Conditional Systemic Risk with Penalized Copula

31st July 2015

Ostap Okhrin[*] **Alexander Ristig**[†] **Jeffrey Sheen**[‡] **Stefan Trück**[§]

Abstract

Financial contagion and systemic risk measures are commonly derived from conditional quantiles by using imposed model assumptions such as a linear parametrization. In this paper, we provide model free measures for contagion and systemic risk which are independent of the specification of conditional quantiles and simple to interpret. The proposed systemic risk measure relies on the contagion measure, whose tail behavior is theoretically studied. To emphasize contagion from extreme events, conditional quantiles are specified via hierarchical Archimedean copula. The parameters and structure of this copula are simultaneously estimated by imposing a non-concave penalty on the structure. Asymptotic properties of this sparse estimator are derived and small sample properties illustrated using simulations. We apply the proposed framework to investigate the interconnectedness between American, European and Australasian stock market indices, providing new and interesting insights into the relationship between systemic risk and contagion. In particular, our findings suggest that the systemic risk contribution from contagion in tail areas is typically lower during times of financial turmoil, while it can be significantly higher during periods of low volatility.

JEL classification: C40, C46, C51, G1, G2
Keywords: Conditional quantile, Copula, Financial contagion, Spill-over effect, Stepwise penalized ML estimation, Systemic risk, Tail dependence.

1. Introduction

The financial crisis in 2008 revealed the need to model and measure interconnectedness between

[*]Chair of Econometrics and Statistics esp. Transportation, Institute of Economics and Transport, Faculty of Transportation, Dresden University of Technology, Helmholtzstraße 10, 01069 Dresden, Germany, ostap.okhrin@tu-dresden.de

[†]Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany, alexander.ristig@hu-berlin.de

[‡]Department of Economics, Macquarie University, NSW, 2109, Australia, jeffrey.sheen@mq.edu.au

[§]Centre for Financial Risk, Macquarie University, NSW, 2109, Australia, stefan.trueck@mq.edu.au

The research was supported by the Deutsche Forschungsgemeinschaft through the CRC 649 "Economic Risk", Humboldt-Universität zu Berlin and the International Research Training Group 1792. Ristig, Sheen and Trück acknowledge support from the Centre for International Finance and Regulation through the Research Grant "Early-Warning Systems and Managing Systemic Risks using Real-Time Financial and Business Conditions Indicators".

The authors are grateful to Martin Hain and Yarema Okhrin for helpful discussions and suggestions.

financial markets and financial institutions. The reason is that high connectivity among financial markets/institutions causes additional risk which is transmitted between market participants. For example, due to the connectedness of financial firms, risk managers have an increased interest in the risk transmitted to their institutions from other institutions, as the exposed uncertainty has to be taken as given and uncontrollable. Moreover, policy makers and regulators aim at identifying risk vectors that will enable them to react to situations of market stress in an suitable manner. As a result of the crisis, many concepts for measuring spill-over effects (contagion) and systemic risk have been proposed with different advantages and disadvantages. See, e.g., Ishikawa, Kamada, Kurachi, Nasu, and Teranishi (2012) or Brunnermeier and Oehmke (2013) for a summary of recent empirical literature, and Bisias, Flood, Lo, and Valavanis (2012) for a survey on quantitative approaches to the measurement of systemic risks. A few of these approaches are briefly reviewed in the sequel.

Diebold and Yilmaz (2014) suggest a consistent and handy tool for describing connectedness between financial institutions on the basis of corresponding realized log-volatilities. The proposed connectedness and systemic risk measures are constructed from generalized forecast error variance decompositions, whose computation requires an estimate of the covariance matrix of firm-specific idiosyncratic shocks. The measures are naturally forward looking and allow statements concerning bilateral contagion (between two risk factors), multilateral contagion (between several risk factors) and systemic risk (pollution among all risk factors). However, the underlying time series should be a proxy for risk, as the concept of Diebold and Yilmaz (2014) is based on the conditional mean of the underlying time series. For example, the body of the distribution of financial returns does not appropriately reflect risk, for which reason the proposed measures cannot be interpreted as risk-channel when applied to financial returns.

Further approaches study systemic risk based on credit risk models to assess the probability of default. For example, Lucas, Schwaab, and Zhang (2014) and Cherubini and Mulinacci (2015) use CDS prices to investigate spill-over effects of sovereign default risks within the Euro-area and contagion within the European banking system, respectively. While Lucas *et al.* (2014) apply a methodology using dynamic skewed-t distributions, Cherubini and Mulinacci (2015) build their analysis on hierarchical Archimedean copulae (HAC). In general, copulae became a standard tool for modeling non-linear and asymmetric dependence among risk factors, which are also interesting features for describing characteristics of financial systems. One of the key issues in measuring risks within a financial system is to appropriately specify the dependence structure between financial assets. This is of even greater importance, since, for example, Coval, Jurek, and Stafford (2009); Zimmer (2012), conclude that poor dependence models can be considered as one of the reasons for the collapse of CDO markets and related securities, and the subsequent financial crisis.

Based on financial returns, contagion and systemic risk is commonly measured by variations of the expected shortfall and Value-at-Risk (VaR). For example, Acharya, Pedersen, Philippon, and Richardson (2010) present the marginal and systemic expected shortfall which are related to economic theory and employed to assess the extent a financial institution is affected by systemic events. Adrian and Brunnermeier (2011) derive contagion and systemic risk measures from conditional quantile functions and define CoVaR as the VaR of the return distribution of all system constituents conditional on the VaR of a financial institution. The authors mainly investigate the contribution of financial institutions to systemic risk. Theoretical properties of marginal/systemic expected shortfall and CoVaR

are comprehensively discussed in Mainik and Schaanning (2014). Shortcomings of the quantitive Co-VaR approach such as the omitted variables bias and the linear-model specification are addressed in Hautsch, Schaumburg, and Schienle (2015) and Härdle, Wang, and Yu (2015). An alternative approach is discussed in White, Kim, and Manganelli (2015), who extend the conditional autoregressive VaR approach (CAViaR) of Engle and Manganelli (2004) and propose Vectorautoregressions for VaR in order to study dynamics of tail dependence among constituents over time.

Differences between the above mentioned approaches in measuring contagion and systemic risk motivate us to develop a unified framework for describing contagion in tail areas and measuring systemic risk arising from contagion, e.g., from spill-over effects of tail-events. To the best of our knowledge, the difference between conditional and unconditional systemic risk has been ignored in the literature so far, albeit systemic risk due to tail-events can obviously be categorized as conditional systemic risk. This distinction plays a fundamental role in the empirical case study below.

In line with several discussed approaches, our study relies on a portfolio of risk-factors, e.g., negative log-returns. The risk in tail areas is measured with conditional quantile functions. To investigate the effect from one risk factor to another, we define bilateral contagion as the normalized partial derivative of the conditional quantile function with respect to the risk-transmitting component. Quantities of this type are denoted as elasticities in economics and their properties are well established. Moreover, we present the bilateral contagion measure in terms of unconditional quantiles, unconditional quantile densities and a conditional copula-based quantile to study its theoretical properties. For example, contagion in tail areas is shown to be mainly driven by the degree of heterogeneity of involved risks and the underlying dependence structure is shown to be of minor importance in the limit. A heterogenous relation typically causes weak contagion from high-risks to low-risks (e.g., to a risk factor with an exponential tailed distribution) and strong contagion from low-risks to high-risks (e.g., to a risk factor with a heavy tailed distribution).

Our approach straightforwardly yields a matrix of bilateral contagion measures and we derive multilateral measures to explore contagion in tail areas between sub-portfolios. In particular, these are shown to be weighted averages of bilateral contagion measures, where the weights are the corresponding risk measures. Likewise, a conditional systemic risk measure for the entire portfolio is derived. Due to the representation of the systemic risk measure as weighted average, negative dependencies between risks lead naturally to diversified externalities and reduce systemic risk. Moreover, high-risks contribute more to systemic risk, which is driven by both contagion and tail-risk, because of the representation as weighted average of bilateral contagion measures.

To meet the tradeoff between flexibility in tail areas and representing the portfolio with small number of parameters, in our empirical application, we parameterize the conditional quantile function via HAC and unconditional quantile functions. Using results on non-concave penalized Maximum Likelihood (ML) estimation, see Fan and Li (2001), we propose a multi-stage estimation procedure for HAC similar to Okhrin, Okhrin, and Schmid (2013c). In particular, we estimate the parameters and aggregate the structure of HAC simultaneously by imposing a non-concave penalty on the structure. This can be interpreted as penalizing a diversified dependence structure in favor of equi-dependence while accounting for the curvature of the log-likelihood function. Equi-correlation concepts are broadly accepted in the finance literature, e.g., Engle and Kelly (2011), but they have been unattended in the copula literature. The asymptotic properties of our estimation and data-driven aggregation procedure

are derived and small sample properties are illustrated in a simulation study.

As the proposed estimation method allows us to represent the conditional quantile functions with a few parameters, we incorporate time-varying parameters in a rolling window analysis. Changing dependence structures during periods of financial turmoil have been recognized in several studies. For instance, Oh and Patton (2014); Christoffersen and Jacobs (2014) find that financial assets tend to show a stronger dependence during crisis periods than in calm periods. We illustrate the behavior of the proposed contagion and systemic risk measures for nine major stock indices and emphasize the Australasian area. We use daily data on log-returns from January 1, 2007 to April 30, 2014 and show that the dependence structure of the considered system can be traced back to five parameters. We examine bilateral contagion, which supports the theoretical properties of the contagion measure. Our systemic risk measure provides new insights and highlights interesting features which have not been discussed in this context. In particular, while our analysis provides the expected result that an increase in dependence between financial markets also increases systemic risk, the developed *conditional* systemic risk measure also properly describes the part of systemic risk arising from contagion in tail areas. As higher order moments like variances are linked to unconditional quantiles, the conditional systemic risk contribution from contagion in tail areas decreases during times of financial turmoil. In other words, our findings suggest that a potential breakdown caused by contagion in tail areas is unlikely during times of high volatility and more probable during calm times.

The paper is organized as follows. Contagion and systemic risk measures are derived in Section 2. Section 3 discusses the estimation details and Section 4 illustrates the performance of the procedure in a Monte Carlo simulation. Empirical results are presented in Section 5, while Section 6 concludes. Regularity assumptions are stated in Appendix A and proofs are moved to Appendix B.

2. Defining contagion and systemic risk

Let X be a d-dimensional random vector $X = (X_1, \ldots, X_d)^\top$ with cumulative distribution function (cdf) $F(x_1, \ldots, x_d) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$ and define the random vector $X_{\not k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_d)^\top$, i.e., random variable X_k is not included in $X_{\not k}$. The cdf $F(\cdot)$ is assumed to be differentiable and strictly monotonically increasing in each argument. Given this notation, the conditional cdf is denoted by

$$F_{X_k|X_{\not k}=x_{\not k}}(x_k) = \mathrm{P}(X_k \leq x_k | X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, X_{k+1} = x_{k+1}, X_d = x_d). \tag{1}$$

As $F_{X_k|X_{\not k}=x_{\not k}}(x_k)$ is strictly monotonically increasing in x_k, its conditional quantile function is

$$Q_{X_k|X_{\not k}=x_{\not k}}(\alpha) \stackrel{\text{def}}{=} F^{-1}_{X_k|X_{\not k}=x_{\not k}}(\alpha) \quad \text{with} \quad \alpha \in (0,1). \tag{2}$$

In a time series context, the event $\{X_{\not k} = x_{\not k}\}$ might refer to past events, e.g., the previous period's VaR as in Engle and Manganelli (2004). Define $U_j \stackrel{\text{def}}{=} F_j(X_j)$ and $u_j \stackrel{\text{def}}{=} F_j(x_j)$ with $U_j \sim \mathrm{U}(0,1)$ and $u_j \in (0,1)$, $j = 1, \ldots, d$. Following Sklar (1959), $F(\cdot)$ can be decomposed into its marginal cdfs, $F_j(x_j)$, $j = 1, \ldots, d$, and a copula function $C(\cdot)$ describing the dependence between the components of X such that $F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$. Overviews of copulae are given in Joe (1997)

and Nelsen (2006), while recent developments for mathematical and quantitative finance are presented in Jaworski, Durante, and Härdle (2013). Analogously to (1), the conditional copula is given by

$$C_{U_k|U_{\not k}=u_{\not k}}(u_k) = \mathrm{P}(U_k \leq u_k | U_1 = u_1, \ldots, U_{k-1} = u_{k-1}, U_{k+1} = u_{k+1}, U_d = u_d), \tag{3}$$

where $\{U_{\not k} = u_{\not k}\} = \{F_{\not k}(X_{\not k}) = F_{\not k}(x_{\not k})\}$ denotes the event $\{F_1(X_1) = F_1(x_1), \ldots, F_{k-1}(X_{k-1}) = F_{k-1}(x_{k-1}), F_{k+1}(X_{k+1}) = F_{k+1}(x_{k+1}), \ldots, F_d(X_d) = F_d(x_d)\}$. Let $Q_j(\alpha) = F_j^{-1}(\alpha)$, $j = 1, \ldots, d$, be the unconditional marginal quantile functions, $\alpha \in (0, 1)$. Based on the conditional copula and the unconditional quantile functions, the conditional quantile from (2) can be rewritten as

$$Q_{X_k|X_{\not k}=x_{\not k}}(\alpha) = Q_k\{C^{-1}_{U_k|U_{\not k}=u_{\not k}}(\alpha)\} = Q_k\{C^{-1}_{F_k(X_k)|F_{\not k}(X_{\not k})=F_{\not k}(x_{\not k})}(\alpha)\}, \tag{4}$$

where the inverse of $C_{\cdot|\cdot}(u_k)$ is denoted by $C^{-1}_{\cdot|\cdot}(\alpha)$. The latter is called a c-quantile and introduced in Bouyé and Salmon (2009). More recently, Bernard and Czado (2015) provide a comprehensive study about non-linear conditional (c-)quantiles and compare their properties with linear conditional quantiles, see Koenker and Bassett (1978). In particular, the theoretical discussion of Bernard and Czado (2015) is not encouraging if one wishes to approximate non-linear conditional quantiles with linear conditional quantiles especially if the conditioning variable is related to a tail-event.

Let $f_j(x_j) = F_j'(x_j)$ be the unconditional density function and let $q_j(\alpha) = Q_j'(\alpha)$, $\alpha \in (0, 1)$, be the unconditional quantile density function popularized in Parzen (1979) and Jones (1992), $j = 1, \ldots, d$. Based on (4), the derivative of $Q_{X_k|X_{\not k}=x_{\not k}}(\alpha)$ with respect to x_ℓ, $\ell \neq k$, is calculated as

$$q^{\ell}_{X_k|X_{\not k}=x_{\not k}}(\alpha) \stackrel{\text{def}}{=} \frac{\partial}{\partial x_\ell} Q_{X_k|X_{\not k}=x_{\not k}}(\alpha) = \frac{q_k\{C^{-1}_{U_k|U_{\not k}=u_{\not k}}(\alpha)\}}{q_\ell(u_\ell)} \frac{\partial}{\partial u_\ell} C^{-1}_{U_k|U_{\not k}=u_{\not k}}(\alpha). \tag{5}$$

Due to the fact that $F_\ell(X_\ell) \sim \mathrm{U}(0, 1)$, $k \neq \ell$, the conditional quantile $Q_{X_k|X_{\not k}=x_{\not k}}(\alpha) = Q_{X_k|U_{\not k}=u_{\not k}}(\alpha)$ does not depend on the specific laws of X_ℓ, $\ell \neq k$, see, e.g., Bernard and Czado (2015). However, the partial derivative of the conditional quantile function $q^{\ell}_{X_k|X_{\not k}=x_{\not k}}(\alpha)$ depends on the specific law of X_ℓ via the quantile density function $q_\ell(\cdot)$, see Equation 5. Where possible, we follow a short hand notation in the sequel, e.g., $Q_{X_k|X_{\not k}=Q_{\not k}(\alpha)}(\alpha) = Q_{X_k|X_{\not k}=x_{\not k}}(\alpha)\Big|_{x_{\not k}=Q_{\not k}(\alpha)}$ and $Q_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha) = Q_{X_k|U_{\not k}=u_{\not k}}(\alpha)\Big|_{u_{\not k}=\boldsymbol{\alpha}}$, where $Q_{\not k}(\alpha) = \{Q_1(\alpha), \ldots, Q_{k-1}(\alpha), Q_{k+1}(\alpha), \ldots, Q_d(\alpha)\}^\top$ and $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)^\top$ are vectors of same size as $x_{\not k}$ and $u_{\not k}$.

To put the previous statements in an economic context, consider two risks X_k and X_ℓ taking values on the real line, where "good events" like profits are on the negative part and "bad events" such as losses on the positive part of the real line. Let $Q_{X_k|X_\ell}(\alpha)$ be a linear conditional quantile model of the form $Q_{X_k|X_\ell}(\alpha) = a(\alpha) + b(\alpha)X_\ell$ with $\alpha \in (0, 1)$. For small values of α, Adrian and Brunnermeier (2011) propose $Q_{X_k|X_\ell=Q_\ell(\alpha)}(\alpha) = Q_{X_k|U_\ell=\alpha}(\alpha)$ as a risk measure, which is *independent* from the specific law $F_\ell(\cdot)$. In order to measure contagion, the focus, nevertheless, changes from the risk measure given by $Q_{X_k|U_\ell=\alpha}(\alpha)$ to the coefficient $b(\alpha) = q^{\ell}_{X_k|U_\ell=\alpha}(\alpha)$, which *depends* on the quantile density of X_ℓ according to (5). Roughly speaking, $b(\alpha)$ carries information about (i) the *relation* of unconditional risk sensitivities through the quantile density functions and (ii) the sensitivity of dependence caused risk $\frac{\partial}{\partial u_\ell} C^{-1}_{U_k|U_\ell=u_\ell}(\alpha)\Big|_{u_\ell=\alpha}$.

2.1. Bilateral contagion

To stay in an economic context, let $X_1, \ldots, X_d$ be a portfolio of continuously distributed risks, such as negative log-returns of financial institutions or financial markets. By normalizing the derivative of the conditional quantile given in (5) with $Q_\ell(u_\ell)/Q_k\{C^{-1}_{U_k|U_{\not\ell}=u_{\not\ell}}(\alpha)\}$, obtain the standardized contagion measure

$$\mathcal{S}^{u_{\not\ell}}_{k\leftarrow\ell} \stackrel{\text{def}}{=} \frac{Q_\ell(u_\ell)q_k\{C^{-1}_{U_k|U_{\not\ell}=u_{\not\ell}}(\alpha)\}}{q_\ell(u_\ell)Q_k\{C^{-1}_{U_k|U_{\not\ell}=u_{\not\ell}}(\alpha)\}}\frac{\partial}{\partial u_\ell}C^{-1}_{U_k|U_{\not\ell}=u_{\not\ell}}(\alpha), \tag{6}$$

where u_ℓ is a component of the vector $u_{\not\ell}$ according to the introduced notation. The bilateral contagion measure at level $\alpha \in (0,1)$ is then defined as $\mathcal{S}^\alpha_{k\leftarrow\ell} \stackrel{\text{def}}{=} \mathcal{S}^{u_{\not\ell}}_{k\leftarrow\ell}\big|_{u_{\not\ell}=\boldsymbol{\alpha}}$. An expression of form (6) is commonly considered as partial *elasticity*, c.f., Sydsæter and Hammond (1995). Due to non-linearities of conditional quantiles, we import the concept of elasticities in order to interpret the effect on the risk measure $Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)$ by a marginal change in x_ℓ: If the contagion tends to zero, i.e., $|\mathcal{S}^\alpha_{k\leftarrow\ell}| \approx 0$, the risk measure is said to be robust with respect to marginal changes in x_ℓ. Conversely, the risk measure is said to be sensitive or fragile with respect to marginal changes in x_ℓ, if $|\mathcal{S}^\alpha_{k\leftarrow\ell}| \approx \infty$. If $|\mathcal{S}^\alpha_{k\leftarrow\ell}| \approx 1$, the risk of X_k measured through a conditional quantile behaves proportional with respect to changes in x_ℓ. This approach is often described as *ceteris paribus* analysis, i.e., analyzing the effect of a marginal change in x_ℓ, while other variables are held constant. The copula representation of $\mathcal{S}^\alpha_{k\leftarrow\ell}$ is convenient for exploring the theoretical properties of the contagion measure. Nonetheless, $\mathcal{S}^\alpha_{k\leftarrow\ell}$ can also be expressed with an implicit assumption on the copula $C(\cdot)$, i.e.,

$$\mathcal{S}^\alpha_{k\leftarrow\ell} \stackrel{\text{def}}{=} \frac{x_\ell}{Q_{X_k|X_{\not\ell}=x_{\not\ell}}(\alpha)}\frac{\partial}{\partial x_\ell}Q_{X_k|X_{\not\ell}=x_{\not\ell}}(\alpha)\Bigg|_{x_{\not\ell}=Q_{\not\ell}(\alpha)}, \tag{7}$$

where x_ℓ is one component of the vector $x_{\not\ell}$. The computation of bilateral contagion measures via (7) is useful for many applications, where no explicit assumption about the dependence structure is imposed. For ease of notation let $\mathcal{S}^\alpha_{k\ell} = \mathcal{S}^\alpha_{k\leftarrow\ell}$ in the following and let $\{\mathcal{S}^\alpha_{k\ell}\}^d_{k,\ell=1}$ be the contagion matrix collecting all partial elasticities. The contagion matrix has zeros on its diagonal and is (usually) non-symmetric. While zeros on the diagonal are due to absence of contagion to oneself, the asymmetry leads to the following conclusions: If $\mathcal{S}^\alpha_{k\ell}$ and $\mathcal{S}^\alpha_{\ell k}$ have positive signs, the risks X_k and X_ℓ are substitutes. Conversely, if $\mathcal{S}^\alpha_{k\ell}$ and $\mathcal{S}^\alpha_{\ell k}$ have negative signs, the risks are complements. No statement can be made, if $\mathcal{S}^\alpha_{k\ell}$ and $\mathcal{S}^\alpha_{\ell k}$ have different signs.

In order to study the behavior of the proposed contagion measure in tail areas, we introduce the concepts of tail-monotonicity and conditional tail independence. Parzen (1979) calls a density function $f(x)$ with cdf $F(x)$, $Q(u) = F^{-1}(u)$, and tail exponent $\gamma > 0$ tail-monotone, if (i) it is non-decreasing on an interval to the right of $a = \sup\{x : F(x) = 0\}$ and non-increasing on an interval to the left of $b = \inf\{x : F(x) = 1\}$, with $-\infty \le a \le b \le \infty$; (ii) $f(x) > 0$ on $x \in (a,b)$ and $\sup_{x\in(a,b)} F(x)\{1-F(x)\}|f'(x)|/f(x)^2 \le \gamma$. Numerous probability laws have tail-monotone densities such as the Gaussian, Pareto and Cauchy laws. The tail exponent is defined as

$$\gamma = \lim_{u\to 1}(u-1)\left(\log\left[f\left\{Q(u)\right\}\right]\right)' = \lim_{u\to 1}(1-u)\left[\log\{q(u)\}\right]',$$

where $[\log\{f(x)\}]'$ is the score of the underlying probability law and γ characterizes (i) short-, (ii) exponential- and (iii) long tails for (i) $\gamma < 1$, (ii) $\gamma = 1$ and (iii) $\gamma > 1$ respectively. Furthermore, let $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha)$ be the marginal c-quantile, i.e.,

$$C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) = C^{-1}_{U_k|U_{\not\ell}=u_{\not\ell}}(\alpha)\Big|_{u_1=1,\ldots,u_{\ell-1}=1,u_{\ell+1}=1,\ldots,u_d=1},$$

and let $\mathcal{S}^{u_\ell}_{k\ell}$ be the associated contagion measure. The notation $z(x) \sim y(x)$, $x \to a$ means $\lim_{x\to a} z(x)/y(x) = 1$. Bernard and Czado (2015) call X_k and X_ℓ conditionally independent in the right tail, if $Q_{X_k|X_\ell=x_\ell}(\alpha) \sim g(\alpha)$, $x_\ell \to \infty$, $\alpha \in (0,1)$, where the function $g(\alpha)$ is independent of x_ℓ. Asymptotic conditional tail independence generally describes a flat conditional quantile function. For example, the Gaussian copula shows conditional dependence but unconditional independence in both tails. Based on the introduced concepts, we summarize the limiting behavior of $\mathcal{S}^{u_\ell}_{k\ell}$ in the right tail area in the following statement.

Proposition 1. *Let X_k and X_ℓ have tail-monotone densities $f_k(x_k)$ and $f_\ell(x_\ell)$ with tail exponents γ_k and γ_ℓ.*

(a) If X_k and X_ℓ are conditionally dependent such that $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) \to 1$, $u_\ell \to 1$, with $\gamma_k \geq 1$ and $\gamma_\ell > 1$, then $\mathcal{S}^{u_\ell}_{k\ell} \to \frac{\gamma_k-1}{\gamma_\ell-1}$ as $u_\ell \to 1$.

(b) If X_k and X_ℓ are conditionally dependent such that $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) \to 1$, $u_\ell \to 1$, with $\gamma_k > 1$ and $\gamma_\ell = 1$, then $\mathcal{S}^{u_\ell}_{k\ell} \to \infty$ as $u_\ell \to 1$.

(c) If X_k and X_ℓ are conditionally independent in the right tail with $\gamma_k \geq 1$ and $\gamma_\ell \geq 1$, then $\mathcal{S}^{u_\ell}_{k\ell} \to 0$ as $u_\ell \to 1$.

This statement basically rules out contagion in the right tail area, if X_k and X_ℓ are conditionally tail independent and stresses the importance of conditional tail dependence for analyzing contagion in tail areas. Furthermore, if the asymptotic behavior of the marginal c-quantile can be described by $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) \to 1$, $u_\ell \to 1$, and the marginal distributions have long or exponential tails, the specific dependence between X_k and X_ℓ can be neglected in the limit. The contagion effect is dominated by the relation of the probability laws and independent of the level $\alpha \in (0,1)$ as $u_\ell \to 1$. To provide some intuition for heterogenous marginal cdfs and tail-dependence in the left tail area, consider the following bivariate example:

Suppose $X_k \sim \mathrm{N}(0,3)$ and $X_\ell \sim t_3$ with identical first and second moments and tail exponents $\gamma_k = 1$ and $\gamma_\ell = 4/3$. As shown in the left panel of Figure 2.1, $|Q_k(u)| < |Q_\ell(u)|$ and $q_k(u) < q_\ell(u)$ for a small u which is clear given the differences in the tails. In addition, let $\{F_k(X_k), F_\ell(X_\ell)\}^\top \sim C(u_k, u_\ell; \theta)$, where $C(u_k, u_\ell; \theta)$, $\theta = 2$, refers to the Clayton copula supporting dependence in the left tail area. As the Clayton copula is restricted to positive dependence and is exchangeable, i.e., $C(u_k, u_\ell) = C(u_\ell, u_k)$, Figure 2.2 reveals $0 < \frac{\partial}{\partial u_\ell} C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) = \frac{\partial}{\partial u_k} C^{-1}_{U_\ell|U_k=u_k}(\alpha)$ and $\alpha \geq C^{-1}_{U_\ell|U_k=u_k}(\alpha) = C^{-1}_{U_k|U_\ell=u_\ell}(\alpha)$ for the considered levels $\alpha = 0.0001$ and $\alpha = 0.5$. These properties can also be shown analytically. Combining assumptions about marginal cdfs and dependence implies that $Q_k(u_k)/Q_\ell\{C^{-1}_{U_\ell|U_k=u_k}(\alpha)\} < 1$ and $q_\ell\{C^{-1}_{U_\ell|U_k=u_k}(\alpha)\}/q_k(u_k) > 1$ in the left tail. As a result, $S^{u_k}_{\ell k} \to \infty$ as $u_k \to 0$ irrespective of α, which is indicated by the dashed lines in the right panel of Figure 2.1. Moreover, the solid lines in

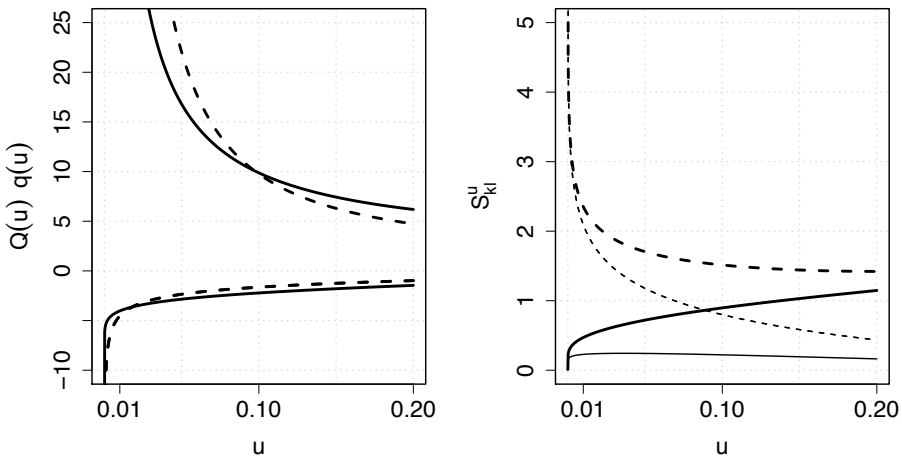

Figure 2.1: The left panel shows quantile function $Q(u)$ (bottom) and quantile density function $q(u)$ (top) for $N(0,3)$ (solid) and t_3 (dashed). The right panel presents $\mathcal{S}^{u_k}_{\ell k}$ (dashed) for $\alpha = 0.0001$ (thin) and $\alpha = 0.5$ (thick).

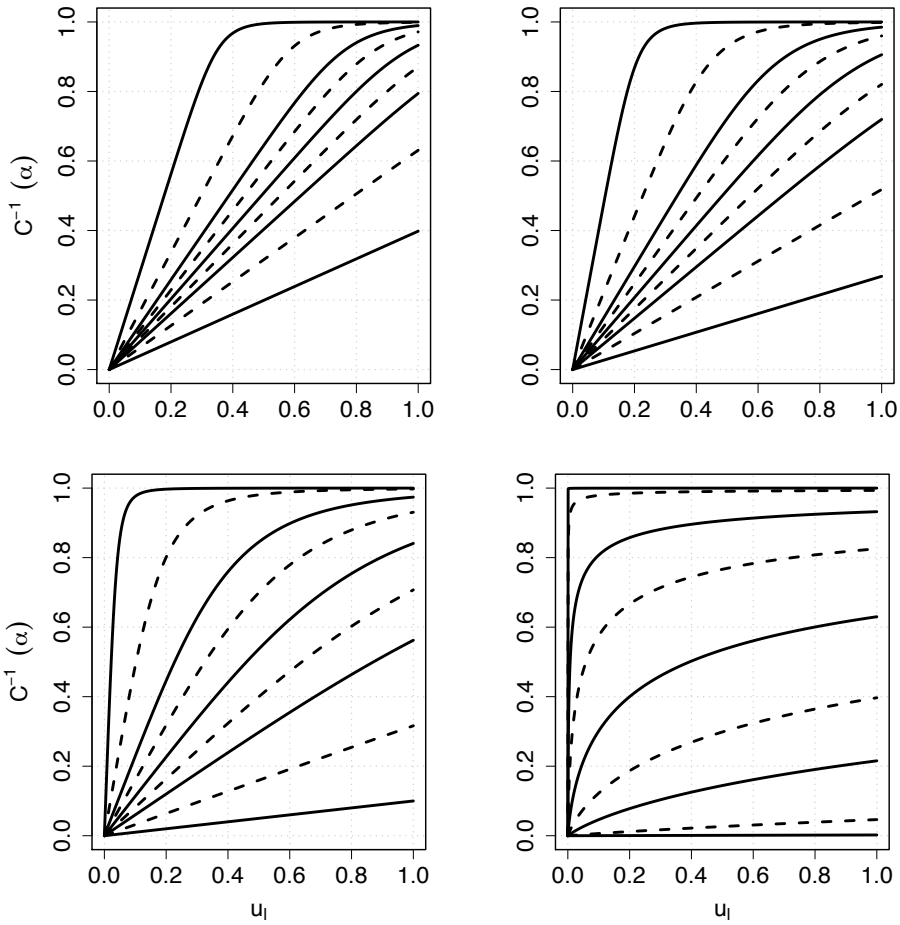

Figure 2.2: Conditional c-quantile functions $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha)$ for the bivariate Clayton copula. The alternating lines (solid and dashed) refer to $\alpha \in \{0.0001, 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99, 0.9999\}$ – bottom-up ordered. Upper panel illustrates the curves for $\theta \in \{9, 6\}$ and lower panel for $\theta \in \{3, 0.5\}$ respectively – left-right ordered.

the right panel of Figure 2.1 clearly illustrate that $\mathcal{S}_{k\ell}^{u_\ell} \to 0$ as $u_\ell \to 0$, although the copula supports dependence in the lower left tail area. This is due to the fast convergence of $q_\ell(u_\ell) \to \infty$ as $u_\ell \to 0$.

The economic interpretation of the example can be summarized as follows: Given a small value of α, a marginal change in x_k referring to X_k with low-risk, e.g., $X_k \sim \mathrm{N}(0,3)$, leads to a significant change in $Q_{X_\ell|U_k=\alpha}(\alpha)$ related to X_ℓ with high-risk, e.g., $X_\ell \sim t_3$. Yet, the reverse statement does not hold! An important implication for managing financial crises is as follows: An intensification of the distress of a low-risk financial market can amplify a financial crisis due to contagion; however low-risk markets are significantly less affected if the increased distress is in high-risk markets. All in all, this example emphasizes the importance of marginal probability laws and their relations to each other for studying contagion in tail areas.

2.2. Contagion from and to sub-portfolios

Deriving multilateral contagion measures is notional more tedious than deriving bivariate contagion measures, but relies on the same idea. To compactly formulate contagion effects from or to a set of risks, denote by $\mathcal{K}_\ell$ and $\mathcal{L}_k$ the sets of indices including all indices expect ℓ and k respectively, i.e., $\mathcal{K}_\ell = \{1,\dots,d\} \setminus \ell$ and $\mathcal{L}_k = \{1,\dots,d\} \setminus k$.

We firstly aim at exploring the simultaneous effect on all variables with index in $\mathcal{K}_\ell$, i.e., on X_1 or $\dots$ or $X_{\ell-1}$ or $X_{\ell+1}$ or $\dots$ or X_d by a marginal change in x_ℓ. The conditional independence of the events $\{X_k|X_{\not\ell} = x_{\not\ell}\}$, $k \in \mathcal{K}_\ell$, justifies building an aggregated function by adding up $Q_{X_k|X_{\not\ell}=x_{\not\ell}}(\alpha)$, $k \in \mathcal{K}_\ell$, which describes the aggregated effect on all risks transmitted by X_ℓ. The contagion effect is then obtained by differentiating and normalizing the corresponding aggregated risk measure. More formally, the derivative of the aggregated function at x_ℓ for $x_{\not\ell} = Q_{\not\ell}(\alpha)$ is given by

$$\frac{\partial}{\partial x_\ell} \sum_{k\in\mathcal{K}_\ell} Q_{X_k|X_{\not\ell}=x_{\not\ell}}(\alpha)\Bigg|_{x_{\not\ell}=Q_{\not\ell}(\alpha)} = \sum_{k\in\mathcal{K}_\ell} q^{\ell}_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha),$$

whose normalization with $Q_\ell(\alpha)/\sum_{k\in\mathcal{K}_\ell} Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)$ leads to the contagion measure

$$\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell} \stackrel{\text{def}}{=} \frac{\sum_{k\in\mathcal{K}_\ell} Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)\mathcal{S}^{\alpha}_{k\ell}}{\sum_{k\in\mathcal{K}_\ell} Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)}. \tag{8}$$

As the contagion measure (8) is a weighted average of bivariate contagion measures $\mathcal{S}^{\alpha}_{k\ell}$, diversification of risks is naturally incorporated. More precisely, contagion effects $\mathcal{S}^{\alpha}_{k\ell}$ to a sensitive risk X_k, i.e., to a risk with a sensitive risk measure $Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)$, contribute more to the aggregated risk of the entire sub-portfolio $\mathcal{K}_\ell$. Moreover, (8) is a signed elasticity and shares the same interpretation as partial elasticities $\mathcal{S}^{\alpha}_{k\ell}$, i.e., $\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell}$ describes the marginal effect on the aggregated function $\sum_{k\in\mathcal{K}_\ell} Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)$ by a marginal change in x_ℓ. For instance, a p%-change in x_ℓ causes a $(\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell} \cdot p)$%-change in the aggregated risk measure.

Secondly, consider the marginal effect on risk measure $Q_{X_k|U_{\not\ell}=\boldsymbol{\alpha}}(\alpha)$ by a simultaneous marginal-change in all x_ℓ with $\ell \in \mathcal{L}_k$. Let $v = Q_{\not\ell}(\alpha)/\|Q_{\not\ell}(\alpha)\|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm with

$\|v\|_2 = 1$. Then, the directional derivative of the function $Q_{X_k|X_{\not k}=x_{\not k}}(\alpha)$ at $x_{\not k}$ along v is given by

$$\nabla_{x_{\not k}} Q_{X_k|X_{\not k}=x_{\not k}}(\alpha)\Big|_{x_{\not k}=v} = \frac{1}{\|Q_{\not k}(\alpha)\|_2} \sum_{\ell \in \mathcal{L}_k} Q_\ell(\alpha) q^{\ell}_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha). \tag{9}$$

Normalizing (9) by $1/Q_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha)$ results in a signed elasticity, which however does not straightforwardly permit classifying risk X_k as stable or fragile when compared with 1. Simply speaking, this is due to the fact that a change in each component of the vector $x_{\not k}$ ("explanatory variables") cannot be compared with a change in a scalar $Q_{X_k|U_{\not k}=u_{\not k}}(\alpha)$ ("dependent variable"). Note, however, that this problem does not arise for the cases discussed above, see Equation 8, since the risk measure is a scalar valued function and a scalar variable is marginally changed. Let $\mathrm{p}_{\not k} = (p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_d)^\top$ collect the considered p_ℓ%-change in each component of $x_{\not k}$. Define $\|y\| = (\sum_{j=1}^q y_j^q)^{1/q}$, where q denotes the number of components in y. Note that $\|y\|$ takes implicitly the size of y into account. Then, normalization of (9) with $\{\|\mathrm{p}_{\not k}\| Q_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha)\}^{-1}$ gives the contagion measure

$$\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k} \stackrel{\text{def}}{=} \frac{1}{\|\mathrm{p}_{\not k}\| \|Q_{\not k}(\alpha)\|_2} \sum_{\ell \in \mathcal{L}_k} \mathcal{S}^{\alpha}_{k\ell}. \tag{10}$$

Given a simultaneous 1%-change in each x_ℓ, i.e., $p_\ell = 1$, $\ell \in \mathcal{L}_k$, the risk measure $Q_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha)$ changes approximately by $(\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k})$%. Due to the normalization with $\|\mathrm{p}_{\not k}\|$ in (10), the contagion measure $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k}$ allows categorizing risk X_k as robust or stable if $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k} < 1$ and sensitive if $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k} > 1$. Note that especially $\|\mathrm{p}_{\not k}\| = 1$ for $d = \infty$, which reflects minor importance of each component in v as $d \to \infty$.

In the spirit of Adrian and Brunnermeier (2011), $\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell}$ and $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k}$ have the following economic interpretation: $\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell}$ measures the pollution of a financial institution X_ℓ to the financial system in distress, i.e., each component $X_{\not k}$ is at its VaR and thus, takes values $X_{\not k} = Q_{\not k}(\alpha)$. Likewise, $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k}$ describes the extent institution X_k is affected in case of a systemic event. A simultaneous change in $x_{\not k}$ can also be interpreted as "joint" shock affecting several risks simultaneously, which relates our concept to factor-based models from credit risk analysis. Note that both presented measures $\mathcal{S}^{\alpha}_{k\leftarrow\mathcal{L}_k}$ and $\mathcal{S}^{\alpha}_{\mathcal{K}_\ell\leftarrow\ell}$ are special cases of a general contagion measure $\mathcal{S}^{\alpha}_{\mathcal{K}\leftarrow\mathcal{L}}$ describing the effect on sub-portfolio $\mathcal{K}$ by a simultaneous change in each component of sub-portfolio $\mathcal{L}$, with $\mathcal{K} \cap \mathcal{L} = \emptyset$.

2.3. Systemic risk

Systemic risk of a portfolio is endogenous by construction, as each included risk contributes to and is affected by systemic risk. As illustrated in the introduction, there are several ways to define systemic risk. Our definition presented below is derived from contagion effects, for which reason we call it conditional systemic risk. The idea is relatively simple: we build all possible "leave-one-out" portfolios and check, whether the "left-out-risk" pollutes the sub-portfolio or is polluted by a simultaneous change in the components of the respective sub-portfolio. Normalization of the aggregated contagion effects yields the result. Our approach characterizes precisely that component of systemic risk arising from contagion in the tail areas. This contagion-based definition is in line with other definitions of systemic risk, e.g., Diebold and Yilmaz (2014), which do not distinguish between conditional and unconditional

systemic risk though.

Formally, we construct an (endogenous) aggregated risk measure for the entire portfolio of risks as

$$\sum_{\ell=1}^{d} \sum_{k_\ell \in \mathcal{K}_\ell} Q_{X_{k_\ell}|X_{\not{\ell}_\ell}=x_{\not{\ell}_\ell}}(\alpha), \tag{11}$$

similar to that for measuring multilateral contagion. Let p be a d-dimensional vector of 1's. Defining $Q(\alpha) = \{Q_1(\alpha), \ldots, Q_d(\alpha)\}^\top$ as well as $v = Q(\alpha)/\|Q(\alpha)\|_2$ and differentiating (11) at x along v produces

$$\nabla_x \sum_{\ell=1}^{d} \sum_{k_\ell \in \mathcal{K}_\ell} Q_{X_{k_\ell}|X_{\not{\ell}_\ell}=x_{\not{\ell}_\ell}}(\alpha)\Bigg|_{x=v} = \frac{1}{\|Q(\alpha)\|_2} \sum_{\ell=1}^{d} \sum_{k_\ell \in \mathcal{K}_\ell} Q_\ell(\alpha) q^{\ell}_{X_{k_\ell}|U_{\not{\ell}_\ell}=\boldsymbol{\alpha}}(\alpha). \tag{12}$$

Using $q^{k_\ell}_{X_{k_\ell}|U_{\not{\ell}_\ell}=\boldsymbol{\alpha}}(\alpha) = 0$ for all k_ℓ, normalization of (12) leads to our notion of systemic risk

$$\mathcal{S}^\alpha \stackrel{\text{def}}{=} \frac{1}{\|\,\mathrm{p}\,\|\|Q(\alpha)\|_2} \frac{\sum_{k,\ell=1}^{d} Q_{X_k|U_{\not{k}}=\boldsymbol{\alpha}}(\alpha)\mathcal{S}^\alpha_{k\ell}}{(d-1)\sum_{k=1}^{d} Q_{X_k|U_{\not{k}}=\boldsymbol{\alpha}}(\alpha)}, \tag{13}$$

where the normalization with $\|\,\mathrm{p}\,\|$ is for the same reason as above. The proposed conditional systemic risk measure is also a weighted average of bivariate contagion measures $\mathcal{S}^\alpha_{k\ell}$. Despite the underlying endogeneity, $\mathcal{S}^\alpha$ shares the interpretation of an elasticity in terms of categorizing a portfolio of risks as stable if $\mathcal{S}^\alpha \approx 0$ and fragile if $\mathcal{S}^\alpha > 1$.

3. Estimation details

In our empirical application below, we impose a structure on the conditional quantiles that supports the modeling of (conditional) tail dependence, as pure non-parametric estimation of introduced contagion and systemic risk measures is accompanied by an inflated variance. Recently, dozens of different copula models with their weak and strong points have been proposed in the literature. Worth mentioning for modeling higher dimensional random vectors are vines, c.f., Bedford and Cooke (2001, 2002); Kurowicka and Joe (2011), factor copulas, see Oh and Patton (2014), and HAC. As shown in Okhrin, Okhrin, and Schmid (2013b); Okhrin *et al.* (2013c), HAC are flexible enough to capture the nature of dependence in financial data and they also support the modeling of tail dependence in a diversified way. Therefore, in this paper we concentrate only on this type of copulas which allow for a more flexible and intuitive dependence structure in comparison to simple Archimedean copulas, and need a smaller number of parameters compared to elliptical copulas (Okhrin *et al.* 2013b). In particular the latter is an important property as it comes to modeling the dependence between a high number of financial time series. In this section, we formally introduce HAC and then discuss the penalized estimation problem with the asymptotic properties of the estimator.

3.1. Hierarchical Archimedean copula

HAC generalize Archimedean copulae, where in the latter arguments are exchangeable, making them inappropriate for measuring scale free dependence for a large dimensional vector X. HAC, however, are recursively built by substituting arbitrary marginal distributions of an Archimedean copula by a further Archimedean copula. This procedure can be subsequently applied for structuring the dependence between the random variables. For instance, Hofert and Scherer (2011) and Choroś-Tomczyk, Härdle, and Okhrin (2013) motivate the structure by clustering economic sectors and Okhrin, Odening, and Xu (2013a) by geographical location. HAC are interesting from a statistical perspective, as the induced dependencies are non-elliptical, non-exchangeable and allow for modeling joint extreme events. Further examples for applications of HAC in quantitative finance and risk management can be found in Savu and Trede (2010) and Härdle, Okhrin, and Okhrin (2013). Beyond that, Hering, Hofert, Mai, and Scherer (2010) induce sub-group specific dependencies via Lévy subordinators and Härdle, Okhrin, and Wang (2015) discuss time-variations, where the structure depends on the hidden state of a Markov chain.

Formally, HAC rely on generator functions $\phi \in \mathfrak{L} = \{\phi : [0;\infty) \to [0,1] \,|\, \phi(0) = 1, \phi(\infty) = 0; (-1)^k\phi^{(k)} \geq 0; k \in \mathbb{N}\}$ and its non-decreasing and convex inverse $(-1)^k\phi^{(k)}(x)$, $x \in [0,\infty)$. For instance, a 4-dimensional Archimedean copula is given by $\phi\{\phi^{-1}(u_1)+\phi^{-1}(u_2)+\phi^{-1}(u_3)+\phi^{-1}(u_4)\}$. As shown in Figure 3.1, inducing a binary structure leads already to a variety of possible HAC, e.g.,

$$C_k\{C_{\ell_1}(u_1,u_2), C_{\ell_2}(u_3,u_4)\} = \phi_k\left[\phi_k^{-1}[\phi_{\ell_1}\{\phi_{\ell_1}^{-1}(u_1)+\phi_{\ell_1}^{-1}(u_2)\}] + \phi_k^{-1}[\phi_{\ell_2}\{\phi_{\ell_2}^{-1}(u_3)+\phi_{\ell_2}^{-1}(u_4)\}]\right],$$

where ϕ_k denotes the generator at the higher hierarchical level and ϕ_{ℓ_j} the generator of the lower nesting level, $j = 1, 2$. Let d_ℓ sub-copulae be rooted at hierarchical level ℓ. If the nesting conditions – (i) $\phi_k, \phi_{\ell_j} \in \mathfrak{L}$ and (ii) $\phi_k^{-1} \circ \phi_{\ell_j}$ having completely monotone derivatives – are satisfied for $j = 1, \ldots, d_\ell$, HAC are properly defined distribution functions according to McNeil (2008). Furthermore, all r-variate marginal distribution functions are HAC, $r \leq d$, which only depend on generators at lower hierarchical levels, see Okhrin *et al.* (2013c). Weaker conditions on the decomposition of generators are stated in Rezapour (2015), but not necessary for our purpose, as we restrict the following discussion to single parameter families of ϕ_{θ_k} and do not allow mixtures of them within one HAC. This restriction simplifies the "hard to check" nesting condition to the parameter ordering $\theta_k \leq \theta_{\ell_j}$, $j = 1, \ldots, d_\ell$, for most parametric families such as Clayton and Gumbel, see Hofert (2011). Even though d_ℓ sub-copulae are rooted at hierarchical level ℓ, only the smallest parameter θ_{ℓ_j}, $j = 1, \ldots, d_\ell$, determines the feasible parameter space of θ_k presented below. Hence, we define $\theta_\ell = \min[\theta_{\ell_j} : j \in \{1, \ldots, d_\ell\}]$, such that $\theta_k \leq \theta_\ell$ holds for all ℓ_j, $j = 1, \ldots, d_\ell$. Furthermore, the parameter at the respective higher nesting level is denoted by $\theta_{k(\ell)}$ in order to emphasize the relation to θ_ℓ.

If the structure of a HAC is not determined by the application, it has to be estimated from data. Having similar dependence between the involved random variables, Okhrin *et al.* (2013b) and Okhrin and Ristig (2014) suggest recursively aggregating a binary tree structure to more complex structures involving nodes with different number of branches. This procedure is reasoned by the associativity property of Archimedean copula, see Nelsen (2006, Theorem 4.1.5), but not statistically studied. However, Okhrin *et al.* (2013b) point out that even a misspecified structure can be interpreted as minimizer of the Kullback-Leibler divergence in the spirit of White (1994). A non-parametric –yet computationally demanding– method for re-building the structure from data is proposed by Segers

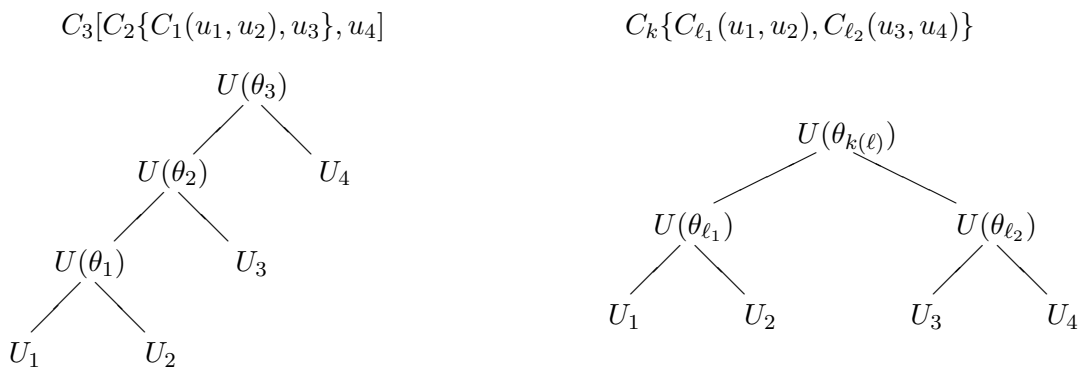

Figure 3.1: Example of a four-dimensional binary and partially nested HAC.

and Uyttendaele (2014).

Górecki, Hofert, and Holeňa (2014) build on the grouping techniques for binary trees of Okhrin *et al.* (2013b) and propose recovering binary structures from pairwise Kendall's correlation coefficients. For this purpose, they introduce pseudo variables based on the diagonal of Archimedean copula, $U_{\ell_j} \stackrel{\text{def}}{=} U(\theta_{\ell_j}) = \phi_{\theta_{\ell_j}}[d_{\ell_j}\phi^{-1}_{\theta_{\ell_j}}\{\max(U_1,\dots,U_{d_{\ell_j}})\}]$, with $U_{\ell_j} \sim \mathrm{U}(0,1)$, $j = 1,\dots,d_{\ell_j}$, and show that U_{ℓ_j} follows a standard uniform distribution as well. Despite arising statistical inefficiencies, this transformation leads to a computationally more tractable procedure than the stage-wise ML estimation method of Okhrin *et al.* (2013b), as the tedious derivation of the HAC's density is not needed. Instead the density of a bivariate Archimedean copula can be used for estimating the parameter at each stage of the procedure to recover the structure of a binary HAC.

Consider the simultaneous estimation of the structure and parameters of $C_k\{C_{\ell_1}(u_1,u_2),C_{\ell_2}(u_3,u_4)\}$: At the first stage of the multi-stage procedure, the dependence parameters for all possible pairs of variables $(U_1.U_2)$, $(U_1.U_3)$, $(U_2.U_3)$, ..., are estimated. Given that $\hat{\theta}_{\ell_1}$ estimated from $(U_1.U_2)$ leads to the strongest fit, the variables U_1 and U_2 are removed from the set of variables and the pseudo-variable $U_{\ell_1} = \phi_{\hat{\theta}_{\ell_1}}[2\phi^{-1}_{\hat{\theta}_{\ell_1}}\{\max(U_1,U_2)\}]$ is added. At the next step, the dependence parameters for all remaining pairs of variables $(U_{\ell_1}.U_2)$, $(U_{\ell_1}.U_3)$, $(U_{\ell_1}.U_4)$ and $(U_3.U_4)$ are estimated. Given that $\hat{\theta}_{\ell_2}$ estimated from $(U_3.U_4)$ leads to the strongest fit, $U_{\ell_2} = \phi_{\hat{\theta}_{\ell_2}}[2\phi^{-1}_{\hat{\theta}_{\ell_2}}\{\max(U_3,U_4)\}]$ is computed and the set of variables is accordingly modified. At the final step, the estimation of the parameter at the root simplifies to the estimation of $\theta_{k(\ell)}$ from the binary Archimedean copula $\phi_{\theta_{k(\ell)}}\{\phi^{-1}_{\theta_{k(\ell)}}(U_{\ell_1}) + \phi^{-1}_{\theta_{k(\ell)}}(U_{\ell_2})\}$. The recursive estimation procedure itself reduces the computational costs enormously, but also the transformation based on the diagonal of Archimedean copula plays a key role. The fact that $U_{\ell_1} \sim \mathrm{U}(0,1)$ and $U_{\ell_2} \sim \mathrm{U}(0,1)$ allows the estimation of $\theta_{k(\ell)}$ using simple ML estimation for binary Archimedean copula irrespective of lower hierarchical levels. Hence, considering the estimation of the parameter at the k-th nesting level is absolutely sufficient for our purpose.

A more general example of a partially nested HAC is given in Figure 3.2. Let $\mathcal{O}_p(1)$ refer to a random variable converging to zero in probability as $n \to \infty$. Knowing the general structure of the HAC and the parameters $\theta_1,\dots,\theta_{d_\ell}$, the parameter θ_ℓ can be consistently estimated like the parameter of a d_ℓ-dimensional Archimedean copula, as $U_1 \stackrel{\text{def}}{=} U(\theta_1),\dots,U_{d_\ell} \stackrel{\text{def}}{=} U(\theta_{d_\ell})$ are uniformly distributed. Given the consistent estimate $\hat{\theta}_\ell$, the parameter $\theta_{k(\ell)}$ can be consistently estimated, as $U_5,\dots,U_d$ and

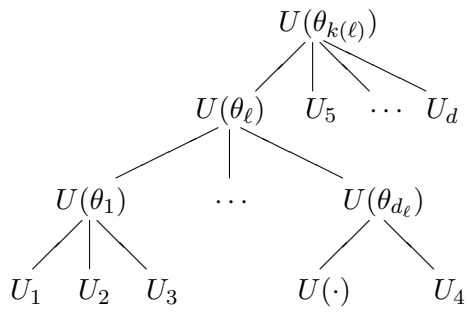

Figure 3.2: Example of a partially nested HAC.

$U(\hat{\theta}_\ell) = U(\theta_\ell) + \mathcal{O}_p(1)$ are uniformly distributed as $n \to \infty$.

As mentioned above, the sketched procedure recovers only binary structures, but not those given in Figure 3.2. Yet, non-binary HAC have clear advantages compared to their binary counterparts. In particular, they are easier to interpret, as less nodes and parameters are involved, and the parameters can be more efficiently estimated, if the true structure is non-binary or even known. Okhrin *et al.* (2013b) propose non-binary structures by joining two subsequent nodes, if $\hat{\theta}_\ell - \hat{\theta}_{k(\ell)} \leq \epsilon$, for a pre-specified parameter ϵ. There are, however, various difficulties in the selection of ϵ. This motivates developing a data-driven method to determine ϵ_n. The parameter ϵ_n results from an implicit penalization of the structure in a penalized ML setting and has an aesthetic statistical interpretation. In particular, a non-concave penalty is imposed on the parameter difference $(\theta_\ell - \theta_{k(\ell)})$. For that reason, we firstly formulate the estimation problem, state secondly the asymptotic properties of the penalized estimator and thirdly, derive ϵ_n.

3.2. Penalized estimation of HAC

Let the univariate marginal distribution functions of the underlying d-dimensional HAC be known. Based on the sketched multi-stage estimation procedures, $\theta_{k(\ell)}$ is the only parameter to be estimated at the k-th nesting level, as parameters from lower nesting levels are estimated in previous estimation stages. In particular, the transformation based on the diagonal of Archimedean copula produces (asymptotically) standard uniformly distributed random variables, so that the estimation problem at the k-th hierarchical level can be traced back to the estimation of the parameter of a d_k-dimensional Archimedean copula, where d_k refers to the number of (pseudo) variables of that respective level.

To emphasize the flexibility of HAC, more than $2.8 \cdot 10^8$ possible structures are available in dimension $d = 10$. Addressing the question of an optimal structure, the huge amount of possible structures makes the calibration of all specifications and subsequent model selection infeasible in practice. As alternative to model selection, two subsequent parameters could also be tested for being equal, but the more tests have to be sequentially conducted, the more demanding is the asymptotic analysis of the estimator. To overcome those complications, while reducing the number of *different* parameters in the model and re-covering the structure optimally, the estimated parameter is shrunken as explained below. In general, we build the procedure on the seminal work of Fan and Li (2001) and suggest

determining/aggregating the structure as well as estimating the involved parameters simultaneously. Shrinkage estimators are popular for simultaneous parameter estimation and model selection in linear and generalized linear models, see Tibshirani (1996), Fan and Li (2001) and Tibshirani (2011), but they are rarely applied to non-linear likelihood-based estimation problems. While the dimension of the dependent variable is typically fixed, Fan and Peng (2004) also consider a diverging number of parameters linked to the number of explanatory variables. We do not discuss a diverging number of parameters, as by definition, the dimension of d-dimensional HAC are connected to at most $(d-1)$ parameters. The advantages and disadvantages of several penalty functions are comprehensively reviewed in Fan and Lv (2010). As the smoothly clipped absolute deviation (SCAD) penalty function, $p_\lambda(|\cdot|)$, has superior properties in comparison with other penalties we restrict the following discussion to the SCAD penalty, whose first derivative is of the form

$$p'_\lambda(|\gamma|) = \lambda\left[\mathbf{I}\{|\gamma| \le \lambda\} + \frac{\max(a\lambda - |\gamma|, 0)}{\lambda(a-1)}\mathbf{I}\{|\gamma| > \lambda\}\right],$$

with $a > 2$. In general, an appropriate penalty function should be singular at the origin, i.e., $\liminf_{n\to\infty}\liminf_{\gamma\to 0+} p'_{\lambda_n}(\gamma)/\lambda_n > 0$, so that the penalized estimator is a thresholding rule.

Let the sequence of random vectors $\{U_i\}_{i=1}^n$ be independent copies with $U_i = \{U_{i1}, \ldots, U_{i\ell_1}(\hat\theta_{\ell_1}), U_{i\ell_{d_\ell}}(\hat\theta_{\ell_{d_\ell}}), \ldots, U_{d_k}\}^\top$ and $U_{ik} \sim \mathrm{U}(0,1)$, $k = 1, \ldots, d_k$. Note that U_i might include pseudo-variables and let each U_i have an identical parametric density function of a d_k-dimensional Archimedean copula $C(u_1, \ldots, u_{d_k}; \theta_{k(\ell)})$ in the family $\{c(u_1, \ldots, u_{d_k}; \theta_{k(\ell)}) : \theta_{k(\ell)} \in \Theta_{k(\ell)} \subseteq \mathbb{R}\}$, where $c(\cdot)$ is the copula density $c(u_1, \ldots, u_{d_k}; \cdot) = \frac{\partial^{d_k}}{\partial u_1, \ldots, \partial u_{d_k}} C(u_1, \ldots, u_{d_k}; \cdot)$. The corresponding log-likelihood contributions are denoted by $\ell_i(\theta_{k(\ell)}) = \log\{c(U_{i1}, \ldots, U_{id_k}; \theta_{k(\ell)})\}$, whose regularity assumptions are listed in Appendix A. These are in line with those in Fan and Li (2001) and Cai and Wang (2014). Given the SCAD penalty and the contributions $\ell_i(\theta_{k(\ell)})$, the penalized log-likelihood at the k-th hierarchical level is given by

$$\mathcal{Q}(\theta_\ell, \theta_{k(\ell)}) = \sum_{i=1}^n \ell_i(\theta_{k(\ell)}) - n p_{\lambda_n}(\theta_\ell - \theta_{k(\ell)}), \tag{14}$$

where $(\theta_\ell - \theta_{k(\ell)})$ is non-negative by construction, as θ_ℓ shortens the feasible parameter space of $\theta_{k(\ell)}$ to $\Theta_{k(\ell)} \setminus (\theta_\ell, \infty)$. Given a consistent estimator of θ_ℓ from the previous estimation stage, denoted by $\hat\theta_\ell$, the objective function is defined as $\mathcal{Q}_{\hat\theta_\ell}(\theta_{k(\ell)}) \stackrel{\text{def}}{=} \mathcal{Q}(\hat\theta_\ell, \theta_{k(\ell)})$ and the penalized estimator is given by $\hat\theta^{\lambda_n}_{k(\ell)} = \arg\max \mathcal{Q}_{\hat\theta_\ell}(\theta_{k(\ell)})$. Similar to Fan and Li (2001, Theorem 2), the sparsity and oracle property of the penalized estimator $\hat\theta^{\lambda_n}_{k(\ell)}$ are summarized in Proposition 2 and 3 respectively. Denote by $\theta_{\ell,0}$ and $\theta_{k(\ell),0}$ the true parameters and assume that $\theta_{\ell,0} = \theta_{k(\ell),0}$, namely two parameters on subsequent levels are equal.

Proposition 2. *Let $\{U_i\}_{i=1}^n$ be independent with log-density $\ell_i(\theta_{k(\ell)})$ for which Assumptions 1-3 hold. If $n^{1/2}\lambda_n \to \infty$ as $n \to \infty$, then $\lim_{n\to\infty} \mathrm{P}(\hat\theta^{\lambda_n}_{k(\ell)} = \theta_{\ell,0}) = 1$.*

Proposition 2 shows that no k-th hierarchical level is added with probability tending to one, if the true model is parsimonious. In other words, the structure is automatically aggregated, if $\hat\theta^{\lambda_n}_{k(\ell)} =$

$\hat{\theta}_\ell$, which has similarities with the fused LASSO proposed in Tibshirani, Saunders, Rosset, Zhu, and Knight (2005). If $\theta_{k(\ell)}$ is the parameter at the lowest nesting level, there exists no parameter θ_ℓ and the parameter $\theta_{k(\ell)}$ is consequently not penalized. This makes it simple to establish the consistency of $\hat{\theta}_\ell$ as shown in Lemma 2, see Appendix B. Now, change the perspective and assume that $\theta_{\ell,0} > \theta_{k(\ell),0}$, namely parameters on two subsequent nodes are truly different. Denote by $\widehat{\mathcal{I}}(\theta_{k(\ell)}) = -n^{-1}\sum_{i=1}^n \ell_i''(\theta_{k(\ell)})$ an estimator of the information matrix $\mathcal{I}(\theta_{k(\ell)})$ defined in Assumption 1, c.f., Appendix A.

Proposition 3. *Let $\{U_i\}_{i=1}^n$ be independent with log-density $\ell_i(\theta_{k(\ell)})$ for which Assumptions 1-3 hold. If $\lambda_n \to 0$ as $n \to \infty$, then*

$$n^{1/2}\left\{\widehat{\mathcal{I}}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_{\ell,0} - \theta_{k(\ell),0})\right\}\left[(\hat{\theta}^{\lambda_n}_{k(\ell)} - \theta_{k(\ell),0}) - \left\{\widehat{\mathcal{I}}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_{\ell,0} - \theta_{k(\ell),0})\right\}^{-1} p'_{\lambda_n}(\theta_{\ell,0} - \theta_{k(\ell),0})\right] \xrightarrow{\mathcal{L}} \mathrm{N}\{0, \mathcal{I}(\theta_{k(\ell),0})\}.$$

As a consequence of Proposition 3, the asymptotic covariance of $n^{1/2}\hat{\theta}^{\lambda_n}_{k(\ell)}$ can be reasonably approximated by $\widehat{\mathcal{I}}(\hat{\theta}^{\lambda_n}_{k(\ell)})$, if $\lambda_n \to 0$ as $n \to \infty$. Under this convergence of λ_n, the estimator $\theta^{\lambda_n}_{k(\ell)}$ enjoys the so called oracle property, i.e., the quality of $\theta^{\lambda_n}_{k(\ell)}$ is as good as if the structure of the HAC was known in advance. It can be straightforwardly deduced from subsequently applying Proposition 2 and 3 that the stage-wise estimation of HAC, as sketched above, recovers the true structure with probability tending to one for $n^{1/2}\lambda_n \to \infty$ as $n \to \infty$ and that the estimators are $n^{1/2}$-consistent for $\lambda_n \to 0$ as $n \to \infty$.

3.3. Attaining sparsity

Even though the maximization of $\mathcal{Q}_{\hat{\theta}_\ell}(\theta_{k(\ell)})$ is an univariate numerical optimization problem, it is a challenging task due to the singularity of the penalty $p_\lambda(\cdot)$ at the origin. Based on similar ideas as presented in Zou and Li (2008), who provide a comprehensive discussion on maximizing non-concave penalized log-likelihood functions, Proposition 4 yields an appealing formula for the penalized estimator $\hat{\theta}^{\lambda_n}_{k(\ell)}$. Denote by $\hat{\theta}_{k(\ell)}$ the ML estimator for $\theta_{k(\ell)}$.

Proposition 4. *Let $\{U_i\}_{i=1}^n$ be independent with log-density $\ell_i(\theta_{k(\ell)})$ for which Assumptions 1-3 hold. Then, $\hat{\theta}^{\lambda_n}_{k(\ell)} = \hat{\theta}_{k(\ell)} + \epsilon_n$, with $\epsilon_n \stackrel{\text{def}}{=} \epsilon(\lambda_n, a_n) = \widehat{\mathcal{I}}(\hat{\theta}_{k(\ell)})^{-1} p'_{\lambda_n}(\hat{\theta}_\ell - \hat{\theta}_{k(\ell)})$.*

Proposition 4 shows that the penalized estimator $\hat{\theta}^{\lambda_n}_{k(\ell)}$ can be expressed as sum of the ML estimator $\hat{\theta}_{k(\ell)}$ and a penalty term presented by the data-driven parameter $\epsilon_n \geq 0$. The parameter ϵ_n is a trade-off between the variability of $n^{1/2}\hat{\theta}_{k(\ell)}$ and the strength of the imposed penalty. The estimator $\hat{\theta}^{\lambda_n}_{k(\ell)}$ deviates enormously from the ML estimator, if the flatness of the log-likelihood decreases $\widehat{\mathcal{I}}(\hat{\theta}_{k(\ell)})$ and ϵ_n; but an increase in the distance between the ML estimates decreases ϵ_n, i.e., $p'_{\lambda_n}(\hat{\theta}_\ell - \hat{\theta}_{k(\ell)})$ is large, as $\lim_{\gamma\to\infty} p'_\lambda(\gamma) = 0^+$.

While Zou and Li (2008) apply the least angle regression (LARS) algorithm, see Efron, Hastie, Johnstone, and Tibshirani (2004), to attain sparsity, we apply the thresholding rule

$$\hat{\theta}_{k(\ell)} = \hat{\theta}_\ell, \quad \text{if} \quad \hat{\theta}_\ell - \hat{\theta}_{k(\ell)} \leq \epsilon_n, \tag{15}$$

at each estimation-stage to obtain a sparse structure. The proposed thresholding rule is elementarily related to the underlying copula, as the structure is aggregated, if $\hat{\theta}_{k(\ell)}^{\lambda_n} \geq \hat{\theta}_\ell$. In other words, the positive bias of the penalized estimator, c.f., Proposition 3, causes a violation of the required nesting condition, so that the estimated HAC fails to be a well defined distribution function. Furthermore, the thresholding rule avoids the a priori specification of a small ϵ, criticized by Zou and Li (2008).

The final task for conducting this penalized multi-stage estimation procedure is an appropriate selection of $(\lambda, a)^\top$. The simultaneous estimation and aggregation of the structure needs a *different* ϵ_n at each estimation stage for the following reasons: (i) While the parameters at lower hierarchical levels are estimated, the structure at higher hierarchical levels is unknown. Therefore, the information from higher nesting levels cannot be taken into account, when ϵ_n is fitted. (ii) Since HAC allow several nodes at the same hierarchical level, two or more sub-structures might be simultaneously built. Obviously, taking the same ϵ_n can be quite misleading. (iii) The sub-structure at higher nesting levels is by construction more complex than the sub-structure at lower levels. A more complex structure, however, should intuitively be stronger penalized in terms of $(\lambda, a)^\top$, as the structure is built with more parameters.

Wang, Li, and Tsai (2007) suggest selecting $(\lambda_n, a_n)^\top$ by minimizing the BIC, see Schwarz (1978), for linear and partially linear models. They, furthermore, show superior asymptotic properties compared with tuning parameters chosen in another optimal way. In particular, the true model is consistently identified, if $(\lambda_n, a_n)^\top$ minimizes the BIC. We basically transfer this idea to our highly non-linear likelihood-based model. However, the asymptotic properties of the tuning parameters themselves are not discussed and are beyond the scope of this paper. The tuning parameter are optimally determined by

$$(\lambda_n, a_n)^\top = \arg \max_{(\lambda,a)^\top} 2 \sum_{i=1}^{n} \ell_i \left\{ \hat{\theta}_{k(\ell)} + \epsilon(\lambda, a) \right\} - q_k \log(n), \tag{16}$$

where q_k denotes the involved number of parameters up to the k-th hierarchical level. Selecting the tuning parameters according to (16) penalizes parameters at higher nesting levels automatically stronger, as the number of parameters q_k is part of the BIC.

4. Simulation Study

The simulation study relies on m Monte Carlo replications, which are needed to estimate 1000 structures correctly. This is due to the fact that only estimates stemming from the same structure can be compared and used to compute summary statistics. In order to compare the results of the simulation study among Archimedean families (Clayton, Frank, Gumbel, Joe), let $\tau : \Theta_{k(\ell)} \to [0, 1]$ transform the parameter $\theta_{k(\ell)}$ into Kendall's correlation coefficient, see Joe (1997); Nelsen (2006). We illustrate the performance of the estimation procedure for two types of models: (i) 5-dimensional HAC with $((\text{U3.U4.U5})_{\theta_\ell}.\text{U1.U2})_{\theta_{k(\ell)}}$, where $\theta_\ell = \tau^{-1}(0.7)$ and $\theta_{k(\ell)} = \tau^{-1}(0.3)$ refer to the group specific dependence parameter of the random vectors $(U_3, U_4, U_5)^\top$ and (U_ℓ, U_1, U_2) respectively, where $U_\ell = \phi_{\theta_\ell}[3\phi_{\theta_\ell}^{-1}\{\max(U_3, U_4, U_5)\}]$. The parameters in terms of Kendall's $\tau(\cdot)$ are chosen as in Segers and Uyttendaele (2014, Section 8.1) for 4-variate HAC. The sample size is $n = 250$. (ii) As $\tau(\cdot)$

Family	$s(\hat{\theta}) = s(\theta_0)$	$\tau(\hat{\theta}_1)$ (sd)	$\tau(\hat{\theta}_2)$ (sd)	$\#\{\hat{\theta}\}$
Clayton	0.82	0.70 (0.01)	0.30 (0.02)	3.04
Frank	0.85	0.70 (0.01)	0.30 (0.02)	3.03
Gumbel	0.85	0.70 (0.01)	0.30 (0.02)	3.02
Joe	0.88	0.70 (0.01)	0.30 (0.02)	3.04

Table 4.1: $s(\hat{\theta}) = s(\theta_0)$ reports the fraction of correctly specified structures, $\tau(\hat{\theta}_k)$ (sd), $k = 1, 2$, refers to the sample average of Kendall's $\tau(\cdot)$ evaluated at the estimates and sd to the sample standard deviation thereof. If the structure is misspecified, $\#\{\hat{\theta}\}$ gives the number of parameters on average included in the misspecified HAC. Monte Carlo sample size is $n = 250$.

is a non-linear transformation, we investigate differences in the aggregation performance for different strength of dependence. In detail, we consider 3-dimensional HAC $((\mathrm{U2.U3})_{\theta_\ell}.\mathrm{U1})_{\theta_{k(\ell)}}$, where θ_ℓ refers to the dependence between $(U_2, U_3)^\top$ and $\theta_{k(\ell)}$ to the dependence of between $(U_\ell, U_1)^\top$, with $U_\ell = \phi_{\theta_\ell}[2\phi_{\theta_\ell}^{-1}\{\max(U_2, U_3)\}]$. The parameters are chosen such that $\theta_\ell = \tau^{-1}(\omega_\ell)$, with $\omega_\ell \in \{0.9, 0.7, 0.5, 0.3, 0.1\}$, and $\theta_{k(\ell)} = \tau^{-1}(\omega_{k(\ell)})$, with $k(\ell) = \ell, \ldots, 5$. The sample size is $n = 100$.

Table 4.1 summarizes the results of the 5-dimensional setup. The true structure is found in more than 82% of the cases and the parameters are unbiasedly estimated with a small empirical standard deviation. If the true structure is not identified, HAC are constructed from 3 parameters in most of the cases as shown in the last column of Table 4.1. Note that a correct classification of the structure requires several aggregation steps, which enlarges the room for mistakes. We would like to emphasize that the results are sensitive with respect to the selection of the tuning parameters $(\lambda, a)^\top$. In practice, (16) is computed by a global stochastic optimization algorithm namely simulated annealing. Our experiments have shown the longer the simulated annealing algorithm iterates the more precise are the estimation results with respect to a correct specification of the structure. However, more iterations make the entire procedure more computationally intensive.

The major findings of the 3-dimensional setting are presented in Table 4.2. In contrast to the previous simulation study, there is only one possible error source to obtain a misspecified structure. The penalized estimator is overall unbiased and the correct structure is detected in most of the cases, especially if the distance between $\tau(\theta_{k(\ell),0})$ and $\tau(\theta_{\ell,0})$ is large. Nevertheless, the low classification rate of the 3-dimensional Archimedean Clayton copula $(\mathrm{U1.U2.U3})_{\tau^{-1}(0.1)}$ should be mentioned as well as the bias of the estimator of the lower parameter of the Frank copula $((\mathrm{U2.U3})_{\tau^{-1}(0.9)}\mathrm{U1})_{\tau^{-1}(0.3)}$.

5. Systemic risk analysis of stock markets

In our empirical study, we apply the ideas on measuring financial contagion and conditional systemic risk to major stock indices of Australasia as well as the leading indices of Europe and the US. In particular we consider log-returns of index closing prices for the US Dow Jones Industrial Average Index (DJIA), the Euro STOXX 50 Index (SX5E), the Japanese Nikkei 225 Index (N225), the Shanghai Stock Exchange Composite Index (SSEC), the Australian All Ordinaries Index (XAO), the Singapore Stock Market Index (STI), the Korea Composite Stock Price Index (KOSPI), the Hong Kong Hang

		Clayton			Frank		
$\tau(\theta_{k(\ell),0})$	$\tau(\theta_{\ell,0})$	$s(\hat{\theta}) = s(\theta_0)$	$\tau(\hat{\theta}_{k(\ell)})$	$\tau(\hat{\theta}_\ell)$	$s(\hat{\theta}) = s(\theta_0)$	$\tau(\hat{\theta}_{k(\ell)})$	$\tau(\hat{\theta}_\ell)$
0.10	0.10	0.31	0.12	0.12	0.83	0.10	0.10
0.10	0.30	0.93	0.10	0.30	0.77	0.09	0.31
0.10	0.50	1.00	0.10	0.50	1.00	0.11	0.50
0.10	0.70	1.00	0.10	0.70	1.00	0.10	0.70
0.10	0.90	1.00	0.10	0.90	1.00	0.11	0.91
0.30	0.30	0.88	0.30	0.30	0.88	0.30	0.30
0.30	0.50	0.98	0.30	0.30	0.93	0.30	0.50
0.30	0.70	1.00	0.30	0.70	1.00	0.30	0.70
0.30	0.90	1.00	0.30	0.90	1.00	0.25	0.92
0.50	0.50	0.89	0.50	0.50	0.88	0.50	0.50
0.50	0.70	1.00	0.50	0.70	1.00	0.50	0.70
0.50	0.90	1.00	0.50	0.90	1.00	0.47	0.91
0.70	0.70	0.90	0.70	0.70	0.90	0.70	0.70
0.70	0.90	1.00	0.70	0.90	1.00	0.69	0.90
0.90	0.90	0.84	0.90	0.90	0.86	0.90	0.90
		Gumbel			Joe		
0.10	0.10	0.87	0.10	0.10	0.87	0.10	0.10
0.10	0.30	0.85	0.09	0.31	0.91	0.09	0.31
0.10	0.50	1.00	0.10	0.50	1.00	0.10	0.50
0.10	0.70	1.00	0.10	0.70	1.00	0.10	0.70
0.10	0.90	1.00	0.10	0.90	1.00	0.11	0.90
0.30	0.30	0.90	0.30	0.30	0.89	0.30	0.30
0.30	0.50	0.95	0.30	0.30	0.98	0.30	0.50
0.30	0.70	1.00	0.30	0.70	1.00	0.30	0.70
0.30	0.90	1.00	0.30	0.90	1.00	0.30	0.90
0.50	0.50	0.90	0.50	0.50	0.90	0.50	0.50
0.50	0.70	1.00	0.50	0.70	1.00	0.50	0.70
0.50	0.90	1.00	0.50	0.90	1.00	0.50	0.90
0.70	0.70	0.90	0.70	0.70	0.90	0.70	0.70
0.70	0.90	1.00	0.70	0.90	1.00	0.70	0.90
0.90	0.90	0.88	0.90	0.90	0.86	0.90	0.90

Table 4.2: The columns $\tau(\theta_{k(\ell),0})$ and $\tau(\theta_{\ell,0})$ refer to parameter values in terms of Kendall's $\tau(\cdot)$ at the lower and higher hierarchical level, respectively. $s(\hat{\theta}) = s(\theta_0)$ reports the fraction of correctly specified structures, $\tau(\hat{\theta}_{k(\ell)})$ and $\tau(\hat{\theta}_\ell)$ refer to the sample averages of Kendall's $\tau(\cdot)$ evaluated at the estimate of the higher and lower hierarchical level. Monte Carlo sample size is $n = 100$.

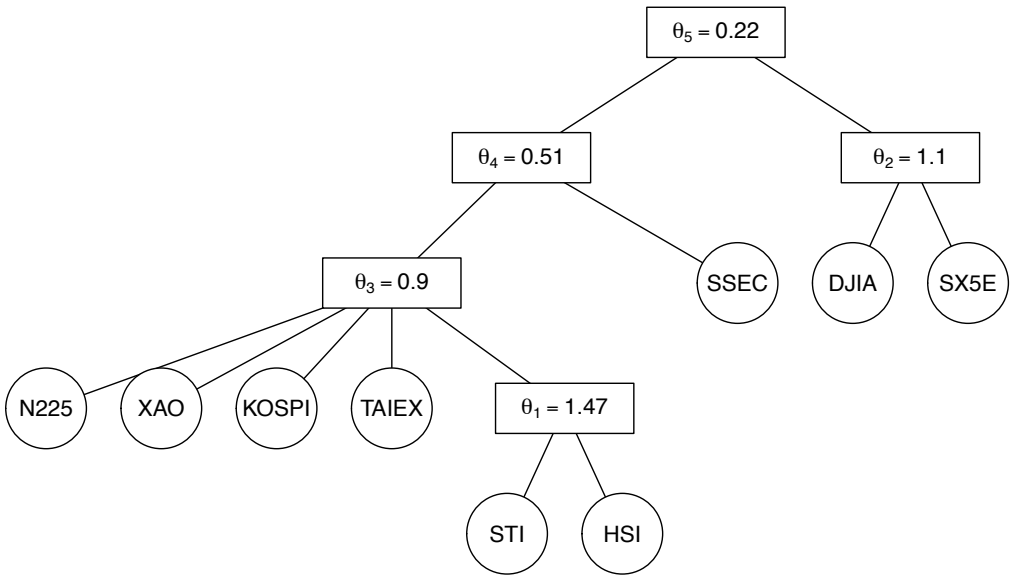

Figure 5.1: The sparsely estimated HAC for the entire data.

Seng Index (HSI) and the Taiwan Capitalization Weighted Stock Index (TAIEX). Based on the different size and geographical proximity of the nine stock market indices considered, i.e., six markets in East and Southeast Asia, the Australian market in relatively close proximity to these markets as well as the significantly more distant major markets in Europe and the US, we expect this choice also to provide an interesting setup for the proposed penalized estimation of the HAC structure. Note also that trading takes place at rather similar time intervals at the Australasian markets, while closing prices for the European and US index are available only much later in the day.

In a classical copula-GARCH setup as, e.g., proposed in Jondeau and Rockinger (2006), the multivariate time series of log-returns is formally modeled by

$$\begin{aligned} X_t &= \mu_t\left(X_{t-1}, \ldots\right) + \sigma_t\left(X_{t-1}, \ldots\right)\varepsilon_t, \\ \varepsilon_t|\mathcal{F}_{t-1} &\sim C\{F_{\varepsilon_1}(x_1), \ldots, F_{\varepsilon_d}(x_d); \theta_t\}, \end{aligned} \tag{17}$$

where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ denotes the information set at time point $t = 1, \ldots, T$. The marginal time series $\{X_{tj}\}_{t=1}^T$, $j = 1, \ldots, d$, are assumed to follow ARMA-APARCH models with skewed-t distributions $F_{\varepsilon_j}(\cdot)$. The parameter vectors of the conditional mean $\mu_t(\cdot)$, the standard deviation $\sigma_t(\cdot)$ and the marginal distributions $F_{\varepsilon_j}(\cdot)$, $j = 1, \ldots, d$, are skipped for notational convenience. A brief description of the data, model selection procedure and performance of the marginal models are reported in Appendix C showing that $\{\varepsilon_{tj}\}_{t=1}^T$, are approximatively iid for $j = 1, \ldots, d$. The dependence between the indices is modeled with a Clayton-based HAC $C(\cdot; \theta_t)$ depending on the parameter sequence $\{\theta_t\}_{t=1}^T$. Even though Archimedean dependence structures might not perfectly describe the dependence among financial returns on the entire support, the Clayton-family is appropriate for analyzing left tail areas. Recall that dependence in the left tail area is a necessary requirement for analyzing spill-over effects in that area according to Proposition 1.

Since the simulation study suggests that a large amount of data is required to recover the "true" structure with high confidence, we firstly use the entire data to fix the structure and secondly apply the rolling window method to estimate time variations of θ_t. The estimation outcome based on the entire sample is presented in Figure 5.1, which shows that the dependencies within the financial system are modeled by five parameters. As there are just five parameters to be estimated in each window, we restrict the window-width to three month, i.e., 63 observations. For comparison, a t-copula approach would require 36 parameters for modeling the dependence structure as well as one parameter for all tail areas. The estimated HAC proposes equi-dependencies ($\theta_3 = 0.9$) between the Australasian indices N225, XAO, KOSPI, TAIEX and the sub-tree of STI and HSI. Slightly stronger dependence is found for the sub-system of DJIA, SX5E. As the copula in (17) models the cross-sectional dependence among the components of X_t at time point $t = 1, \ldots, T$, the weak dependence at the root node of the HAC-structure reflects the different trading times of the indices as well as possibly the geographical distance between European, US and Australasian markets. Note, however that the dependence between the European SX5E and the US DJIA index is estimated to be relatively high, although these markets trade at different times and are also geographically distant. If the estimate at the root-node showed stronger dependence between the sub-systems, it would be interesting to investigate contagion between the systems. Yet, the weak dependence at the root node rather suggests a separate analysis of the sub-systems.

Let us first consider the relationship between log-returns from the US and the European index: exploring contagion between the DJIA and the SX5E, Table C.1 reveals that the distribution of the residuals of the DJIA has *slightly* thicker tails and is more left-skewed than the distribution of SX5E residuals. Given a small value α, this little heterogeneity in the margins implies for the unconditional quantile functions of the log-returns $|Q_{\text{SX5E}}(\alpha)| < |Q_{\text{DJIA}}(\alpha)|$ and for the unconditional quantile densities $q_{\text{SX5E}}(\alpha) < q_{\text{DJIA}}(\alpha)$, since the effect of the time-varying mean and volatility can be neglected as $\alpha \to 0$. Note that the quantile and quantile density functions of the log-returns generally depend on the time-varying mean and volatility. The estimated spill-over effects $S^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$ and $S^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$ are presented in Figure 5.2, where $S^{\alpha}_{k\leftarrow\ell} = S^{\alpha}_{k\ell}$. Obviously, the effect to DJIA by a marginal change in SX5E, $S^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$, is larger than $S^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$ for all levels of $\alpha \in \{0.1, 0.01, 0.0001\}$. Similar to the example in Section 2.1, this result relies on the ratio of quantile densities, $q_{\text{DJIA}}\{C^{-1}_{\text{DJIA}|\cdot}(\alpha)\}/q_{\text{SX5E}}(\alpha)$, which drives the contagion effect $S^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$ as $\alpha \to 0$. Moreover, as $\alpha \to 0$, Figure 5.2 also reveals that the contagion effects $S^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$ and $S^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$ converge to a constant which is independent of the copula parameter $\theta_{2,t}$.

To analyze systemic risk within the sub-system DJIA and SX5E denote by $Q(\alpha)$ the vector of unconditional quantiles, i.e., $Q(\alpha) = \{Q_{\text{DJIA}}(\alpha), Q_{\text{SX5E}}(\alpha)\}^{\top}$. Moreover, recall that a conditional quantile $Q_{X_k|X_{\not k}=x_{\not k}}(\alpha)$ measures the impact on X_k given the event $\{X_{\not k} = x_{\not k}\}$ and that a quantile $Q_k(\alpha)$ measures risk unrelated to a specific event, $\alpha \in (0, 1)$. Loosely speaking, $\|Q(\alpha)\|_2$ can be regarded as measure for unconditional risk. The fact that the systemic risk measure $\mathcal{S}^{\alpha}$ is a normalized weighted average, see (13), suggests that stronger contagion $\mathcal{S}^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$ implies larger contributions of the DJIA to the systemic risk measure $\mathcal{S}^{\alpha}$. This is because of two facts: (i) $\mathcal{S}^{\alpha}_{\text{DJIA}\leftarrow\text{SX5E}}$ is larger than $\mathcal{S}^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$ because of slightly heterogenous margins. (ii) As the DJIA has slightly thicker tails than the SX5E, the weights given by the risk measure $Q_{X_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha) = Q_k\{C^{-1}_{U_k|U_{\not k}=\boldsymbol{\alpha}}(\alpha)\}$, $k = \text{DJIA}$, are larger than those for the SX5E.

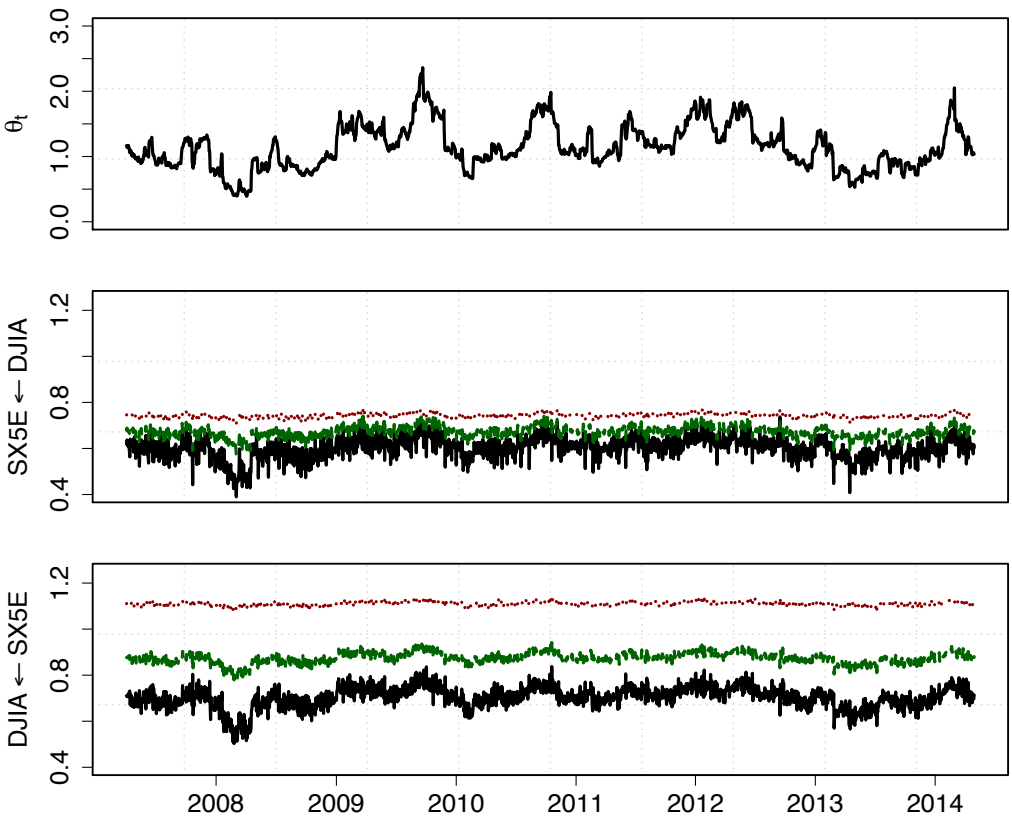

Figure 5.2: The upper panel shows estimates of the dependence parameter $\hat{\theta}_{2,t}$, the centered panel illustrates the risk transmitted from DJIA to SX5E $\mathcal{S}^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$ and the lower panel the risk transmitted from SX5E to DJIA $\mathcal{S}^{\alpha}_{\text{SX5E}\leftarrow\text{DJIA}}$. Solid lines refer to $\alpha = 0.1$, dashed lines to $\alpha = 0.01$ and dotted lines to $\alpha = 0.0001$.

The lower panel of Figure 5.3 brings systemic risk to light which is present at all levels α over the entire period. There are two major components driving $\mathcal{S}^\alpha$: (i) the dependence parameter θ_2 and (ii) the unconditional risk $\|Q(\alpha)\|_2$. Interestingly, the quantities θ_2 and $\|Q(\alpha)\|_2$ have a contrary effect on $\mathcal{S}^\alpha$: stronger dependence between the markets will increase systemic risk, but a higher level of unconditional risk will typically decrease $\mathcal{S}^\alpha$. While the co-movement of systemic risk and dependence satisfies underlying expectations, the relation between $\|Q(\alpha)\|_2$ and $\mathcal{S}^\alpha$ is not obvious. Being precise, $\mathcal{S}^\alpha$ does not measure the general level of systemic risk, but addresses that part of systemic risk arising from contagion in tail areas. For the following consideration let us now assume that the dependence between the markets is constant over time: if volatility in model (17) increases, i.e., $\sigma_{j,t+1}(\cdot) > \sigma_{j,t}(\cdot)$, the absolute value of the unconditional quantile also increases, i.e., $|\mathcal{Q}_{k,t+1}(\alpha)| > |\mathcal{Q}_{k,t}(\alpha)|$, since α is constant and the effect of $\mu_{k,t}$ is usually negligible. To observe a constant systemic risk from t to $(t+1)$ despite increased volatility, joint tail events at $(t+1)$ have to be more extreme than those at time point t. This connection between quantiles and moments leads to the following summary: If, on the one hand, the unconditional risk level is already high, systemic risk due to contagion in the tail area is typically low. If, on the other hand, the unconditional risk level is low, a potential breakdown due to contagion in tail areas becomes more likely, therefore, increasing systemic risk.

For example, let us have a closer look at the second half of year 2008 displayed in Figure 5.3. Despite the financial crisis, systemic risk arising from contagion in the tail area is extremely low compared to other periods of the sample. This is true for all levels of α considered. These results are due to the almost constant dependence parameter θ_{t2} during that period and the enormously high unconditional risk $\|Q(\alpha)\|_2$. Inspection of the data generating process (17) shows that volatilities $\sigma_{tj}(X_{j,t-1},\ldots)$, $t = 1,\ldots,T$, $j = 1,\ldots,d$, are directly related to the unconditional risk measures $Q_j(\alpha)$. There might be a systemic factor driving the volatility of both stock indices DJIA and SX5E during the crisis, but our study does not support the hypothesis of high systemic risk due to contagion in the left tail area.

Analyzing systemic risk within the sub-system rooting at $\theta_4 = 0.51$ of Figure 5.1 shows a similar relation between unconditional risk $\|Q(\alpha)\|_2$ and conditional systemic risk arising from contagion in the tail areas $\mathcal{S}^\alpha$. Note that now $\|Q(\alpha)\|_2$ is computed from the stock indices HSI, KOSPI, N225, SSEC, STI, TAIEX and XAO. Again we find that systemic risk from spill-over effects of an unexpected shock is typically high during periods of low unconditional risk $\|Q(\alpha)\|_2$. Nevertheless, Figure 5.4 shows that the conditional systemic risk is less pronounced than in the sub-system of DJIA and SX5E, which is related by the overall higher level of $\|Q(\alpha)\|_2$. Using the terminology from Section 2, the sub-system can be categorized as robust since $\mathcal{S}^\alpha < 1$, $\alpha = 0.0001$, which emphasizes the absence of systemic risk due to contagion in tail areas.

6. Conclusion

In this paper, we have proposed a consistent and unified framework for describing financial contagion and measuring systemic risk arising from contagion in tail areas. Properties of the developed bilateral contagion measure are comprehensively discussed and derived. Contagion measures describing effects from and to sub-portfolios are derived by aggregating the respective bilateral children. In particular, the effect from a sub-portfolio can be interpreted as a joint shock hitting that subset of the portfolio. A measure to quantify systemic risk is derived by similar arguments and the proposed measures are

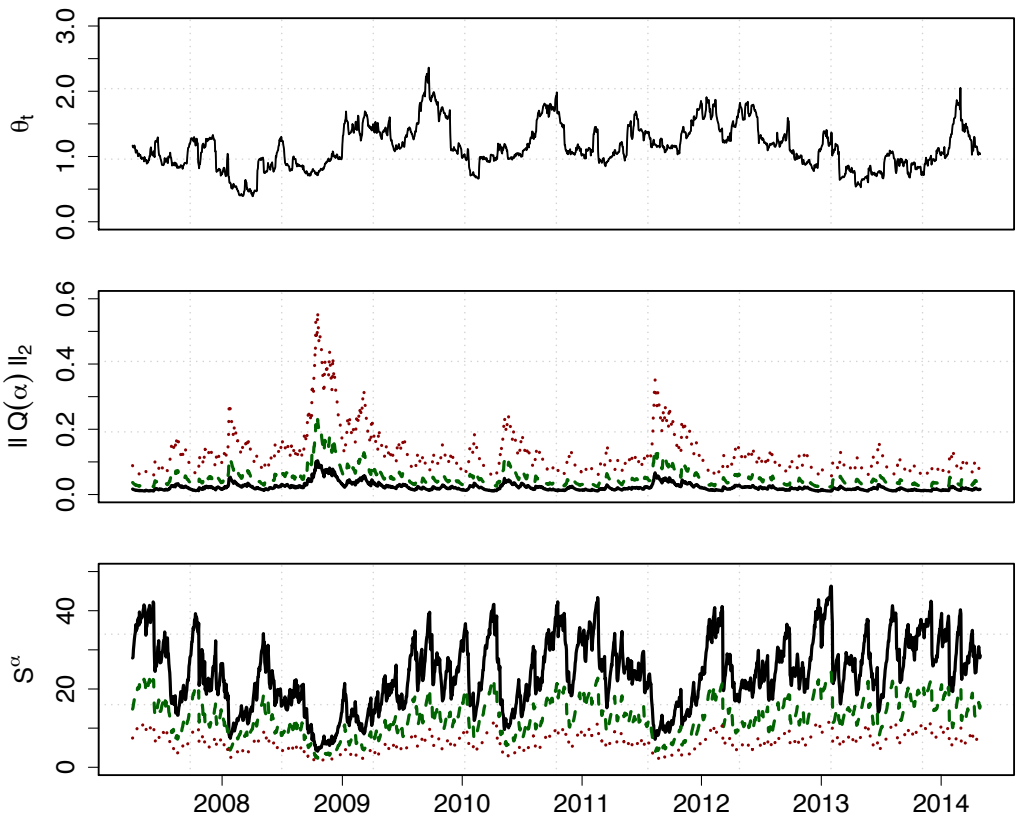

Figure 5.3: The upper panel shows estimates of the dependence parameter $\hat{\theta}_{2,t}$, the centered panel shows the unconditional risk measure $\|Q(\alpha)\|_2$, and the lower panel systemic risk $\mathcal{S}^\alpha$ within the sub-portfolio SX5E and DJIA. Solid lines refer to $\alpha = 0.1$, dashed lines to $\alpha = 0.01$ and dotted lines to $\alpha = 0.0001$.

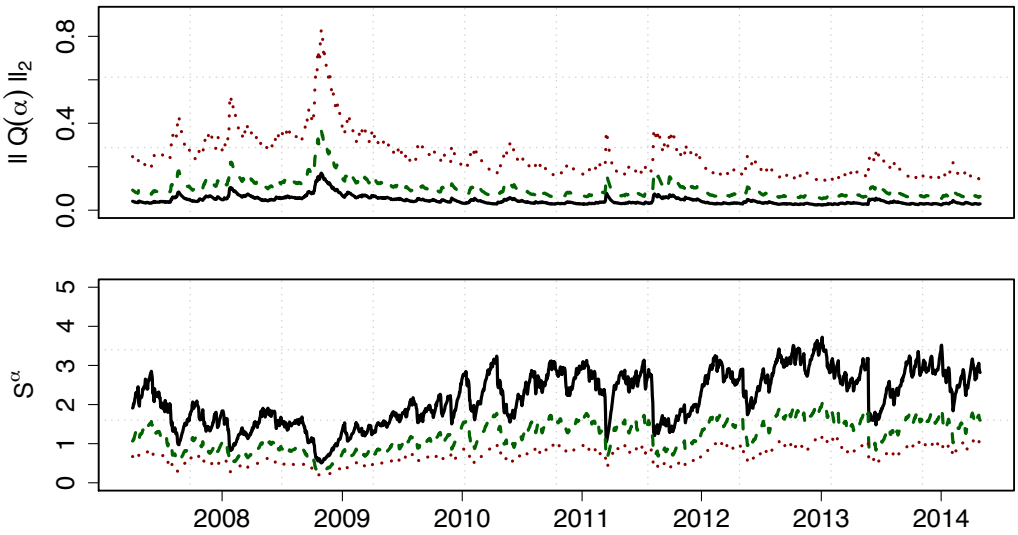

Figure 5.4: The upper panel shows the unconditional risk measure $\|Q(\alpha)\|_2$ and the lower panel systemic risk $\mathcal{S}^\alpha$ for the sub-portfolio HSI, KOSPI, N225, SSEC, STI, TAIEX and XAO. Solid lines refer to $\alpha = 0.1$, dashed lines to $\alpha = 0.01$ and dotted lines to $\alpha = 0.0001$.

related to existing approaches in the literature. To combine a small number of parameters in total with flexibility in tail areas, we suggest modeling dependencies among random variables with hierarchical Archimedean copula and provide a new estimation procedure for that type of copula. Based on theory about penalized Maximum Likelihood estimation, we discuss the asymptotic properties of the estimator which are supported by a simulation study. Last but not least, we have applied the developed tools in an empirical study based on a rolling window analysis of major stock indices. Next to the expected result that an increase in dependence between financial markets also increases systemic risk, our study also reveals new relations between (conditional) systemic risk and the overall (unconditional) risk level. We find that the systemic risk contribution from contagion in tail areas decreases during times of financial turmoil, i.e., we propose that a potential systemic breakdown caused by contagion in the tail areas is less likely during times of high volatility and could rather occur due to a shock in a quiet market environment.

A. Regularity assumptions

Assumption 1. *The model is identifiable and $\theta_{k(\ell),0}$ is an interior point of the compact parameter space $\Theta_{k(\ell)}$. We assume that $\mathrm{E}_{\theta_{k(\ell)}}\{\ell_i'(\theta_{k(\ell)})\} = 0$ and information equality holds,*

$$\mathcal{I}(\theta_{k(\ell)}) \stackrel{\text{def}}{=} \mathrm{E}_{\theta_{k(\ell)}}\left\{\ell_i'(\theta_{k(\ell)})^2\right\} = -\,\mathrm{E}_{\theta_{k(\ell)}}\left\{\ell_i''(\theta_{k(\ell)})\right\} \quad for \quad i = 1, \ldots, n.$$

Assumption 2. *The Fisher information $\mathcal{I}(\theta_{k(\ell)})$ is finite and strictly positive at $\theta_{k(\ell),0}$.*

Assumption 3. *There exists an open subset Ω of $\Theta_{k(\ell)}$ containing the true parameter $\theta_{k(\ell),0}$ such that*

for almost all U_i, $i = 1, \ldots, n$, the density $c(U_{i1}, \ldots, U_{id_k}; \theta_{k(\ell)})$ admits all third derivatives $c'''(\cdot; \theta_{k(\ell)})$ for all $\theta_{k(\ell)} \in \Omega$. Furthermore, there exist functions $M(\cdot)$ such that $\left|\ell_i'''(\theta_{k(\ell)})\right| \leq M(U_i)$, for all $\theta_{k(\ell)} \in \Omega$, with $\mathrm{E}\{M(U_i)\} < \infty$.

B. Mathematical appendix

Proof of Proposition 1.
Note that $\lim_{u_\ell \to 1} \mathcal{S}_{k\ell}^{u_\ell}$ equals

$$\mathcal{S}_{k\ell}^{u_\ell} = \frac{\Big[\log Q_k\{C_{U_k|U_\ell=u_\ell}^{-1}(\alpha)\}\Big]'}{\Big[\log Q_\ell(u_\ell)\Big]'} \quad \text{as} \quad u_\ell \to 1. \tag{18}$$

To prove part (a), recall that the quantile density function is defined as derivative of the quantile function, i.e., $q(u) = Q'(u)$. Parzen (1979) shows that $q(u) \sim (1-u)^{-\gamma}$, $u \to 1$, where $q(\cdot)$ denotes the quantile density function of a probability law with tail-monotone density function $f(\cdot)$ and tail exponent γ. Since $\int (1-u)^{-\gamma} du = (\gamma-1)^{-1}(1-u)^{1-\gamma} + K$, we conclude that $Q(u) \sim (\gamma-1)^{-1}(1-u)^{1-\gamma}$, $u \to 1$, where K is a constant independent of u. As a consequence, (18) can be reformulated as

$$\begin{aligned}\mathcal{S}_{k\ell}^{u_\ell} &= \frac{(1-\gamma_k)\Big[\log\{1 - C_{U_k|U_\ell=u_\ell}^{-1}(\alpha)\}\Big]'}{(1-\gamma_\ell)\Big[\log(1-u_\ell)\Big]'} \\ &= \frac{\gamma_k - 1}{\gamma_\ell - 1} \frac{1-u_\ell}{1 - C_{U_k|U_\ell=u_\ell}^{-1}(\alpha)} \frac{\partial}{\partial u_\ell} C_{U_k|U_\ell=u_\ell}^{-1}(\alpha) \quad \text{as} \quad u_\ell \to 1.\end{aligned} \tag{19}$$

Since the limit of (19) is not well defined, we apply l'Hôpital's rule and represent (19) as

$$\mathcal{S}_{k\ell}^{u_\ell} = \frac{\gamma_k - 1}{\gamma_\ell - 1} \frac{1}{\frac{\partial}{\partial u_\ell} C_{U_k|U_\ell=u_\ell}^{-1}(\alpha)} \frac{\partial}{\partial u_\ell} C_{U_k|U_\ell=u_\ell}^{-1}(\alpha) \to \frac{\gamma_k - 1}{\gamma_\ell - 1} \quad \text{as} \quad u_\ell \to 1.$$

To prove part (b), we build on another result of Parzen (1979), who shows for the case $\gamma = 1$ that

$$q(u) \sim (1-u)^{-1}\{-\log(1-u)\}^{\beta-1}, u \to 1, \quad \text{and} \quad \beta \in [0,1],$$

where β is the so called shape parameter. We exclude extreme cases from our analysis and restrict the discussion to the case $\beta \in (0,1)$. The cases $\beta = 0$ and $\beta = 1$ refer to the extreme-value and to the exponential distribution respectively. Integration, i.e., $Q(u) \sim \beta^{-1}\{-\log(1-u)\}^{\beta}$, $u \to 1$, allows rewriting (18) as

$$\mathcal{S}_{k\ell}^{u_\ell} = \frac{1-\gamma_k}{\beta_\ell} \frac{(1-u_\ell)\log(1-u_\ell)}{1 - C_{U_k|U_\ell=u_\ell}^{-1}(\alpha)} \frac{\partial}{\partial u_\ell} C_{U_k|U_\ell=u_\ell}^{-1}(\alpha) \quad \text{as} \quad u_\ell \to 1, \tag{20}$$

where the parameter β_ℓ describes the shape of the law of X_ℓ. Since the limit of (20) is not well

defined, we apply l'Hôpital's rule and obtain

$$\mathcal{S}_{k\ell}^{u_\ell} = \frac{1-\gamma_k}{\beta_\ell} \frac{1+\log(1-u_\ell)}{\frac{\partial}{\partial u_\ell} C^{-1}_{U_k|U_\ell=u_\ell}(\alpha)} \frac{\partial}{\partial u_\ell} C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) \to \infty \quad \text{as} \quad u_\ell \to 1,$$

where the positive sign of the limit is due to $\gamma_k > 1$.

To prove part (c) for $\gamma_\ell > 1$, note that $C^{-1}_{U_k|U_\ell=u_\ell}(\alpha) \sim F_k\{g(\alpha)\}$, $u_\ell \to 1$, by the definition of conditional independence, so that (19) can be rewritten

$$\mathcal{S}_{k\ell}^{u_\ell} = \frac{\gamma_k - 1}{\gamma_\ell - 1} \frac{1-u_\ell}{1-F_k\{g(\alpha)\}} \frac{\partial}{\partial u_\ell} F_k\{g(\alpha)\} \to 0 \quad \text{as} \quad u_\ell \to 1.$$

The proof for $\gamma_\ell = 1$ follows by similar arguments. □

Lemma 1 (C.f., Fan and Li (2001) and Cai and Wang (2014))**.** *Under the assumptions of Proposition 2, if $n^{1/2}\lambda_n \to \infty$ as $n \to \infty$, then for $\theta_{k(\ell)}$ satisfying $(\theta_{k(\ell)} - \theta_{k(\ell),0}) = \mathcal{O}_p(n^{-1/2})$,*

$$\lim_{n\to\infty} \mathrm{P}\left\{\mathcal{Q}_{\theta_\ell}(\theta_{k(\ell)}) \le \mathcal{Q}_{\theta_\ell}(\theta_\ell)\right\} = 1. \tag{21}$$

Proof of Lemma 1.

Due to the natural constraints on the parameters, we need to show that for $\varepsilon_n = An^{-1/2}$, with $A > 0$,

$$\mathcal{Q}'_{\theta_\ell}(\theta_{k(\ell)}) > 0 \quad \text{for} \quad 0 < \theta_\ell - \theta_{k(\ell)} < \varepsilon_n. \tag{22}$$

Taylor expansion of $\ell'_i(\theta_{k(\ell)})$ about $\theta_{k(\ell),0}$ leads to

$$\begin{aligned}\mathcal{Q}'_{\theta_\ell}(\theta_{k(\ell)}) = &\sum_{i=1}^n \ell'_i(\theta_{k(\ell),0}) + \sum_{i=1}^n \ell''_i(\theta_{k(\ell),0})(\theta_{k(\ell)} - \theta_{k(\ell),0}) + \frac{1}{2}\sum_{i=1}^n \ell'''_i(\bar{\theta}_{k(\ell)})(\theta_{k(\ell)} - \theta_{k(\ell),0})^2 \\ &+ np'_{\lambda_n}(\theta_\ell - \theta_{k(\ell)}),\end{aligned}$$

with $\bar{\theta}_{k(\ell)}$ lying between $\theta_{k(\ell)}$ and $\theta_{k(\ell),0}$. By classical arguments, c.f., Lehmann and Casella (1998),

$$n^{-1}\sum_{i=1}^n \ell'_i(\theta_{k(\ell),0}) = \mathcal{O}_p(n^{-1/2}), \quad \text{and} \quad n^{-1}\sum_{i=1}^n \ell''_i(\theta_{k(\ell),0}) = -\mathcal{I}(\theta_{k(\ell),0}) + o_p(1),$$

and by using $(\theta_{k(\ell)} - \theta_{k(\ell),0}) = \mathcal{O}_p(n^{-1/2})$, we obtain

$$\begin{aligned}\mathcal{Q}'_{\theta_\ell}(\theta_{k(\ell)}) &= \mathcal{O}_p(n^{1/2}) + np'_{\lambda_n}(\theta_\ell - \theta_{k(\ell)}) \\ &= n\lambda_n\left\{\mathcal{O}_p(n^{-1/2}/\lambda_n) + p'_{\lambda_n}(\theta_\ell - \theta_{k(\ell)})/\lambda_n\right\}.\end{aligned}$$

As $\liminf_{n\to\infty} \liminf_{\theta_{k(\ell)}\to\theta_\ell^-} p'_{\lambda_n}(\theta_\ell - \theta_{k(\ell)})/\lambda_n > 0$ for the considered penalty and $\mathcal{O}_p(n^{-1/2}/\lambda_n) \to 0$ as $n \to \infty$, (22) holds. □

Lemma 2. *Under the assumptions of Proposition 2 and 3, if $\lambda_n \to 0$ as $n \to \infty$, then $\hat{\theta}_\ell - \theta_{\ell,0} = o_p(1)$.*

Proof of Lemma 2.
Let $c(u, v; \theta)$ be the density function of a bivariate Archimedean copula. The stage-wise estimation procedure is initialized with estimating the parameter θ_1 for all possible pairs $(U_1, U_2)^\top, \ldots, (U_1, U_d)^\top, \ldots, (U_{d-1}, U_d)^\top$, i.e., $\hat{\theta}_{1,gh} = \arg\max_{\theta_1} \sum_{i=1}^n \log\{c(U_{ig}, U_{ih}; \theta_1)\}$ for $g = 1, \ldots, d-1$, $h = g+1, \ldots, d$, and selecting the largest estimate among all estimated parameters. Suppose $\hat{\theta}_{1,(d-1)d}$ is the estimator with the largest value and set $\hat{\theta}_1 = \hat{\theta}_{1,(d-1)d}$. Then, $\hat{\theta}_1 - \theta_{1,0} = \mathcal{o}_p(1)$ follows by classical ML theory. Remove U_{d-1} and U_d from the set of variables. To finish the first estimation-stage set $U_{d-1} = \phi_{\hat{\theta}_1}\left[2\phi^{-1}_{\hat{\theta}_1}\{\max(U_{d-1}, U_d)\}\right]$ and add $U_{d-1} \sim \mathrm{U}(0,1)$ to the set of variables again.

At the next estimation stage θ_2 is estimated for all possible pairs $(U_1, U_2)^\top, \ldots, (U_1, U_{d-1})^\top, \ldots, (U_{d-2}, U_{d-1})^\top$, i.e., $\hat{\theta}^{\lambda_n}_{2,gh} = \arg\max_{\theta_2} \sum_{i=1}^n \log\{c(U_{ig}, U_{ih}; \theta_2)\} - np_{\lambda_n}(\hat{\theta}_1 - \theta_2)$, for $g = 1, \ldots, d-2$, $h = g+1, \ldots, d-1$, and selecting the largest estimate among all estimated parameters. Suppose $\hat{\theta}^{\lambda_n}_{2,(d-2)(d-1)}$ is the estimator with largest value, set $\hat{\theta}^{\lambda_n}_2 = \hat{\theta}^{\lambda_n}_{2,(d-2)(d-1)}$ and assume $\theta_{1,0} = \theta_{2,0}$. Due to the consistency of $\hat{\theta}_1$, we obtain $\mathcal{Q}_{\hat{\theta}_1}(\theta_2) = \mathcal{Q}_{\theta_{1,0}}(\theta_2) + o_p(1)$. If $\lambda_n \to 0$ as $n \to \infty$, the penalized estimator $\hat{\theta}^{\lambda_n}_{2,n}$ is $n^{1/2}$-consistent, c.f., Fan and Li (2001); Cai and Wang (2014), and Lemma 1 implies $\mathrm{P}\{\mathcal{Q}_{\theta_{1,0}}(\hat{\theta}^{\lambda_n}_2) \leq \mathcal{Q}_{\theta_{1,0}}(\theta_{1,0})\} \to 1$ as $n \to \infty$. If $\theta_{1,0} > \theta_{2,0}$, it can be shown by applying arguments of the proof of Proposition 3, that the bias arising from penalized ML estimation vanishes asymptotically as $\lambda_n \to 0$, so that $(\hat{\theta}^{\lambda_n}_2 - \theta_{2,0}) = \mathcal{o}_p(1)$ as $n \to \infty$. The statement follows by iteratively repeating the previous steps up the ℓ-th hierarchical level. □

Proof of Proposition 2.
Due to Lemma 2, we obtain $\mathcal{Q}_{\hat{\theta}_\ell}(\theta_{k(\ell)}) = \mathcal{Q}_{\theta_{\ell,0}}(\theta_{k(\ell)}) + o_p(1)$. For any $\theta_{k(\ell)}$ satisfying $(\theta_{k(\ell)} - \theta_{k(\ell),0}) = \mathcal{O}_p(n^{-1/2})$, Lemma 1 implies $\mathrm{P}\{\mathcal{Q}_{\theta_{\ell,0}}(\hat{\theta}^{\lambda_n}_{k(\ell)}) \leq \mathcal{Q}_{\theta_{\ell,0}}(\theta_{\ell,0})\} \to 1$ as $n \to \infty$, which completes the proof. □

Proof of Proposition 3.
Note that the estimator $\hat{\theta}^{\lambda_n}_{k(\ell)}$ satisfies $0 = \sum_{i=1}^n \ell'_i(\hat{\theta}^{\lambda_n}_{k(\ell)}) + np'_{\lambda_n}(\hat{\theta}_\ell - \hat{\theta}^{\lambda_n}_{k(\ell)})$. Taylor expansion of $\ell'_i(\cdot)$ about $\theta_{k(\ell),0}$ and $p'_{\lambda_n}(\cdot)$ about $\theta_0^- = (\theta_{\ell,0} - \theta_{k(\ell),0})$ leads to

$$0 = \sum_{i=1}^n \ell'_i(\theta_{k(\ell),0}) + \sum_{i=1}^n \ell''_i(\theta_{k(\ell),0})(\hat{\theta}^{\lambda_n}_{k(\ell)} - \theta_{k(\ell),0}) + n\left[p'_{\lambda_n}(\theta_0^-) + p''_{\lambda_n}(\theta_0^-)\{(\hat{\theta}_\ell - \hat{\theta}^{\lambda_n}_{k(\ell)}) - \theta_0^-\}\right],$$

which can be rewritten as

$$n^{1/2}\left\{\widehat{\mathcal{I}}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_0^-)\right\}\left[(\hat{\theta}^{\lambda_n}_{k(\ell)} - \theta_{k(\ell),0}) - \left\{\widehat{\mathcal{I}}(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_0^-)\right\}^{-1} p'_{\lambda_n}(\theta_0^-)\right]$$
$$= n^{-1/2}\sum_{i=1}^n \ell'_i(\theta_{k(\ell),0}) + p''_{\lambda_n}(\theta_0^-)n^{1/2}(\hat{\theta}_\ell - \theta_{\ell,0}). \quad (23)$$

Note that the existence of a $n^{1/2}$-consistent estimator $\hat{\theta}^{\lambda_n}_{k(\ell)}$ requires the condition $p''_{\lambda_n}(\theta_{\ell,0} - \theta_{k(\ell),0}) \to 0$ being satisfied for $\theta_{\ell,0} > \theta_{k(\ell),0}$ as $n \to \infty$, c.f., Fan and Li (2001, Theorem 1). As this property holds for the SCAD penalty, we obtain $p''_{\lambda_n}(\theta_0^-) = p''_{\lambda_n}(\theta_{\ell,0} - \theta_{k(\ell),0}) = \mathcal{o}(1)$. Using Lemma 2, it can be easily shown that $n^{1/2}(\hat{\theta}_\ell - \theta_{\ell,0}) = \mathcal{O}_p(1)$. Hence, the right hand side of (23) converges to $\mathrm{N}\{0, \mathcal{I}(\theta_{k(\ell),0})\}$

by a central limit theorem as $n \to \infty$. $\square$

Proof of Proposition 4.
Since $\hat{\theta}_\ell \geq \theta_{k(\ell)}$ and $\hat{\theta}_\ell \geq \hat{\theta}_{k(\ell)}$, a linear approximation of $p_\lambda(|\hat{\theta}_\ell - \theta_{k(\ell)}|)$, for $(\hat{\theta}_\ell - \theta_{k(\ell)}) \approx (\hat{\theta}_\ell - \hat{\theta}_{k(\ell)})$, gives

$$p_\lambda(|\hat{\theta}_\ell - \theta_{k(\ell)}|) \approx p_\lambda(\hat{\theta}_\ell - \hat{\theta}_{k(\ell)}) + p'_\lambda(\hat{\theta}_\ell - \hat{\theta}_{k(\ell)}) \left\{ (\hat{\theta}_\ell - \theta_{k(\ell)}) - (\hat{\theta}_\ell - \hat{\theta}_{k(\ell)}) \right\}. \tag{24}$$

Quadratic approximation of $\ell_i(\theta_{k(\ell)})$ around $\hat{\theta}_{k(\ell)}$ and $\ell'(\hat{\theta}_{k(\ell)}) = 0$, leads to the minimization problem

$$\arg\min_{\theta_{k(\ell)}} \frac{1}{2}(\theta_{k(\ell)} - \hat{\theta}_{k(\ell)})^2 \left\{ -\sum_{i=1}^{n} \ell''_i(\hat{\theta}_{k(\ell)}) \right\} + n p'_{\lambda_n}(\hat{\theta}_\ell - \hat{\theta}_{k(\ell)}) \left\{ (\hat{\theta}_\ell - \theta_{k(\ell)}) - (\hat{\theta}_\ell - \hat{\theta}_{k(\ell)}) \right\}.$$

Ignoring non-relevant parts and solving it with respect to $\theta_{k(\ell)}$ gives the unique solution $\hat{\theta}^{\lambda_n}_{k(\ell)} = \hat{\theta}_{k(\ell)} + \widehat{\mathcal{I}}(\hat{\theta}_{k(\ell)})^{-1} p'_{\lambda_n}(\hat{\theta}_{\ell(k)} - \hat{\theta}_k)$, where information equality gives $-n^{-1}\sum_{i=1}^{n} \ell''_i(\hat{\theta}_{k(\ell)}) \approx n^{-1}\sum_{i=1}^{n} \ell'_i(\hat{\theta}_{k(\ell)})^2 = \widehat{\mathcal{I}}(\hat{\theta}_{k(\ell)})$, which completes the statement of Proposition 4. $\square$

C. Data description and fit of the marginal distributions

The daily log-returns obtained from Datastream cover the principal equity indices of the following nine areas/countries over the period January 1, 2007 - April 30, 2014: USA (DJIA), Europe (SX5E), Japan (N225), China (SSEC), Australia (XAO), Singapore (STI), Korea (KOSPI), Hongkong (HSI) and Taiwan (TAIEX). Missing values in the marginal time series, which consist of 1913 observations each, are replaced by the sample average of the surrounding 10-20 observations. ARMA-APARCH models with skew-t^{χ}_{ν} distributed error terms are employed to remove temporal dependence, see Ding, Granger, and Engle (1993) and Fernández and Steel (1998). The models are selected according to the smallest BIC, where up to three autoregressive and moving average lags are considered in the ARMA and APARCH components. The shape and skewness parameter is denoted by $\nu \in [1, \infty)$ and $\chi \in (0, \infty)$ respectively.

The estimated values presented in Table C.1 indicate that the distributions of the residuals have heavy tails and are slightly left-skewed, where $\chi = 1$ indicates symmetry. Instead of presenting the entire estimation results, we only report results on diagnostic tests concerning the autocorrelation in the first and second moment of the standardized residuals as well as the Anderson Darling test for assessing the model fit, see Box and Pierce (1970) and Anderson and Darling (1952). The corresponding p-values are listed in Table C.1. Even though there is weak evidence for autocorrelation in ε_i^2 for the N225 and STI as well as in ε_i for the SSEC and TAIEX, the model fit is regarded as sufficiently good, as each series passes the Anderson Darling test at the 5% significance level.

Index	χ	ν	$Q_{10}(\varepsilon_i)$	$Q_{15}(\varepsilon_i)$	$Q_{10}(\varepsilon_i^2)$	$Q_{15}(\varepsilon_i^2)$	AD GoF
DJIA	0.85	6.22	0.96	0.85	0.43	0.76	0.08
HSI	0.92	8.24	0.70	0.26	0.19	0.32	0.28
KOSPI	0.87	7.28	0.19	0.49	0.42	0.17	0.44
N225	0.89	10.55	0.69	0.77	0.91	0.03	0.23
SSEC	0.91	4.55	0.03	0.10	0.07	0.16	0.21
STI	0.90	12.89	0.15	0.16	0.06	0.03	0.83
SX5E	0.91	7.94	0.92	0.85	0.15	0.20	0.66
TAIEX	0.86	5.67	0.19	0.02	0.58	0.58	0.15
XAO	0.84	16.88	0.79	0.86	0.95	0.96	0.69

Table C.1: Parameter estimates for skewed Student-t distribution and p-values for conducted Ljung-Box tests, $Q_l(\cdot)$, for lags $l \in \{10, 15\}$, and the Anderson-Darling goodness of fit test (AD GoF) for daily log-returns of the considered stock market indices.

References

Acharya VV, Pedersen LH, Philippon T, Richardson M (2010). "Measuring Systemic Risk." *Working Paper 1002*, Federal Reserve Bank of Cleveland.

Adrian T, Brunnermeier MK (2011). "CoVaR." *Working Paper 17454*, National Bureau of Economic Research.

Anderson TW, Darling DA (1952). "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes." *The Annals of Mathematical Statistics*, **23**(2), 193–212.

Bedford T, Cooke RM (2001). "Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines." *Annals of Mathematical and Artificial Intelligence*, **32**, 245–268.

Bedford T, Cooke RM (2002). "Vines – a New Graphical Model for Dependent Random Variables." *The Annals of Statistics*, **30**(4), 1031–1068.

Bernard C, Czado C (2015). "Conditional Quantiles and Tail Dependence." *Journal of Multivariate Analysis*, **138**(0), 104–126.

Bisias D, Flood M, Lo AW, Valavanis S (2012). "A Survey of Systemic Risk Analytics." *Working Paper 0001,2012*, U.S. Department of Treasury.

Bouyé E, Salmon M (2009). "Dynamic Copula Quantile Regressions and Tail Area Dynamic Dependence in Forex Markets." *The European Journal of Finance*, **15**(7–8), 721–750.

Box GEP, Pierce DA (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association*, **65**(332), 1509–1526.

Brunnermeier MK, Oehmke M (2013). "Bubbles, Financial Crises, and Systemic Risk." In *Handbook of the Economics of Finance*. Elsevier, Amsterdam.

Cai Z, Wang X (2014). "Selection of Mixed Copula Model via Penalized Likelihood." *Journal of the American Statistical Association*, **109**(506), 788–801.

Cherubini U, Mulinacci S (2015). "Systemic Risk with Exchangeable Contagion: Application to the European Banking System." *Papers*, arXiv.org. URL http://EconPapers.repec.org/RePEc:arx:papers:1502.01918.

Choroś-Tomczyk B, Härdle WK, Okhrin O (2013). "Valuation of Collateralized Debt Obligations with Hierarchical Archimedean Copulae." *Journal of Empirical Finance*, **24**(C), 42–62.

Christoffersen P, Jacobs K (2014). "Dynamic Dependence and Diversification in Corporate Credit."

Coval J, Jurek J, Stafford E (2009). "The Economics of Structured Finance." *Journal of Economic Perspectives*, **23**(1), 3–25.

Diebold FX, Yilmaz K (2014). "On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms." *Journal of Econometrics*, **182**(1), 119–134.

Ding Z, Granger CWJ, Engle RF (1993). "A Long Memory Property of Stock Market Returns and a New Model." *Journal of Empirical Finance*, **1**(1), 83–106.

Efron B, Hastie T, Johnstone I, Tibshirani R (2004). "Least Angle Regression." *The Annals of Statistics*, **32**(2), 407–451.

Engle RF, Kelly B (2011). "Dynamic Equicorrelation." *Journal of Business & Economic Statistics*, **30**(2), 212–228.

Engle RF, Manganelli S (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." *Journal of Business & Economic Statistics*, **22**(4), 367–381.

Fan J, Li R (2001). "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association*, **96**(456), 1348–1360.

Fan J, Lv J (2010). "A Selective Overview of Variable Selection in High Dimensional Feature Space." *Statistica Sinica*, **20**(1), 101–148.

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

Fernández C, Steel MFJ (1998). "On Bayesian Modeling of Fat Tails and Skewness." *Journal of the American Statistical Association*, **93**(441), 359–371.

Górecki J, Hofert M, Holeňa M (2014). "On the Consistency of an Estimator for Hierarchical Archimedean Copulas." In J Talašová, J Stoklasa, T Talášek (eds.), *32nd International Conference on Mathematical Methods in Economics*, pp. 239–244. Palacký University, Olomouc.

Härdle WK, Okhrin O, Okhrin Y (2013). "Dynamic Structured Copula Models." *Statistics and Risk Modeling*, **30**, 361–388.

Härdle WK, Okhrin O, Wang W (2015). "Hidden Markov Structures for Dynamic Copulae." *Econometric Theory*, **FirstView**, 1–35.

Härdle WK, Wang W, Yu L (2015). "TENET: Tail-Event Driven NETwork Risk."

Hautsch N, Schaumburg J, Schienle M (2015). "Financial Network Systemic Risk Contributions." *Review of Finance*, **19**(2), 685–738.

Hering C, Hofert M, Mai JF, Scherer M (2010). "Constructing Hierarchical Archimedean Copulas with Lévy Subordinators." *Journal of Multivariate Analysis*, **101**, 1428–1433.

Hofert M (2011). "Efficiently Sampling Nested Archimedean Copulas." *Computational Statistics & Data Analysis*, **55**, 57–70.

Hofert M, Scherer M (2011). "CDO Pricing with Nested Archimedean Copulas." *Quantitative Finance*, **11**, 775–87.

Ishikawa A, Kamada K, Kurachi Y, Nasu K, Teranishi Y (2012). "Introduction to the Financial Macro-econometric Model." *Bank of Japan Working Paper Series 12-E-1*, Bank of Japan.

Jaworski P, Durante F, Härdle WK (2013). *Copulae in Mathematical and Quantitative Finance*, volume 213 of *Lecture Notes in Statistics*. Springer.

Joe H (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Jondeau E, Rockinger M (2006). "The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application." *Journal of International Money and Finance*, **25**(5), 827–853.

Jones M (1992). "Estimating Densities, Quantiles, Quantile Densities and Density Quantiles." *Annals of the Institute of Statistical Mathematics*, **44**(4), 721–727.

Koenker R, Bassett G (1978). "Regression Quantiles." *Econometrica*, pp. 33–50.

Kurowicka D, Joe H (2011). *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Company, Incorporated.

Lehmann EL, Casella G (1998). *Theory of Point Estimation*. 2 edition. Springer.

Lucas A, Schwaab B, Zhang X (2014). "Conditional Euro Area Sovereign Default Risk." *Journal of Business & Economic Statistics*, **32**(2), 271–284.

Mainik G, Schaanning E (2014). "On Dependence Consistency of CoVaR and some other Systemic Risk Measures." *Statistics & Risk Modeling*, **31**(1), 49–77.

McNeil AJ (2008). "Sampling Nested Archimedean Copulas." *Journal of Statistical Computation and Simulation*, **78**, 567–581.

Nelsen RB (2006). *An Introduction to Copulas*. Springer, New York.

Oh DH, Patton AJ (2014). "Modelling Dependence in High Dimensions."

Okhrin O, Odening M, Xu W (2013a). "Systemic Weather Risk and Crop Insurance: The Case of China." *Journal of Risk and Insurance*, **80**(2), 351–372.

Okhrin O, Okhrin Y, Schmid W (2013b). "On the Structure and Estimation of Hierarchical Archimedean Copulas." *Journal of Econometrics*, **173**, 189–204.

Okhrin O, Okhrin Y, Schmid W (2013c). "Properties of Hierarchical Archimedean Copulas." *Statistics & Risk Modeling*, **30**(1), 21–54.

Okhrin O, Ristig A (2014). "Hierarchical Archimedean Copulae: The HAC Package." *Journal of Statistical Software*, **58**(4), 1–20. URL http://www.jstatsoft.org/v58/i04/.

Parzen E (1979). "Nonparametric Statistical Data Modeling." *Journal of the American Statistical Association*, **74**(365), 105–121.

Rezapour M (2015). "On the Construction of Nested Archimedean Copulas for d-Monotone Generators." *Statistics & Probability Letters*, **101**(0), 21–32.

Savu C, Trede M (2010). "Hierarchies of Archimedean Copulas." *Quantitative Finance*, **10**(3), 295–304.

Schwarz G (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, **6**(2), 461–464.

Segers J, Uyttendaele N (2014). "Nonparametric Estimation of the Tree Structure of a Nested Archimedean Copula." *Computational Statistics & Data Analysis*, **72**(0), 190–204.

Sklar A (1959). "Fonctions de Répartition à n Dimension et Leurs Marges." *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 299–231.

Sydsæter K, Hammond P (1995). *Mathematics for Economic Analysis*. Prentice-Hall International editions. Prentice Hall.

Tibshirani R (1996). "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 267–288.

Tibshirani R (2011). "Regression Shrinkage and Selection via the Lasso: a Retrospective." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(3), 273–282.

Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005). "Sparsity and Smoothness via the Fused Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(1), 91–108.

Wang H, Li R, Tsai CL (2007). "Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method." *Biometrika*, **94**(3), 553–568.

White H (1994). *Estimation, Inference and Specification Analysis*. 1 edition. Cambridge University Press, Cambridge.

White H, Kim TH, Manganelli S (2015). "VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles." *Journal of Econometrics*, **187**(1), 169–188.

Zimmer DM (2012). "The Role of Copulas in the Housing Crisis." *Review of Economics and Statistics*, **94**(2), 607–620.

Zou H, Li R (2008). "One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models." *The Annals of Statistics*, **36**(4), 1509–1533.