  CIlow <- apply(tmp,2,quantile,0.025)
  CIhigh <- apply(tmp,2,quantile,0.975)
  chi <-chiEmp(data,nq=nq,qmin=qmin,qmax=qmax)

  par(cex.lab=2,cex.axis=2,cex.main=2,mar=c(5,4.4,4,2)+0.9)
  plot(chi[,1],chi[,2],ylim=c(0,1),xlab="q",
       ylab=ylabel,lwd=2, main = maintxt, col = "#56B4E9", pch = 19)
  lines(chi[,1],CIlow,lty=3,lwd=2, col = "gray")
  lines(chi[,1],CIhigh,lty=3,lwd=2, col = "gray")
  abline(h = chimod, lwd = 2)
  abline(v = 0.9, col = "lightcoral", lwd = 2, mar=c(5,4.4,4,2)+0.9)
}

chiEmp <-function(data,nq=25,qmin, qmax){
  n<-nrow(data)
```

```
  datatr<-(apply(data,2,rank) - 0.5)/n
  qlim<-c(qmin,qmax)
  u <- seq(qlim[1], qlim[2], length = nq)
  cu<-sapply(c(1:nq), function(i) mean(apply(datatr,1,min) >= u[i]))
  return(cbind(u,cu/(1-u)))
}

dim_Gamma <- function(Gamma) {
  dimension <- dim(Gamma)

  if ((length(dimension) == 2) & (dimension[1] == dimension[2])) {
    dimension[1]
  } else {
    stop("Not a square matrix!")
  }
}

select_edges <- function(graph) {
  d <- igraph::vcount(graph)
  adj_mat <- igraph::as_adjacency_matrix(graph, sparse = FALSE) > 0


  sel.edges <- matrix(0, nrow = 0, ncol = 2)

  for (i in 1:(d - 1)) {
    for (j in (i + 1):d) {

      # set new_edge
      new_edge <- c(i, j)

      # check if new_edge is already in the graph
      is_already_edge <- adj_mat[i, j] | adj_mat[j, i]
```

```
      # if not, add it to the graph; else skip to the next
      if (!is_already_edge) {
        extended_graph <- igraph::add_edges(graph = graph, edges = new_edge
      } else {
        next
      }

      # check if new graph is decomposable
      is_chordal <- igraph::is_chordal(extended_graph)$chordal

      # measure the length of the path from i to j in the old graph
      length_path <- length(as.vector(
        igraph::shortest_paths(graph, from = i, to = j)$vpath[[1]]
      ))

      if (is_chordal & length_path != 2) {
        sel.edges <- rbind(sel.edges, new_edge, deparse.level = 0)
      }
    }
  }

  return(sel.edges)
}

mparetomargins <- function(data, set_indices) {
  data_sub <- data[, set_indices]
  idx <- which(apply(data_sub, 1, max) > 1)
  return(data[idx, set_indices])
}
```

```
fmpareto_HR <- function(data,
                        p = NULL,
                        cens = FALSE,
                        init,
                        maxit = 100,
                        graph = NULL,
                        method = "BFGS") {
  if (!is.null(p)) {
    # if p provided -> data not Pareto -> to convert
    data <- data2mpareto(data, p)
  } else {
    # if p not provided -> data already Pareto
    data <- data
  }

  # censoring at 1 since data already normalized
  p <- 1
  d <- ncol(data)
  if (length(p) == 1) {
    p <- rep(p, d)
  }

  # negative log likelihood function
  if (cens) {
    # censor below the (multivariate) threshold
    data.p <- censor(data, p)
    r <- nrow(data.p)

    L <- apply(data.p > matrix(p, ncol = d, nrow = r, byrow = TRUE), 1, wh

    if (is.matrix(L)) {
      L <- split(t(L), 1:r)
```

```
}

I <- which(lapply(L, length) > 0 & lapply(L, length) < d)
J <- which(lapply(L, length) == d)

nllik <- function(par) {
  if (!is.null(graph)) {
    Gtmp <- complete_Gamma(graph = graph, Gamma = par)
    par <- Gtmp[upper.tri(Gtmp)]
  }

  G <- par2Gamma(par)
  S <- Gamma2Sigma(G, k = 1)

  if (any(par <= 0) | !matrixcalc::is.positive.definite(S)) {
    return(10^50)
  }

  else {
    if (length(I) > 0) {
      y1 <- mapply(logdVK_HR,
                   x = as.list(data.frame(t(data.p)))[I],
                   K = L[I], MoreArgs = list(par = par)
      )
    }
    else {
      y1 <- 0
    }
    if (length(J) > 0) {
      y2 <- logdV_HR(x = data.p[J, ], par = par)
    }
    else {
```

```
        y2 <- 0
      }
      y <- sum(y1) + sum(y2) - (length(I) + length(J)) * log(V_HR(p, par
      return(-y)
    }
  }
}
else {
  r <- nrow(data)
  L <- apply(data > matrix(p, ncol = d, nrow = r, byrow = TRUE), 1, which

  if (is.matrix(L)) {
    L <- split(t(L), 1:r)
  }

  I <- which(lapply(L, length) > 0) # 1:r
  nllik <- function(par) {
    if (!is.null(graph)) {
      Gtmp <- complete_Gamma(graph = graph, Gamma = par)
      par <- Gamma2par(Gtmp)
    }

    G <- par2Gamma(par)
    S <- Gamma2Sigma(G, k = 1)

    if (any(par <= 0) | !matrixcalc::is.positive.definite(S)) {
      return(10^50)
    }
    else {
      if (length(I) > 0) {
        y1 <- logdV_HR(x = data[I, ], par = par)
      }
```

```
        else {
          y1 <- 0
        }
        y <- sum(y1) - length(I) * log(V_HR(p, par = par))
        return(-y)
      }
    }
  }

  # optimize likelihood
  opt <- stats::optim(init, nllik,
                      hessian = TRUE,
                      control = list(maxit = maxit), method = method
  )

  z <- list()
  z$convergence <- opt$convergence
  z$par <- opt$par
  if (is.null(graph)) {
    z$Gamma <- par2Gamma(z$par)
  } else {
    z$Gamma <- complete_Gamma(graph = graph, Gamma = z$par)
  }
  z$nllik <- opt$value
  z$hessian <- opt$hessian
  return(z)
}

censor <- function(x, p) {
  f2 <- function(x, p) {
    x_is_less <- x <= p
    y <- x
```

```
    y[x_is_less] <- p[x_is_less]
    return(y)
  }
  return(t(apply(x, 1, f2, p)))
}
```