Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis in Econometrics and Management Science:
Quantitative Finance

Tail Dependence Structures of Commodity Futures Prices

Author:
V.H. Vloeberghs

Student number:
568012

Supervisor:
Dr. P. Wan

Second assessor:
Prof. dr. C. Zhou

July 14, 2022

Abstract

This research studies the upper dependence structure of extremes in the energy-agriculture-metal price nexus for the connected commodities system. It uses the theory on extremal graphical models introduced by Engelke and Hitz (2020) to study the 21 futures returns and focuses on the method for identifying the structure. Three methods are used. A forward greedy procedure, the extremal graphical lasso and a newly proposed clustering algorithm based on the Leiden community finding algorithm. A simulation study finds that using these methods can improve upon the accuracy of an empirical estimator of the tail correlation and that the extremal graphical lasso is the most accurate model, although it tends to overestimate the number of edges. The clustering algorithm is found to be sensitive to mislabelling of the nodes. From the empirical application, it is found that the results of the three approaches are close to each other and that the strongest connections are found between nodes that are in the same group. In the connections between groups, we see that natural gas serves as the strongest connection between groups. The strength of these connections is lower with respect to the non-extreme correlations. This supports the neutrality hypothesis.

Contents

1 Introduction

In the period between January 2006 and May 2008, the world's food prices increased dramatically. During that time, the price of corn rose by 180%, soybeans by 127% and rice by 134%, according to the front-month futures. Rises in food prices affect households with low levels of assets in particular (Alem and Söderbom, 2012) and have caused an increase in poverty in low-income countries during that time (Marktanner and Noiset, 2013). This period came to be known in history as the world food price crisis. Multiple causes have been linked to the events: droughts in grain-producing nations (Fellmann et al., 2014), rice export restrictions in Vietnam and India (Headey, 2011), but most notably the rise in oil prices and biofuels demand (Headey and Fan, 2008). In that same period, the price of Brent crude oil rose by 104%. The linkage between oil and food prices is called the oil-food nexus. There are two main mechanisms through which the two are coupled.

The first one is a cost-push effect coming from the increase in production expenses. Agriculture is an energy-intensive sector that relies in multiple ways on petroleum products. Transportation and processing of agricultural products require fuels for the machinery, such as diesel and petrol. Furthermore, fertilisers and other agrarian chemicals use fossil resources, such as natural gas as input.

The second effect is a demand-pull effect that results from the rising demand for biofuels. First-generation biofuels come directly from biomass and compete with other agricultural products, such as food, for arable land. Consumption of biofuels has increased by over 50% since 2010, according to BP's Statistical Review of World Energy 2021 (BP, 2021), and is expected to increase in the coming decades (IEA, 2021). Reboredo (2012), Kristoufek et al. (2012), Yahya et al. (2019) show that the link between food and energy prices has increased over the years, fuelling the debate on the food vs. fuel dilemma.

Aside from energy and agricultural commodities, a third commodity type is metals. It has also been shown that it is closely linked to energy prices due to the energy-intensive nature of processing the material (Reboredo and Ugolini, 2016). For this commodity type, we typically make the distinction between industrial metals, such as copper, tin and zinc, and precious metals, such as gold, silver and platinum.

While the literature has been abundant on the regular co-movements of the three commodity markets, the focus has been less on their tail dependence. The research that studies the tails usually relies on copulas. This approach is limited in the number of commodities series that can be considered in cross-section due to estimation efficiency in the parameters. Although the futures market returns of all three commodity types have been shown to exhibit heavy-tailedness, few studies directly model the threshold exceedances using Pareto distributions. This research adds to the current literature by studying the upper dependence structure of extremes in the energy-agriculture-metal price nexus for the connected commodities system, consisting of both raw resources and products. To the best of the authors' knowledge, this has not been attempted in the literature before. As we are interested in the extremal dependence structure, we focus on methods for identifying the graph structure from the data. Two existing and one original method are used for this purpose. The research is topical due to the increase in the use of

biofuels and recent energy and food price shocks due to the invasion of Ukraine. The recent development of a general theory of conditional independence for multivariate Pareto distributions enables us to do this on a scale that has not been attempted before. It is found that tail dependence is the strongest for commodities of the same type and that interaction between groups is limited. Natural gas is the commodity type that has the strongest tail dependence with other commodity types as it serves as the main energy input in energy-intensive energy such as metal production and agriculture. The rest of the inter-group connection is limited, therefore supporting the neutrality hypothesis.

Section 2 discusses the current state of knowledge on the dependence between commodities and research on its extremes. This paper adds to this literature by increasing the number of considered commodities and utilising a model that makes use of extreme conditional independence and graphical models. This method is described in section 3. Three different approaches are taken to find the graphical structure and their characteristics and performance are compared in a simulation study in section 4. After that, the data set on commodity futures returns is introduced and filtered in section 5 in order to be used for the empirical application. The results of the empirical application are shown in section 6. Final conclusions are drawn in section 7 and some recommendations for future work are made.

2 Literature

A considerable amount of literature has been dedicated to the linkage between energy and agricultural commodity prices. Nonetheless, no univocal consensus has been established. Zhang et al. (2010) finds no direct long-run price relationship between different foods and fuels and limited if any short-run relations. This supports the so-called neutrality hypothesis. The data that are used spans from 1989 to 2008. However, Reboredo (2012), Kristoufek et al. (2012) and Yahya et al. (2019) show that the dependence between food and energy has increased after the global food price crisis 2006-2008 due to the demand increase in biofuels, even though only weak connections are found. Baumeister and Kilian (2014) treats May 2006 as the date of a tentative structural break between oil and agricultural prices due to the passing of the Energy Policy Act 2005, which promoted the use of ethanol and concludes that the link is demand-driven instead of being caused by higher production costs. Nazlioglu (2011) disagrees with the neutrality hypothesis and shows with nonlinear Granger causality methods that there is a persistent unidirectional causality from oil to corn and soybeans. A follow-up paper, Nazlioglu and Soytas (2012), further investigates the dynamic relationship between oil and 24 agricultural commodities using panel data. They find strong support for the role of oil prices on agricultural commodities, contrary to most literature around that time. They contribute their contradicting result to better use of information with their panel cointegration and causality analysis compared to previous individual time-series analyses. They state that even though it might not be possible to find a causal relation of a specific agricultural commodity to oil prices, a larger group of agricultural commodities may have a connection to oil prices. That explanation is further explored in this research by creating a large system of both agricultural and energy resources and products.

Regarding the tail dependence, Wang et al. (2014) find that the explanatory abilities of oil shocks on agricultural commodity prices have increased in the post-crisis period and are greater than that of aggregate demand shocks. Ji et al. (2018) conclude that lower tail dependence between oil and food prices is much stronger in bearish regimes indicating systematic risk spillovers during extreme downwards movements. On the link between large oil price movements and industrial and precious metals, Reboredo and Ugolini (2016) find spillover effects on all the metals. They also find that upward oil price movements were larger than downwards movements. Albulescu et al. (2020) study extreme dependencies among energy, agriculture and metal commodities indices by applying a copula-based Kendall's tau approach. They find stronger dependence between energy and other commodities in lower tails.

Although there has been extensive research on this topic, all of these studies, except Nazlioglu and Soytas (2012), use only a limited amount of commodities. The main reason is to keep the models parsimonious. However, Engelke and Hitz (2020) recently developed a general theory of conditional independence for multivariate Pareto distributions that enables the definition of graphical models and sparsity for extremes. This allows us to increase dimensions considerably.

3 Methodology

The methodology that is followed is presented in this section. To be able to use futures returns series for this extreme value theory model, the autocorrelation of the returns and squared returns has to be removed. The procedure for this is described in subsection 3.1.

To be able to use extremal graphical models, some theoretical background on extreme value theory and graphical models is needed. Both concepts are combined in the theory on extremal graphical models from Engelke and Hitz (2020). These three subjects are summarised in subsection 3.2.

The structure of the graph has to be found first to be able to apply these concepts. Identifying the graph is the focus of this research. Three methods are employed for that purpose: a forward greedy procedure that is introduced in the Engelke and Hitz (2020) paper, an extremal version of the graphical lasso that is also developed by Engelke in a yet-to-be-published work, and an original method developed to find a leaner and better interpretable graph from the empirical variogram that uses the Leiden community finding algorithm (Traag et al., 2019). The three methods are described in subsection 3.3. The Leiden community finding algorithm is elaborated upon in subsubsection 3.3.4. The implementation of this methodology is done in R and is built upon the package GRAPHICALEXTREMES published on the *Github* page of Sebastian Engelke.

The notation used in this thesis is based on Engelke and Hitz (2020). Bold symbols are used for column vectors and the components are denoted with a subscript, e.g. $x_i, i \in \{1, .., d\}$ is a component of $\mathbf{x}$. Uppercase letters are used for random variables and lowercase for deterministic values.

3.1 Data filtering

To study the behaviour of commodity prices, we consider the returns of commodity futures. These data are easy to obtain and span a large array of real-world market prices. Financial time series, such as these, exhibit some particular characteristics called stylised facts. These include profound serial correlation of absolute returns, heteroscedasticity and clustering of extreme returns (McNeil et al., 2015). This complicates its analysis as numerous theories, inter alia extreme value theory, rely on the assumption of independent and identically distributed random variables. A way to mitigate this obstacle is to use ARMA-GARCH filtering, as proposed by McNeil and Frey (2000). The goal is to have a variable which can be used to apply extreme value theory. It uses the univariate GARCH framework, first developed by Bollerslev et al. (1992), and is given as

$$\begin{aligned} X_i^{(t)} &= \mu_i + \sum_{j=1}^{p} \phi_j X_i^{(t-j)} + \sum_{k=1}^{q} \theta_k \varepsilon_i^{(t-k)} + \varepsilon_i^{(t)} \\ \varepsilon_i^{(t)} &= \sigma_i^{(t)} Z_i^{(t)} \\ \left(\sigma_i^{(t)}\right)^2 &= \omega_i + \sum_{j=1}^{r} \alpha_j \left(X_i^{(t-j)}\right)^2 + \sum_{k=1}^{s} \beta_k \left(\sigma_i^{(t-k)}\right)^2, \end{aligned}$$

where $Z_i^{(t)}$ is independent and identically distributed. $Z_i^{(t)}$ corresponds to the initial shock in an extreme event. We collect them for all dimensions as elements of vector $\mathbf{Z}$. The time notation is dropped here because this random vector is considered to be time-invariant. The cross-sectional dependence of its elements is what needs to be modelled in order to understand the price relations of commodities in extreme events. An assumption on the distribution of $Z_i^{(t)}$ has to be made. There are two approaches that are considered. Firstly, we assume the Student's t-distribution and estimate the degree of freedom ν as a parameter. In the second approach, we acknowledge the misspecification of the distribution as normal, but recognise that the estimation should still lead to a consistent estimator. The Gaussian likelihood should therefore not be seen as a proper likelihood, but more as an objective function. This procedure is called Quasi Maximum Likelihood (QML) (McNeil et al., 2015).

The number of lags in the ARMA-GARCH model is selected by considering the AIC of both distributions separately. The best specifications under both distribution assumptions are then are evaluated for goodness-of-fit using a Q-Q plot and the Jarque-Bera test. Finally, the serial correlation of the returns and squared returns are investigated with the correlogram and tested with the Ljung-Box test. The results of this procedure are shown in section 5 and the ARMA(1,0)-GARCH(1,1) model is selected.

The cross-sectional dependence of the elements of $\mathbf{Z}$ in the upper tail has to be modelled in order to understand extreme price relations. Due to the high number of dimensions of this vector, traditional methods fail. Instead, we use extremal graphical models which are more suitable for this type of data due to their visual representation and their ability to reduce the amount of the parameters that need to be estimated by using the concept of conditional independence. The background on extremal graphical models is summarised in the next part.

3.2 Background on extremal graphical models

The cross-sectional dependence of the initial shocks of futures returns, i.e. $\mathbf{Z}$, is what needs to be modelled in order to understand the price relations of commodities in extreme events. Extremal graphical models are a useful tool for this purpose as they allow for high-dimensional data and produce a visually interpretable output diagram. They lie at the interface of both extreme value theory and graphical models. Firstly an introduction is given on the former, followed by one on the latter. Only after that, it can be explained how these concepts are combined in Engelke and Hitz (2020). This background uses a different local notation compared to the other parts in this research as it spans a more general theory on extremal graphical models. That is done to conform with the notation used in most literature concerning these topics. While $\mathbf{X}$ indicates a general random vector is this part, we apply the models on the residual vector $\mathbf{Z}$ in this application.

3.2.1 Extreme value theory

Data on extreme events is scarce due to its very nature. This makes it difficult to use traditional estimators for building a model. For this reason, extreme value theory was developed. This branch of statistics focuses on the extreme deviations from the mean. There are two main approaches in this theory: block maxima and peaks-over-threshold.

The former approach studies the tail behaviour of multivariate data by looking at the componentwise maxima. Consider a d-dimensional random vector $\mathbf{X}$ which has independent copies $\mathbf{X}^{(i)}$, $i = 1, ..., n$. We denote the componentwise maxima with $\mathbf{M_n} = (M_{1n}, ..., M_{dn}) = (\max_{i=1}^{n} X_1^{(i)}, ..., \max_{i=1}^{n} X_d^{(i)})$. Under a linear normalisation with constants $b_{jn} \in \mathbb{R}$ and $a_{jn} > 0$, there is convergence towards a limiting distribution, such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_{jn} - b_{jn}}{a_{jn}} \leqslant x\right) = G_j(x) = \exp - (1 + \xi_j x)_+^{-1/\xi_j}, \quad x \in \mathbb{R},$$

where $z_+ = max(z, 0)$. G_j is called the generalised extreme value distribution, and the heavyiness of the tail is governed by the parameter ξ_j. When focused on the analysis of the dependence structure, one typically estimates the marginals first and then normalises them to standard Pareto.

The transformation is performed as $\check{X}_j = 1/\left\{1 - \hat{F}_j\left(X_j\right)\right\}$, where $\hat{F}_j$ is the empirical distribution of the j^{th} element. All observations for which none of the elements exceed a chosen threshold $\hat{F}^{-1}(p)$ are eliminated. Here $\hat{F}$ corresponds to the empirical distribution of all data points across vectors. The threshold p is set 0.85 in this research, as this yields a favourable balance between bias and variance in the estimator of the tail correlation. The results are validated for robustness by setting the threshold to 0.75 and 0.95 in the sensitivity analysis in section 6. The remaining data points after elimination are then normalised, i.e. $\tilde{X}_j = \frac{\check{X}_j}{\hat{F}^{-1}(p)}$, so that $\hat{F}^{-1}(p) = 1$. For our application, we denote the vector of transformed residuals as $\tilde{\mathbf{Z}}$.

It is said that the standardised vector $\tilde{\mathbf{X}}$ is in the max-domain attraction of the d-dimensional random vector $\boldsymbol{\mathcal{Z}}$ if for any $\mathbf{z}$

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{i=1,...,n} \tilde{X}_{i1} \leqslant nz_1, \ldots, \max_{i=1,...,n} \tilde{X}_{id} \leqslant nz_d\right) = \mathbb{P}(\boldsymbol{\mathcal{Z}} \leqslant \mathbf{z}).$$

If so, $\boldsymbol{\mathcal{Z}}$ is max-stable with standard Fréchet marginals $\mathbb{P}\left(\mathcal{Z}_j \leqslant z\right) = \exp\left(-1/z\right), z \geqslant 0$, and we say

$$\mathbb{P}(\boldsymbol{\mathcal{Z}} \leqslant \mathbf{z}) = \exp\left(-\Lambda(\mathbf{z})\right), \quad \mathbf{z} \in \mathcal{E},$$

where the exponent measure Λ is a Radon measure on the cone $\mathcal{E} = [0, \infty)^d \backslash \{\mathbf{0}\}$.

This approach has limited use in this application as max-stable distributions have been shown to lead only to trivial probabilistic structures for the notion of conditional independence by Papastathopoulos and Strokorb (2016). It is, however, useful to understand where the exponent measure Λ comes from, as this comes back in the other approach.

The latter approach in extreme value theory looks at the threshold exceedances. Resnick (2007) states that the convergence in subsubsection 3.2.1 is equivalent to

$$\lim u\{1 - \mathbb{P}(\mathbf{X} \leqslant u\mathbf{z})\} = \Lambda(\mathbf{z}), \quad \mathbf{z} \in \mathcal{E}, \quad \mathbf{z} \in \mathcal{E}, .$$

This can be extended to the multivariate distribution of threshold exceedances of $\mathbf{X}$ as

$$\mathbb{P}(\mathbf{Y} \leqslant \mathbf{z}) = \lim_{u \to \infty} \mathbb{P}\left(\frac{\mathbf{X}}{u} \leqslant \mathbf{z} \mid \|\mathbf{X}\|_\infty > u\right) = \frac{\Lambda(\mathbf{z} \wedge \mathbf{1}) - \Lambda(\mathbf{z})}{\Lambda(\mathbf{1})},$$

where $\|\mathbf{x}\|_\infty = \max_{i \in V} |x_i|$. Limiting random vector $\mathbf{Y}$ has a distribution which is called the multivariate Pareto distribution. u serves as a kind of threshold and is equal to one in the standard Pareto distribution.

We assume that the distribution of $\mathbf{Y}$ has a positive and continuous density $f_{\mathbf{Y}}$, which is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\partial^d}{\partial y_1 \dots \partial y_d} \mathbb{P}(\mathbf{Y} \leqslant \mathbf{y}) = \frac{\lambda(\mathbf{y})}{\Lambda(\mathbf{1})},$$

where $\Lambda(\mathbf{1})$ is a normalisation constant. This means that the density $f_{\mathbf{Y}}$ is proportional to the density λ of the component measure. For the density of λ, we introduce Hüssler-Reiss distributions.

Hüssler-Reiss distributions (Hüsler and Reiss, 1989) are a particular type of distributions that are parameterised by a symmetric, strictly conditionally negative defined matrix Γ. It has non-negative entries and zeros on the diagonals. The density of the component measure λ can be expressed for any $k \in \{1, .., d\}$ as (see Engelke et al. (2015))

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

where ϕ_p corresponds to the centred p-dimensional normal distribution and

$$\tilde{\mathbf{y}}_{\setminus k} = \left\{\log\left(y_i / y_k\right) + \Gamma_{ik}/2\right\}_{i=1,\dots,d}.$$

The strictly positive definite matrix $\Sigma^{(k)}$ is defined as follows

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

The entries of Γ, Γ_{ij}, correspond to the strength of the dependence between the i^{th} and j^{th} component (Engelke and Hitz (2020); Engelke and Volgushev (2020)). A value of zero corresponds to complete dependence and it goes to infinity for independence. This means that the dependence structure of the multivariate Pareto distribution is summarised in one matrix, similarly to the covariance matrix in a multivariate normal distribution. This makes them a natural analogue of Gaussian distributions in the world of asymptotically dependent extremes (Engelke and Hitz, 2020). While having all (extremal) dependence information summarised in one matrix is rather useful for computers, it is cumbersome for a human to interpret directly, especially for a high number of dimensions. For this reason, graphical models were developed. This type of graph consists of nodes and edges that form a network to represent the dependence structure. Traditionally, they were created for non-extreme data, but Engelke and Hitz (2020) show how they can be combined with the Hüssler-Reiss Pareto distribution to have an equivalent for threshold exceedances called extremal graphical models.

3.2.2 Graphical models

The dependence structure between a set of variables can be summarised in a graphical model. A graph $\mathcal{G}$ consist of a set of vertices or nodes V, corresponding to variables, and a set of edges E, connecting the vertices. The absence of an edge between two nodes indicates conditional independence. For Figure 1, this means that X_1 is independent of X_2 if conditioned on X_3. This is denoted as $X_1 \perp X_2 \mid X_3$. An undirected graph does not give any orientation to the edges, i.e. edge (i, j) is the same as (j, i) for $i, j \in V$. This type of graph is the only one used throughout this research.

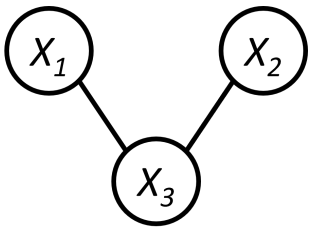


Figure 1: Example of an undirected graph

A random vector $\mathbf{X}$, corresponding to graph $\mathcal{G} = (V, E)$, has elements $X_i, i \in V$ that have a continuous state space $\mathcal{X}_i \subset \mathbb{R}$. $\mathbf{X}$ takes values in the Cartesian product $\mathcal{X} = \times_{i \in V} \mathcal{X}_i$.

If the random vector $\mathbf{X}$ has a positive and continuous Lebesgue density $f_{\mathbf{X}}$, it follows from the Hammersley-Clifford theorem that the density is factorised as

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{C \in \mathcal{C}} \psi_C \left(\mathbf{x}_C \right), \quad \mathbf{x} \in \mathcal{X},$$

for suitable functions ψ_C on $\times_{i \in C} \mathcal{X}_i$.

A decomposable graph is a graph that can be successively decomposed into its cliques (Lauritzen, 1996). A clique is a subset of nodes where each distinct pair of nodes is connected to each other. If the graph is decomposable, then the factorisation can be written in terms of marginal densities of the cliques C, that are part of the clique set $\mathcal{C}$ and separators D for the separator set $\mathcal{D}$. This is done as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\prod_{C \in \mathcal{C}} f_C \left(\mathbf{x}_C \right)}{\prod_{D \in \mathcal{D}} f_D \left(\mathbf{x}_D \right)}, \quad \mathbf{x} \in \mathcal{X}.$$

Gaussian graphical models correspond to the graphs of multivariate normal distributions. The edge set can be derived from the precision matrix Θ for a random vector $\mathbf{X}$ that follows $N(\mu, \Sigma)$. The precision matrix is defined as the inverse of the covariance matrix Σ. If Θ_{ij} is zero, we know that there is no edge between node i and j. This makes it easier to interpret the dependence structure of a high-dimensional data set, as we can directly translate a summary statistic, in this case the covariance matrix, to a visual representation. Extremal graphical models have the same idea, but for threshold exceedances. Although there are a few more steps involved, they allow for the same easy interpretation as the Gaussian graphical models.

3.2.3 Extremal graphical models

The concept of conditional independence of extremes allows the construction of graphical models, such as in the case of Gaussian graphical models. This gives three major advantages. Firstly, the tail dependence structure of a model can be shown in a graph that is easily interpretable. That facilitates visual analysis of the results for high dimensional data. Secondly, sparsity can be introduced, which allows for a more parsimonious and manageable model in higher dimensions. Thirdly, we can estimate the parameters of the model separately on lower-dimensional subsets of the data.

As explained above, the graph can be directly derived from the precision matrix Θ of the multivariate normal distribution for Gaussian graphical models. For the Hüsler–Reiss Pareto distribution, it is derived from the precision matrix $\Sigma^{(k)}$ in Equation 1. Engelke and Hitz (2020) state that it is done as follows

$$Y_i \perp_{\mathrm{e}} Y_j \mid \mathbf{Y}_{\setminus\{i,j\}} \Leftrightarrow \begin{cases} \Theta^{(k)}_{ij} = 0, & \text{if } i,j \neq k, \\ \sum_{l\neq k} \Theta^{(k)}_{lj} = 0, & \text{if } i = k, j \neq k, \\ \sum_{l\neq k} \Theta^{(k)}_{il} = 0, & \text{if } j = k, i \neq k, \end{cases} \tag{3}$$

where $Y_i \perp_{\mathrm{e}} Y_j \mid \mathbf{Y}_{\setminus\{i,j\}}$ indicates conditional independence between vertex i and j, i.e. the absence of edge (i,j).

To be able to estimate the parameters of the model separately on lower-dimensional subsets of the data, we only consider block graphs, which are a class of decomposable graphs. These connected graphical structures consist of cliques that are separated by singleton separator sets. Block graphs are especially suitable in our application as we expect that commodities within the same type to be closer connected to each other than to another type, e.g. corn and wheat are more closely related than corn and copper. Engelke and Hitz (2020) show that we can factorise the d-variate Pareto distribution according to the graph as

$$f_{\mathbf{Y}}(\mathbf{y};\theta) = \frac{1}{\Lambda(\mathbf{1};\theta)} \prod_{C\in\mathcal{C}} \frac{\lambda_C\left(\mathbf{y}_C;\theta_C\right)}{\prod_{j\in C} y_j^{-2}} \prod_{i\in V} y_i^{-2}, \quad \mathbf{y}\in\mathcal{L}.$$

This allows us to estimate the parametric family of each clique (θ_C) separately. Each of these cliques has a lower-dimensional Hüsler-Reiss distribution of its own, characterised by the variogram matrix Γ_C. The parameters of each clique can be estimated using the maximum likelihood estimation of the clique log-likelihood. Although this in theory leads to a decrease in estimation efficiency compared to maximising the joint likelihood, Engelke and Hitz (2020) show that this difference is rather small. In addition, using clique likelihood estimation allows for increasing the dimension of the model to a higher level as long as the clique size remains small.

Not all components of $\mathbf{X}$ may have been converged to the limiting distribution $\mathbf{Y}$. In order to avoid biased estimates of the dependence parameters θ_C, we can use censoring on the data (Smith et al. (1997); Ledford and Tawn (1997)). This practice uses a modified likelihood contribution to only include the information that a threshold is not exceeded by a certain component instead of using its exact value. Let $\mathbf{X_C}$ be a data point corresponding to the clique $C\in\mathcal{C}$ for which $\|\mathbf{X}_C\|_\infty > u$, where u is sufficiently high.

J is the set of indices $j \in C$ that does not exceed the threshold u. The censored likelihood contribution is defined as

$$f_C^{\text{cens}}\ (\mathbf{y}_C; \theta_C) = \int_{[0,1]^{|J|}} f_C\left(\mathbf{y}_C; \theta_C\right) \mathrm{d}\mathbf{y}_J, \quad \mathbf{y}_C \in \mathcal{L}_C.$$

We use this likelihood contribution to estimate the parameters θ_C using maximum likelihood estimation.

The theory up to now assumes that we know the structure of the graph, so that we can factorise the likelihood into the likelihood of the cliques and separators. In practice, however, this has to be found separately. The focus of this research is on three methods to find the extremal graphical structure from the data on futures returns.

3.3 Finding the extremal graphical dependence structure from futures return data

In order to fit the parameters using maximum likelihood estimation and to draw conclusions on the tail dependence network, we need to know the structure of the graph. Although it is possible to estimate the variogram from the data directly with Equation 6, only fully connected graphs are found from this matrix. Sparsity has to be introduced in order to use the concept of conditional independence effectively. Despite the fact that there are multiple methods for the Gaussian case, there are only a limited set of options for the extremal case due to the novelty of this theory. This research focuses on three methods and compares their findings and performances. From this part onward, we focus on the application of graphical models on commodity futures return data and switch, therefore, back to the notation used in subsection 3.1.

3.3.1 Forward greedy

The first method uses the minimum spanning tree as a starting point and adds edges using a greedy forward search based on the BIC.

As a starting point we use the minimum spanning tree, which is defined as the tree that minimises the sum of distances or weights on that tree. Often, it is used to minimise the costs of building a kind of network that connects a set of nodes, e.g. laying cabling to connect homes in a neighbourhood. However, with a proper choosing of the weights, it can serve as the backbone for building the extremal graphical structure. It is defined as

$$\mathcal{T}_{\text{mst}} = \underset{\mathcal{T}=(V,E)}{\arg\min} \sum_{(i,j)\in E} w_{ij}. \tag{4}$$

This minimisation problem is solved in this implementation by the greedy algorithm of Prim (1957).

It is crucial to select suitable weights so that we can recover the true underlying tree structure corresponding to the conditional independence relations. In Engelke and Hitz (2020), the negative maximised bivariate-loglikelihood are used as edge weights. This has, however, two drawbacks. Firstly, for higher

dimensions d it can become severely costly to compute d^2 likelihood optimisations. Furthermore, a set of parametric bivariate models has to be chosen advance (Engelke and Volgushev, 2020). As a solution, Engelke and Volgushev (2020) propose to use the elements of extremal variogram Γ, as $\Gamma_{ij}^{(m)}$ can be interpreted as the distance between Y_i^m and Y_j^m, where $\mathbf{Y}^m$ is defined as $\mathbf{Y}$ conditioned on the event that $\{Y_m > 1\}$ for root node m. This distance is large when Y_i^m and Y_j^m are weakly dependent and vice versa. $\Gamma_{ij}^{(m)}$ can be estimated with the GARCH-filtered data as

$$\hat{\Gamma}_{ij}^{(m)} := \widehat{\mathrm{Var}}\left(\log \tilde{Z}_i - \log \tilde{Z}_j : \log \tilde{Z}_m > 1\right). \tag{5}$$

The results for all root nodes m is averaged out, i.e.

$$\widehat{\Gamma} = \frac{1}{d}\sum_{m=1}^{d} \widehat{\Gamma}^{(m)}. \tag{6}$$

The elements of $\widehat{\Gamma}$ are used as weights for the minimum spanning tree, i.e.

$$\widehat{w_{ij}} = \widehat{\Gamma}_{ij}$$

They have shown that this approach can consistently recover the true underlying tree.

A tree, however, is not likely to describe the dependence structure sufficiently for higher dimensional data. More edges can be added to form a block graph. Edges are added one by one and it is checked whether the graph is still a block graph with a maximum clique size of three. This limit is set to limit the number of potential edges and reduces the computational load for loglikelihood calculations. The pseudocode for this approach is given in algorithm 1.

Algorithm 1 Forward greedy algorithm

```
Calculate estimator Γ̂ from data
Assign weights w_ij = Γ̂_ij to each edge (i, j), i, j ∈ V, i ≠ j
Find T_mst by solving the minimisation problem in Equation 4
Compile list of possible new edges P for which graph G is block graph with maximum clique size ≤ 3
while P ≠ ∅ do
    for edge (i, j) in P do
        Add (i, j) to G form an extended graph G_(i,j)
        if G_(i,j) is block graph then
            Calculate BIC for G_(i,j) using the clique loglikehood
        end
    end
    Keep edge with lowest corresponding BIC, i.e. G = arg min_{G_(i,j)} BIC
    Remove that edge from P
end
```

This heuristic method is adopted from the original work on extremal graphical models from Engelke and Hitz (2020), with the alteration that it uses BIC instead of AIC. This is done to increase the penalty for adding edges to find sparser graphs.

The forward greedy method has been shown to provide decent results. However, it can be quite slow for high dimensions as many edges have to be tried and it does not offer a way to control the sparsity of the graph explicitly. An alternative to this method is also developed by Sebastian Engelke and is called extremal graphical lasso.

3.3.2 (Reconnecting) extremal graphical lasso

The second method forces sparsity on the off-diagonal elements precision matrix $\Sigma^{(k)}$ for all k and uses it to construct the variogram Γ in order to recover the graph.

The extremal variogram is the summary statistic for the tail dependence of a vector for Hüssler-Reiss distributions in a similar way as the precision matrix is for dependence in a multivariate normal distribution. It reveals the underlying graphical structure. The empirical extremal variogram can be estimated from the data using Equation 6 and Equation 5. In practice, however, the result rarely has any zero entries in the matrix. This leads to a fully connected graph. Introducing sparsity to this matrix has some advantages. It allows for an easier interpretation of the graph, can factorise the densities in lower-dimensional margins and yields more efficient inference (Engelke and Ivanovs, 2021). A well-established method in Gaussian graphical models is to use graphical lasso (Friedman et al., 2007). It maximises the penalised likelihood of the precision matrix Θ as

$$\widehat{\Theta}_\rho = \underset{\Theta \geq 0}{\operatorname{argmax}} \log\det\Theta - \operatorname{tr}\left(\widehat{\Sigma}\Theta\right) - \rho\|\Theta\|_1,$$

where ρ corresponds to the penalising parameter and $||.||_1$ to the l_1 norm of the matrix. This pulls the off-diagonal elements towards zero, which corresponds to the independence of the corresponding vector elements. A larger ρ leads to a sparser precision matrix, which translates to fewer edges.

A similar method for Hüssler-Reiss distributions has been developed by Engelke, Lalancette and Volgushev in a yet-to-be-published paper, called Extremal Graphical LASSO (eglasso). It maximises the penalised likelihood of the precision matrix given by

$$\widehat{\Theta}_\rho^{(k)} = \underset{\Theta \geq 0}{\operatorname{argmax}} \log\det\Theta - \operatorname{tr}\left(\widehat{\Sigma}^{(k)}\Theta\right) - \rho \sum_{\substack{i\neq j \\ i,j\neq k}} |\Theta_{ij}|\,. \tag{7}$$

There is, however, a slight complication compared to the graphical lasso for Gaussian graphical models. While the whole extremal graphical structure is revealed by each $\Theta^{(k)}$, independent of the choice of the root node k, the connections with the root node are not revealed by zero entries. Instead, they correspond to zero row sums (see Equation 3). This condition is not enforced by the extremal graphical lasso. This would mean that we can not introduce sparsity on the connections with the root node if we would only

use one choice of k. As a solution to this problem, they suggest looping through all possible choices of root nodes and deciding on the connections by a simple majority vote

$$(i,j) \in \widehat{E}_\rho \Leftrightarrow \quad \frac{1}{d-2} \# \left\{ k \in V \backslash \{i,j\} : \left(\widehat{\Theta}_\rho^{(k)} \right)_{ij} \neq 0 \right\} \geq 1/2.$$

A potential problem that is observed with the eglasso algorithm, proposed by Engelke, Lalancette and Volgushev, is that for high values of ρ the graph may become disconnected. This is a problem as finding the loglikelihood requires a connected graph. That puts a limit on the amount of sparsity that can be introduced with their method. For that reason, I propose an augmentation of their approach called Reconnecting Extremal Graphical LASSO (reglasso). This method adds the edge that connects two disconnected subgraphs corresponding to the smallest element in variogram Γ. This is repeated until the graph is connected again. This allows us to use extremal graphical lasso for a higher ρ. The full algorithm for reglasso is given in algorithm 2.

Algorithm 2 Reconnecting extremal graphical lasso

Calculate estimator $\hat{\Gamma}$ from data

Set voting matrix $\mathcal{V}$ to null matrix

for root node $k \in V$ **do**

Calculate $\hat{\Sigma}^{(k)}$ from $\hat{\Gamma}$ using Equation 2

Estimate sparse precision matrix $\hat{\Theta}_\rho^{(k)}$ from $\hat{\Sigma}^{(k)}$ using extremal graphical lasso (Equation 7)

Count near-zero elements in $\hat{\Theta}_\rho^{(k)}$, i.e. $\mathcal{V}_{ij} = \mathcal{V}_{ij} + I\{| \left(\widehat{\Theta}_\rho^{(k)} \right)_{ij} | < 1e-10\}$, $i,j \in V, i \neq j$

end

Add edge (i,j) to graph $\mathcal{G}$ if $\frac{\mathcal{V}_{ij}}{d-2} < 1/2$, $i,j \in V, i \neq j$

while $\mathcal{G}$ is disconnected **do**

Add edge $(i,j) = \underset{(i,j)}{\operatorname{argmin}} \widehat{\Gamma}_{ij}$, $i,j \in V, i \neq j$, i and j are not connected

end

ρ offers a way to control the amount of sparsity in the graph, a higher ρ corresponds to a graph with fewer edges. This is an improvement compared to the greedy forward procedure, which lacks this flexibility. The algorithm is repeated for different values of ρ. Each value of ρ corresponds to a different graph and thus a different model. The BIC is calculated for each graph and the one with lowest value is selected. The va;ues that are used in this analysis range from 0.01 to 0.15.

Although extremal graphical lasso has made an improvement in terms of flexibility compared to forward greedy procedure, it lacks its easy interpretable nature due to the indiscriminate enforcement of sparsity. This yields graphs with many edges and nodes that are not clustered in groups. Therefore, we would like to have a method that has both characteristics: explicit control of sparsity and a simple and interpretable graph.

3.3.3 Clustering algorithm

The third method finds closely-connected nodes with a community-finding algorithm called the Leiden algorithm and fully connects them as cliques. The disconnected cliques are sewn together using the strongest single connection between cliques based on the empirical variogram.

Extremal graphical lasso has multiple improvements with respect to the forward greedy procedure. It does not require the minimum spanning tree, has more flexibility due to the ρ parameter and is faster than the greedy algorithm. However, it also has some drawbacks. Reglasso is indiscriminate in introducing sparsity. It, therefore, does not promote the formation of decomposable graphs, which facilitates the use of clique likelihood estimation. This makes it slower to estimate the parameters of the model and reduces the estimation efficiency. In addition to that, it makes it more difficult to interpret the graph as the connections are not clustered. For those reasons, I propose an heuristic approach called the clustering algorithm. This method is focused specifically on forming and connecting cliques in order to leverage the potential of clique likelihood estimation. This yields a better interpretable graph that shows different groups of nodes and how they are connected.

It identifies the cliques in a block graph by finding clusters of strongly connected nodes using a community finding algorithm. A community is a set of nodes that is more strongly connected to nodes of the same community than to others. This implementation uses the Leiden algorithm (Traag et al., 2019) for this purpose. It is described in subsubsection 3.3.4. The algorithm uses the weights of the edges in the graph to find communities. The larger the weight of a certain edge is, the stronger the connection between its nodes is and, therefore, the more likely it is that two nodes belong to the same community. For this reason, the inverse of the elements in variogram Γ are used as weights. This is in contrast to finding the minimum spanning tree where the element itself is used. The reason for this is that the community finding algorithm is a maximisation problem, while finding the minimum spanning tree is a minimisation problem.

Cliques are formed from the found communities leading to a collection of separate fully-connected subgraphs. These subgraphs have to be sewn together to have a connected graph that can be used for the application of extremal graphical models. It proceeds by connecting every subgraph with the strongest potential separator that connects two previously disconnected subgraphs until the graph is connected. A group of connected cliques forms a new subgraph. A connected graph has only one subgraph. The strength of the connection is assessed by considering the corresponding element in the empirical variogram $\widehat{\Gamma}$. The full algorithm is given in algorithm 3.

Algorithm 3 Clustering algorithm

Calculate estimator $\hat{\Gamma}$ from data

Assign weights $w_{ij} = \frac{1}{\hat{\Gamma}_{ij}}$ to each edge (i, j), $i, j \in V, i \neq j$,

Divide vertices into communities based on w_{ij}, with the Leiden algorithm (Traag et al., 2019)

Form complete subgraphs from communities, i.e. form (unconnected) cluster graph $\mathcal{G}$

Collect unconnected subgraphs of $\mathcal{G}$ in set $\mathcal{U}$

Collect all potential separators in set $\mathcal{S}$

while $\mathcal{G}$ is disconnected **do**

  Add largest $w_{ij} \in \mathcal{S}$ for which i and j are not in the same subgraph

  Update $\mathcal{U}$

  Remove (i, j) from $\mathcal{S}$

end

The size of the clusters is determined by the resolution parameter γ, of which its origin is explained in subsubsection 3.3.4. A lower resolution leads to larger cliques and, therefore, more edges. The algorithm is repeated for different values of γ. Each value of γ corresponds to a different graph. The BIC is calculated for each graph and the one with lowest value is selected. The values range from 0.01 to 1.00.

Similarly to reglasso it does not require the minimum spanning tree, has flexibility due to a parameter controlling the sparsity, but has to additional benefit of labelling groups of nodes and forming decomposable graphs. This yields an easily interpretable graph. It is built upon a famous algorithm in network analysis that is called the Leiden algorithm. The workings of this algorithm are explained next.

3.3.4 Communities and Leiden algorithm

The clustering algorithm uses a community finding algorithm, called the Leiden algorithm, to identify the cliques in the graph. The operation of this algorithm is explained here.

In real-world network structures, it is possible that nodes in a particular group are stronger linked to each other than those of other groups, e.g. groups of friends in a social network, functionally related proteins in a protein network and commodity types in a network of trade resources (Palla et al., 2005). Those groups are called communities. It is useful to identify communities in a large networks as it helps for the interpretability of the network. Multiple strategies exist for this purpose, e.g. Newman and Girvan (2004), Blondel et al. (2008) and Traag et al. (2011). One popular algorithm is called the Leiden algorithm (Traag et al., 2019). It is an iterative algorithm that is fast and uncovers good partitions in the network.

The algorithm consists of three phases that are iterated until a local optimum is reached. The phases are: (1) local moving of nodes (2) refinement of partition (3) aggregation of partition.

We start with the singleton partition, i.e. the partition where every node has its own community. In phase (1), a queue of all nodes in random order is made. The first node is removed from the queue and added to the community which yields the highest increase in the objective or quality function $\mathcal{H}$. The

objective function is the modularity and is given by

$$\mathcal{H} = \frac{1}{2m} \sum_{ij} (w_{ij} - \gamma n_i n_j) \delta(\sigma_i, \sigma_j),$$

where m is the total edge weight, γ the resolution parameter, n_i the degree, of node i, i.e. the number of connections to other nodes, σ_i the community of node i and $\delta(.,.)$ the Kronecker delta function, which is one if both inputs have the same value, and zero otherwise. The weights w_{ij} are the inverse of the corresponding elements of variogram, as is explained above. The size of the communities is governed by resolution parameter γ. It functions as a kind of threshold in that the communities should have a density of at least γ. Lower resolutions lead to fewer and larger communities and vice versa. If the node is moved to a different community, all the neighbours of the node that do not yet belong to this community, are added to the rear of the queue. The steps are repeated until the queue is empty. The local moving of nodes corresponds to step a) in Figure 2.

In phase (2), the partition is set again to the singleton partition. Nodes that belonged to the same community in phase (1) are merged randomly if they yield an increase in the quality function $\mathcal{H}$. The larger the increase in $\mathcal{H}$, the more likely they are to merge. Randomness allows for a broader exploration of the solution space and is controlled with the hyperparameter β, which is set to 0.01, as this is the default setting of the package and yields decent results. This phase corresponds to step c) in Figure 2.

In phase (3), we aggregate the communities found in phase (2) into nodes and assign them to same communities as the ones found phase (1). This is step d) in Figure 2. After that, the phases are repeated until no improvements can be made, i.e. no movement of the nodes yields an increase in the quality function $\mathcal{H}$.

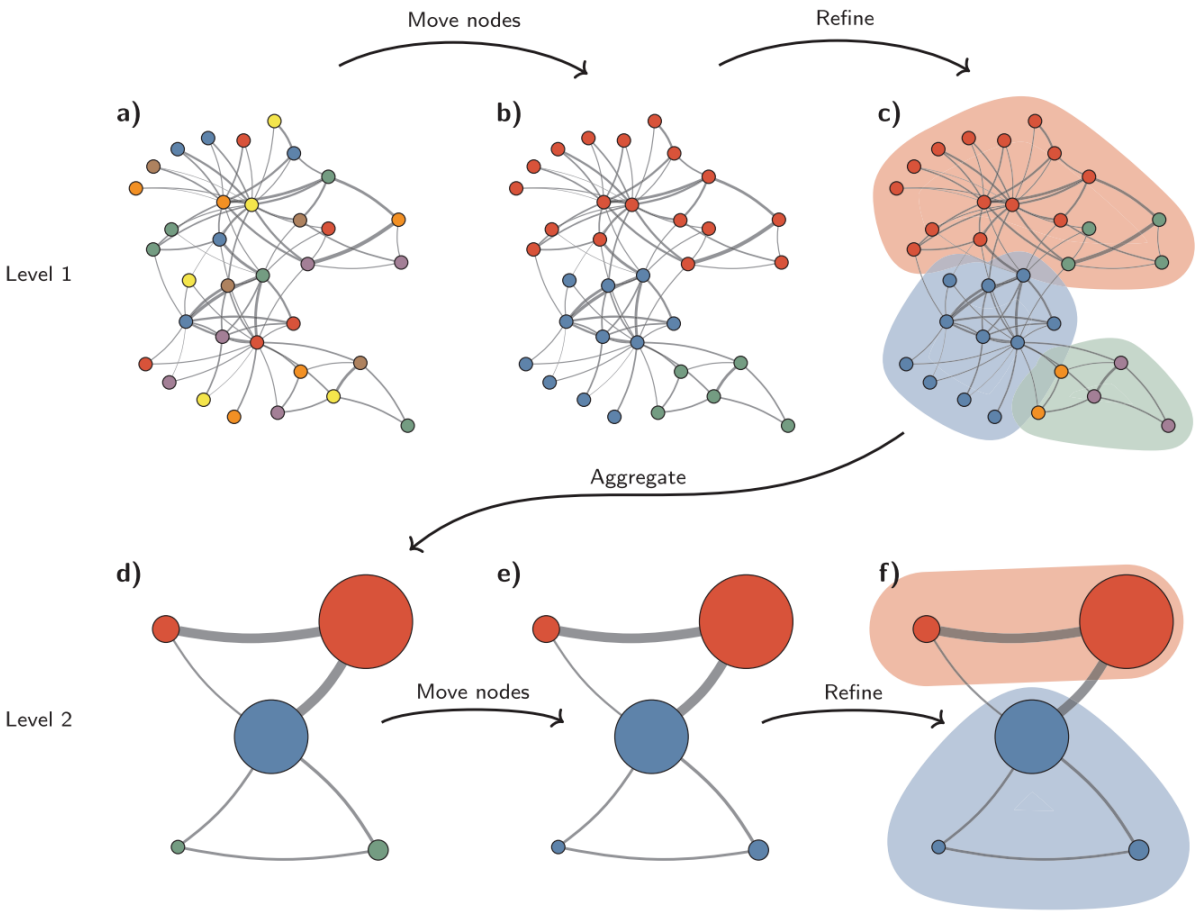


Figure 2: Visualisation of the Leiden community-finding algorithm. Reprinted from Traag et al. (2019)

Aside from finding the graphical structure in a new way using the newly proposed clustering algorithm, it also helps with interpreting the graph as it groups sets of nodes in communities.

The three methods to find the graphical structure should be compared how they perform relative to each other. The performance measures used for this are explained next.

3.3.5 Performance measures

The different approaches are compared for the correctness of the structure and data fit. Five measures in total are used for the evaluation total number of edges, number of not-found edges, number of faulty edges, Bayes information criterion and mean absolute percentage error with respect to the true tail correlation matrix. Different combinations of these metrics are used depending on the type of analysis.

The exact structure is known in a simulation study. This makes it possible to evaluate for each approach whether certain edges are not found in the estimation that is there in the exact graph. This is the number of not-found edges. Similarly, it can be assessed if there are edges found in the estimation that are not there in the exact graph. These two metrics indicate how accurate the estimation of the model was under a certain approach and whether it tends to under-or overestimate connections between nodes. In the empirical analysis the true graph is not known. We use the total number of edges instead to give an indication on the complexity of the graph.

The fourth measure is the Mean Absolute Percentage Error of the tail correlations. One of the most important pieces of information from the model is its implied tail correlation matrix. It gives the bilateral tail coefficient or tail correlation between each of the variables and is derived for each element from the model as

$$\chi_{ij} = 2 - \Lambda_{ij}(1,1).$$

For a model to be effective, we want this measure to be close to the true correlation coefficients. In the empirical setting, it can be estimated for a threshold u as

$$\widehat{\chi}_{ij} = \frac{\sum_{k=1}^{n} I\left\{\tilde{Z}_i^{(k)} > u, \tilde{Z}_j^{(k)} > u\right\}}{\sum_{k=1}^{n} I\left\{\tilde{Z}_j^{(k)} > u\right\}}$$

It is difficult to estimate these measures using a direct empirical estimator due to the small sample sizes of the extreme correlations. This effect is shown in section 4 where the empirical estimator is compared to the exact one.

The last measure is the Bayes Information Criterion (BIC). This metric gives a value for the model fit of the data while correcting for the increased likelihood due to additional parameters. This metric is only used for the empirical application because each generated data set in the simulation study has a different loglikelihood than the previous one. Its formula is

$$\text{BIC} = p\ln(n) - 2L\left(\hat{\theta}; \tilde{\mathbf{Z}}^{(1)}, \ldots, \tilde{\mathbf{Z}}^{(n)}\right)$$

where p is the number of parameters, n is the number of observations and the second term is twice the negative censored log-likelihood.

These are the metrics that are used in the simulation study and empirical analysis in the next sections. The procedure to generate random samples in the simulation study is explained in the last part of the methodology.

3.4 Simulation study

In order to compare the graph identification capabilities of the three methods, we perform a simulation study. A sample is generated of the limiting variable $\mathbf{Y}$ from the Hüssler-Reiss Pareto distribution for a known graph with a known variogram. We simulate $\mathbf{Y}$ and not $\tilde{\mathbf{Z}}$, because we assume for the empirical analysis that $\tilde{\mathbf{Z}}$ has sufficiently converged to the limiting distribution and we are more interested in the graph discovery characteristics of the three methods. This assumption is validated in the empirical analysis by doing a sensitivity analysis on the threshold. The simulation procedure is the same that is used in Engelke and Hitz (2020) and is implemented in the R package GRAPHICALEXTREMES. The pseudocode for the algorithm is given in algorithm 4.

Algorithm 4 Generation of random sample from Hüssler-Reiss Pareto distribution

Assign empty sample $\mathcal{S} = \emptyset$
Sample n integers from $[1, d]$ in $\mathbf{A}$
for k in V **do**
  Assign $n_k = \sum_{i=1}^{d} I\{A_i = k\}$
  if $n_k > 0$ **then**
    Calculate $\tilde{\Sigma}^{(k)}$ from variogram Γ using Equation 8
    for m in $1 : n_k$ **do**
      Generate the d-dimensional vector W^k, $W^k \sim N(0, \tilde{\Sigma}^{(k)})$
      Generate U, $U \sim \text{Unif}(0, 1)$
      $\tilde{Y}^k = \exp\left(W^k - \text{diag}(\tilde{\Sigma}^{(k)})/2\right)$
      $Y^k = \frac{\tilde{Y}^k}{\sum_i^d \tilde{Y}^k} \frac{1}{U}$
      Add Y^k to $\mathcal{S}$
    end
  end
end

It makes uses of $\tilde{\Sigma}^{(k)}$, which is similar to $\Sigma^{(k)}$, but includes the k^{th} row and column. It is obtained from the variogram Γ as

$$\tilde{\Sigma}^{(k)} = \frac{1}{2}\left\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\right\}_{i,j \in V} \in \mathbb{R}^{d \times d}. \tag{8}$$

This simulation procedure allows us to compare the three methods in the simulation study.

4 Simulation study

A simulation study is performed to analyse the performance and characteristics of the three different approaches. We want to evaluate two requirements for each of the methods.

1. The method has to find the correct underlying graph, rated by the number of Edges that are Not Found (ENF) and the number of Faulty Edges that are added to the graph (FE).
2. The method has to be able to predict the tail correlations more accurately than the empirical estimator, rated with Mean Absolute Percentage Error with respect to the exact tail correlations matrix (MAPE).

For this purpose, four graphs with 10 vertices and a corresponding variogram are created. The four graphs are the following:

1. **Tree**: a connected graph without cycles. It has a unique path between any two vertices.
2. **Block** graph (or clique tree): a connected graphs and decomposable graph, consisting of cliques that are separated by single nodes
3. **Undecomposable** (or non-chordal) graph: a connected graph that can not be decomposed into smaller subgraphs
4. (Connected) **cluster** graph: a connected graphs and decomposable graph, consisting of cliques that are separated by single edges

For each of the four models, 800 data points are generated using the procedure described in subsection 3.4. This data is used to estimate the model again. We do this for 100 simulations in order to calculate the average and standard deviation of the performance measures and to observe how many times each method is able to find the edges.

Table 1 shows the mean and standard deviation for the number of Edges Not Found (ENF) in the model compared to the true graph and the number of Faulty Edges (FE) in model compared to the true graph. This corresponds to requirement 1. Table 2 shows the mean and standard deviation for the Mean Absolute Percentage Error (MAPE) of model-implied tail correlation matrix (χ) compared to the true tail correlation matrix.

Table 1: Number of Edges Not Found (ENF) in the model compared to the true graph (left). Number of Faulty Edges (FE) in model compared to the true graph (right). Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

ENF	Greedy	Reglasso	Clustering
Tree	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Block	8.4 (1.3)	1.6 (1.2)	3.7 (0.4)
Undecomposable	4.5 (0.5)	2.0 (0.3)	3.1 (1.0)
Cluster	9.8 (1.0)	2.6 (0.6)	3.0 (0.0)

FE	Greedy	Reglasso	Clustering
Tree	0.9 (0.3)	0.2 (0.4)	1.8 (3.5)
Block	0.6 (0.7)	6.2 (2.4)	2.7 (2.4)
Undecomposable	2.4 (0.9)	1.1 (1.1)	6.3 (3.9)
Cluster	0.7 (0.5)	3.0 (2.4)	1.0 (1.7)

Table 2: Mean Absolute Percentage Error (MAPE) of model-implied tail correlation compared to the true tail correlation. Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

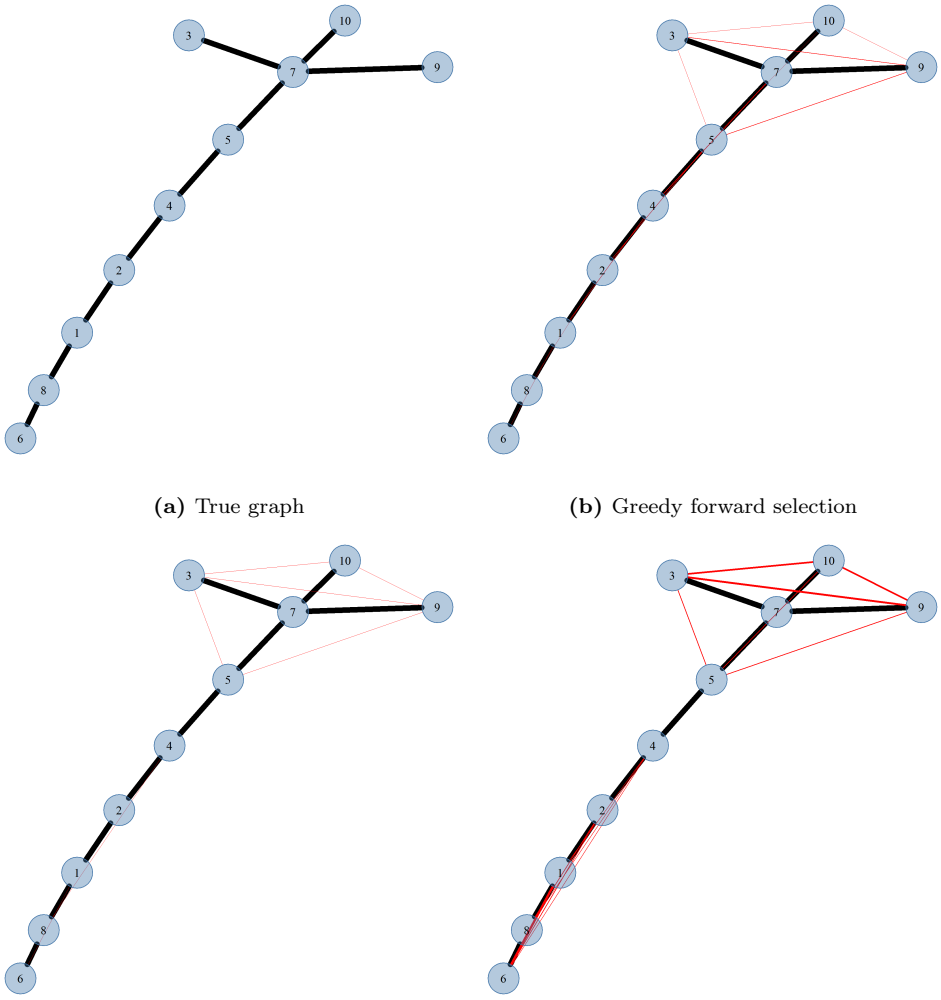
MAPE	Empirical	Greedy	Reglasso	Clustering
Tree	3.7 (1.2)	1.2 (0.6)	1.2 (0.7)	1.2 (0.6)
Block	2.1 (0.7)	1.6 (0.2)	0.9 (0.3)	1.7 (0.2)
Undecomposable	1.8 (0.7)	2.8 (0.4)	0.6 (0.2)	3.3 (1.3)
Cluster	5.5 (1.9)	17.4 (1.6)	2.3 (1.1)	12.8 (1.0)

Tree

Figure 3 shows the true graph and the graphs found by each of the three approaches for the tree. The black edges correspond to edges in the true graph and the red edges to edges that are found in the estimation, but are not part of the true graph. The thickness of the edge indicates in how many simulations it is found.

The full true tree is recovered by each of the three approaches in all of the simulations. This makes it possible to estimate the tail correlations in an efficient way. Table 2 shows that that all three methods are able to estimate the true tail correlations with an average absolute error of close to 1%. That is considerably better than the empirical estimator of the tail correlation, which has an average absolute error of 3.7%.

However, we see a difference between the methods in the number edges that are incorrectly included in the estimated graph. While reglasso only includes 0.2 wrong edges on average, we see that this number is larger for the other two methods.



(a) True graph **(b)** Greedy forward selection

(c) Extremal graphical lasso **(d)** Clustering algorithm

Figure 3: True graph and graphs found for each of the three methods for the tree. Black edges correspond to edges in the true graph. Red edges correspond to wrongly found edges. The thickness of the edges indicates how many times that edge is found. Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

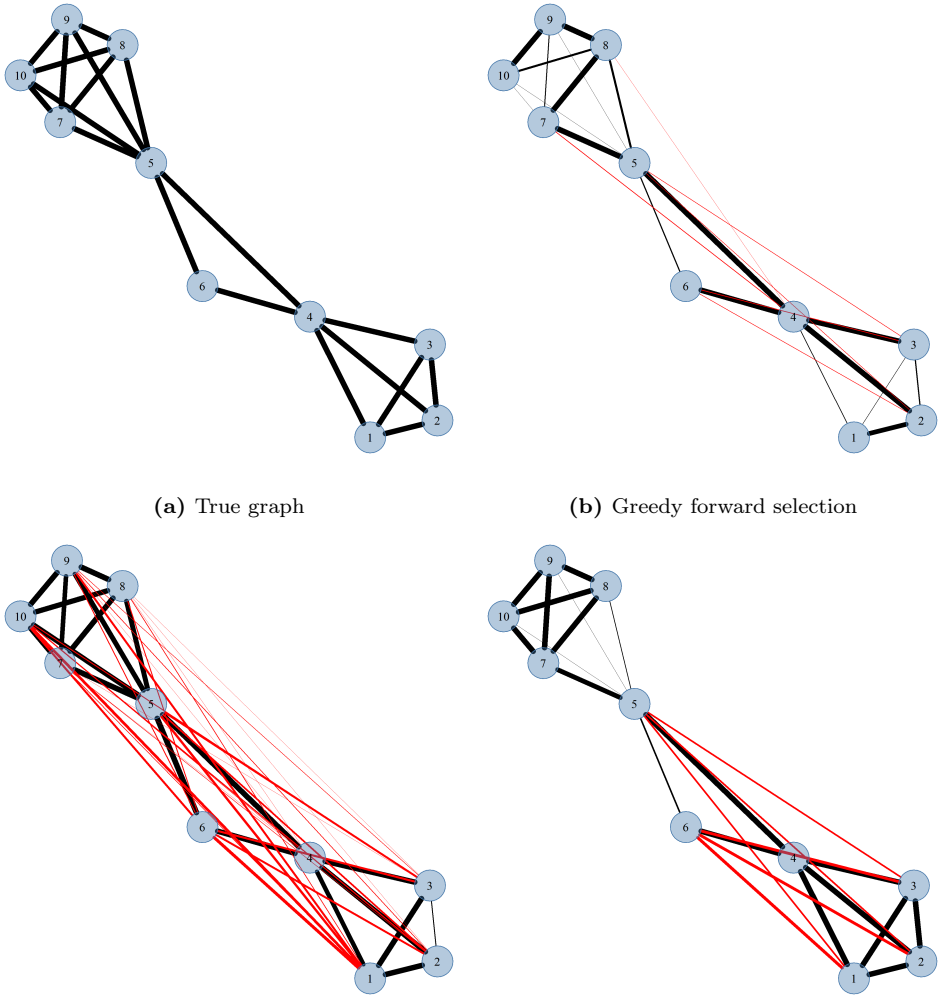
Block graph

For the block graph in Figure 4, we see that reglasso is best able to recover the true graph. This result can also be seen in Table 2, where it outperforms the other two methods and the empirical estimator in terms of accuracy of the tail correlations. It does, however, overestimate the number of edges in the graph significantly. This reduces interpretability of the graph, but has less of a considerable effect on the tail correlation accuracy compared to missing an edge. Although the greedy and clustering procedure are still able to improve upon the empirical estimator for the tail correlations, they do not find certain edges for the following reasons.

For the greedy forward selection, the problem arises from the maximum clique size of three. In the true graph we see three cliques of which two have a size larger than three. This type of graph is not

considered by the algorithm. Instead, it forces the graph is smaller cliques, thereby missing some edges. This is can be seen in Figure 4b.This reduces tail correlation prediction accuracy.

For the clustering algorithm, the problem also arises from a constraint on the graph type. While a block graph separates cliques by a single node, the clustering algorithm looks for connected cluster graphs, which separate cliques by a single edge. This leads to a bias that underestimates the inclusion of certain edges to force the allowed graph type. An example for this is edge $(5, 8)$ in Figure 5d.



(a) True graph **(b)** Greedy forward selection

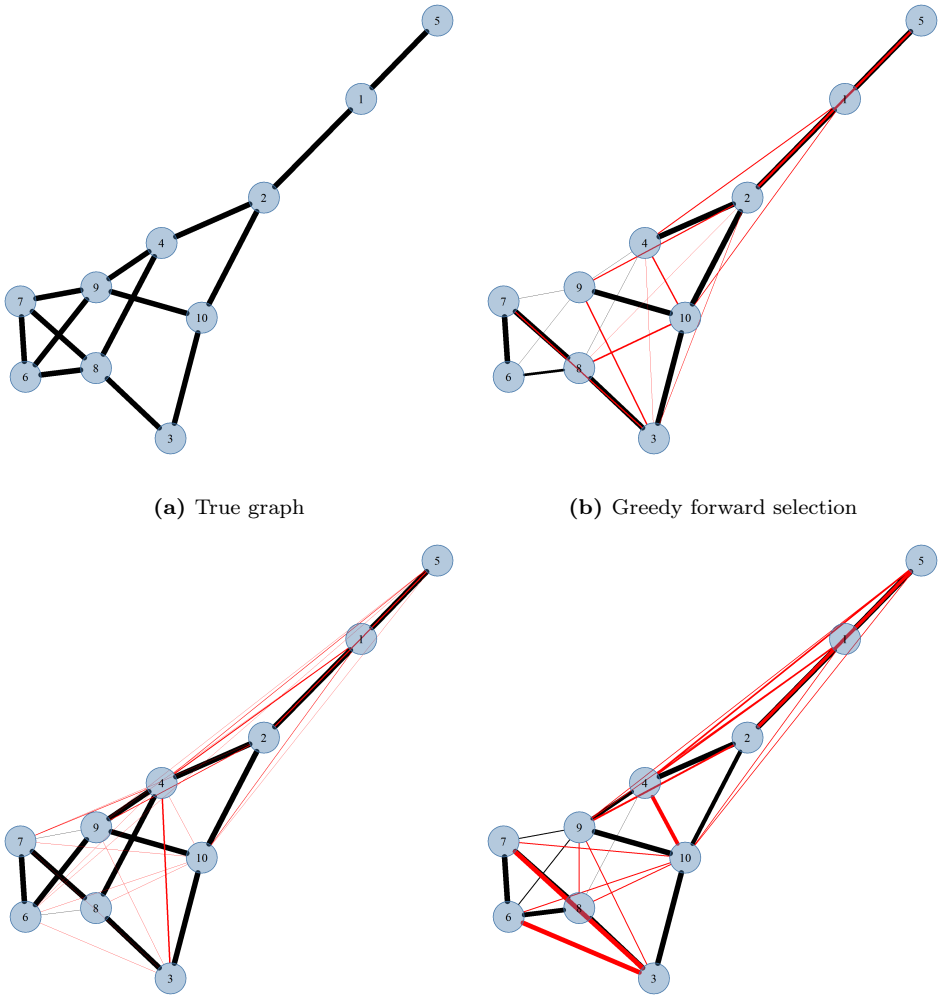
(c) Extremal graphical lasso **(d)** Clustering algorithm

Figure 4: True graph and graphs found for each of the three methods for the block graph. Black edges correspond to edges in the true graph. Red edges correspond to wrongly found edges. The thickness of the edges indicates how many times that edge is found. Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

Undecomposable graph

Reglasso performs favourable for the undecomposable graph in Figure 5, as it is able to find 12 of the 14 edges in almost all simulations. This translates in accurate results for the tail correlations in Table 2, especially when compared to the other two methods and the empirical estimator. Greedy forward selection

and the clustering algorithm underperform in terms of finding the correct graph and estimating the tail correlation due to the fact that they are looking for a decomposable graph.



(a) True graph

(b) Greedy forward selection

(c) Extremal graphical lasso

(d) Clustering algorithm

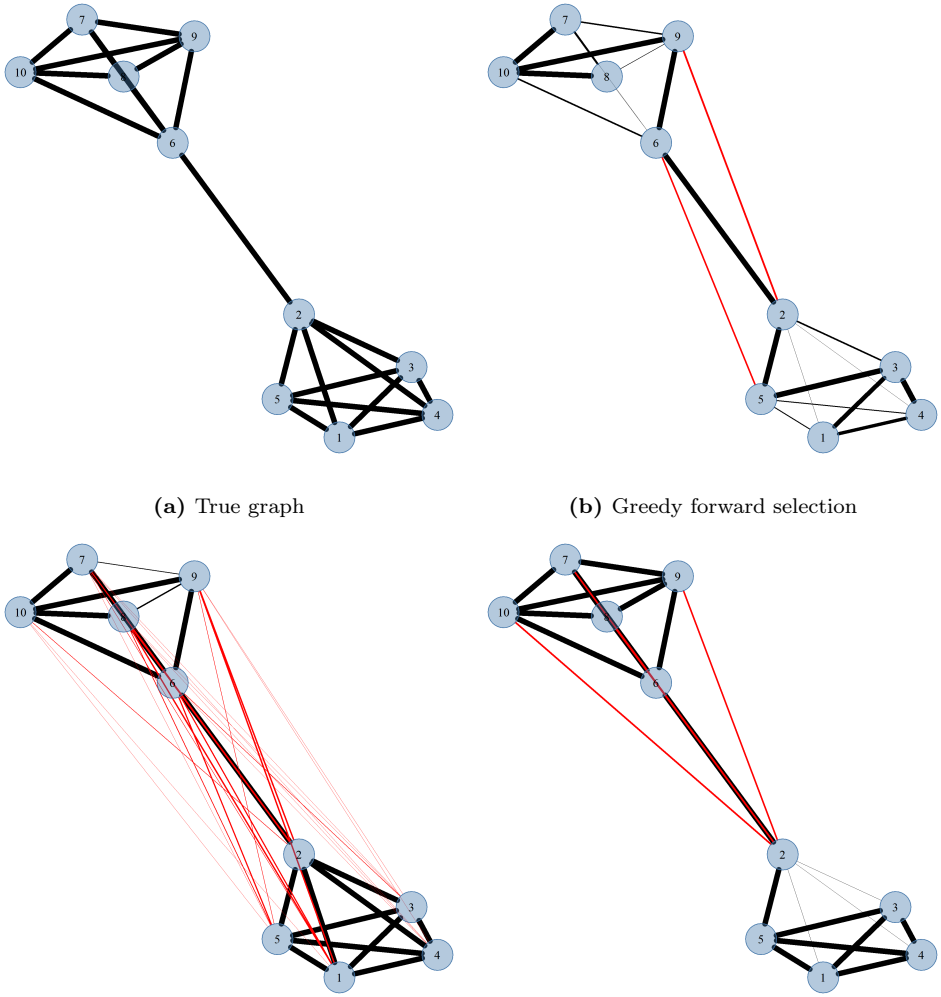
Figure 5: True graph and graphs found for each of the three methods for the undecomposable graph. Black edges correspond to edges in the true graph. Red edges correspond to wrongly found edges. The thickness of the edges indicates how many times that edge is found. Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

Connected cluster graph

For the connected cluster graph in Figure 6, we come to the following results.The greedy forward selection is not able to recover the full graph in most of the simulation. The reason is the same as for the previous graphs. The graph identification accuracy is limited because of the limitation of maximum clique size of three. This yields a poor estimation of the tail correlation that under performs the empirical estimator in terms of accuracy.

Furthermore, we observe that the tail correlation error of the clustering is also high. This contradicts the initial hypothesis that the clustering algorithm should perform well on these types of graphs, as it

especially designed for them. The explanation for this can be deducted from Figure 6d. Vertex 2 is assigned to the wrong cluster. While it is part of the bottom cluster in the true graph, we see that it is added to the top cluster or as a separate cluster in partition of the Leiden algorithm. Although this conceptually seems as a small mistake, it has large implications for the graph. Mislabelling this one node yields three missing edges and up to three wrong edges. This reduces tail correlation estimation accuracy to a point that it is outperformed by the empirical estimator. Although the sensitivity to mislabelling is detrimental for estimating the tail correlations, it may not be that much of a problem when the goal is only to interpret the graphical structure visually.



(a) True graph

(b) Greedy forward selection

(c) Extremal graphical lasso

(d) Clustering algorithm

Figure 6: True graph and graphs found for each of the three methods for the connected cluster graph. Black edges correspond to edges in the true graph. Red edges correspond to wrongly found edges. The thickness of the edges indicates how many times that edge is found. Results for a graph with 10 vertices, a generated sample size of 800 observations and 100 repetitions of the simulation.

In conclusion, it is observed that the (reconnecting) extremal graphical lasso has the highest tail correlation estimation accuracy of the three methods and outperforms the empirical estimator. The reason for this is that is able to recover most of the true edges of all methods. It does, however, tend to overestimate the number of edges in the graph. Despite that not having a big impact on the tail correlation accuracy, it has a negative effect on the visual interpretation of the graph. The other two methods can, therefore, complement this method in this aspect. Especially the clustering algorithm, which labels the nodes into groups, can help to understand the extremal dependence structure in the data. For that reason, it is useful to apply all of them to the same set of data for the empirical application.

5 Data

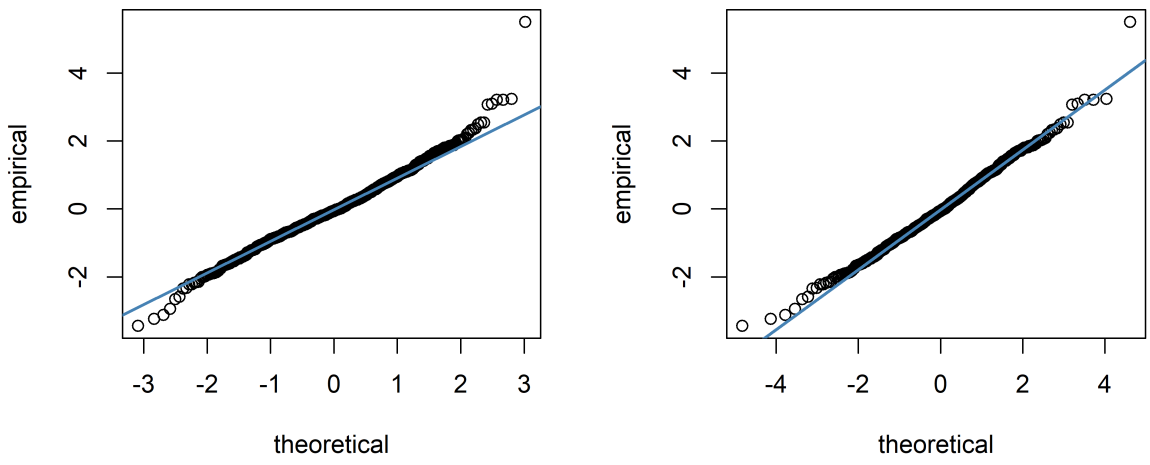
The data that are used for the empirical application consists of weekly future prices of 21 commodities from the energy, agriculture and metal industry. The time period that is considered ranges from January 6, 2006 through January 1, 2021, corresponding to 783 time points. We have to restrict ourselves to the most liquid exchange-traded futures. This is measured in open interest, which is the total number of futures contracts held by market participants at the end of the trading day. This is multiplied with the closing price and contract size to get the dollar value of the open contracts at a particular date. The minimum of the dollar value of open contracts is evaluated to be sufficiently high for all dates and across all commodities. The data is gathered from the Bloomberg terminal. The list of considered commodities is as follows

- Brent crude
- Natural gas
- Heating oil
- RBOB gasoline
- Corn
- Soybeans
- Sugar
- Wheat
- Rice
- Coffee
- Lumber
- Copper
- Aluminium
- Tin
- Zinc
- Gold
- Silver
- Platinum

Financial time series, such as these, adhere to a set of stylised facts, which include profound serial correlation of absolute returns, heteroscedasticity and clustering of extreme returns. In order to use extreme value theory, we have to perform a transformation to make the series i.i.d. ARMA-GARCH filtering is used for this purpose. The exact procedure is described in subsection 3.1 and only the results are shown here. The reader is referred to Appendix A for the tables with results for the AIC lag selection and to Appendix B for the Jarque-Bera and Ljung-Box tests.

The AIC is used to determine the number of lags that have to be included in the ARMA(p,q)-GARCH(r,s) model. The ARMA(1,0)-GARCH(1,1) has most often the lowest AIC for both the normal and Student-t distribution assumptions. This configuration is, therefore, used from here on.

Regarding the distribution assumption, we consider Figure 7, which shows the Q-Q plots of the filtered data under the normal distribution and t-distribution for sugar future returns series, as an example. It shows that the t-distribution is better able to capture the heavy tails in the data. This is further confirmed by the Jarque-Bera test, which rejects the null hypothesis of normality for 19 of the 21 series for a significance level of 1%.

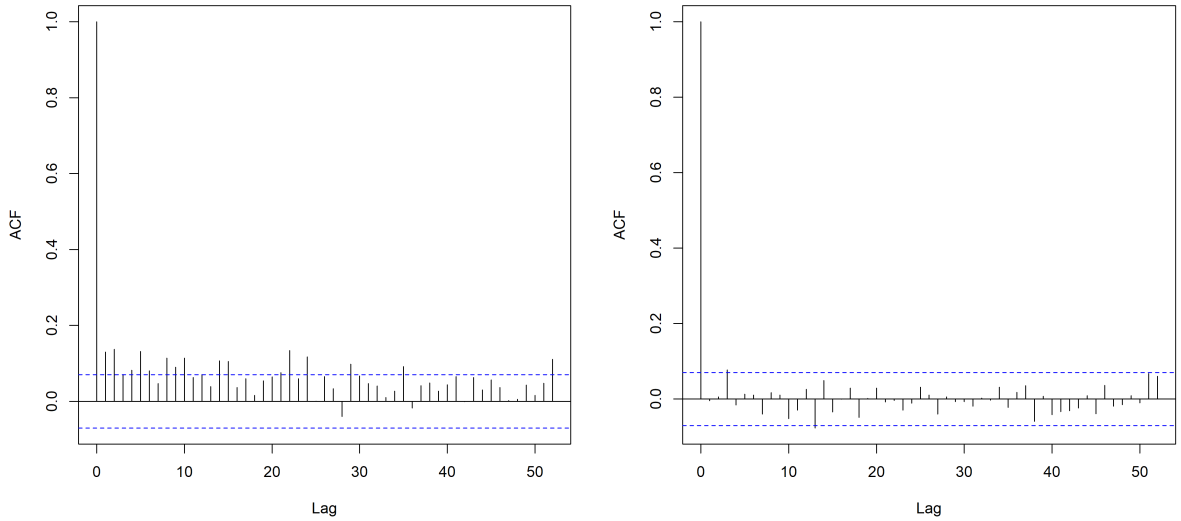


(a) Q-Q plot for normal distribution

(b) Q-Q plot for t-distribution with 9.1 d.o.f.

Figure 7: Q-Q plot of the filtered returns versus the theoretical distribution: normal (left) and t-distribution (right)

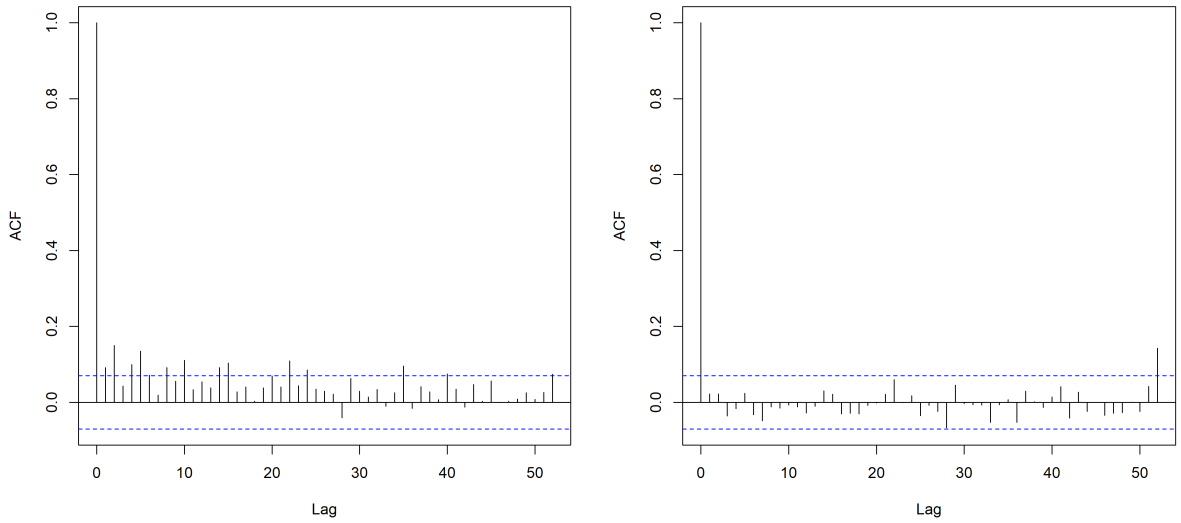
Finally, we examine whether the filtering has removed the serial correlation of the returns and squared returns. For illustrative purposes, Figure 8 and Figure 9 show the correlogram of the lags of the returns and squared returns of the sugar future respectively. Most of the serial correlation is removed by the filtering operations as indicated by the lower number of lags that exceeds the 95% confidence band. This is further confirmed by the Ljung-Box test on returns and squared returns which shows that the null hypothesis of no serial correlation is not rejected after filtering. This is in contrast to the unfiltered series where this null hypothesis is rejected most of the time, indicating serial correlation. The results of the visual inspection and statistical tests on the filtered data give enough confidence to satisfy the i.i.d. assumption required for extreme value theory.



(a) Correlogram for the unfiltered returns

(b) Correlogram for the filtered returns

Figure 8: Correlogram of the autocorrelation function and the lags for the unfiltered (left) and filtered (right) returns. Blue lines represent 95 % confidence band.



(a) Correlogram for the unfiltered squared returns

(b) Correlogram for the filtered squared returns

Figure 9: Correlogram of the autocorrelation function and the lags for the unfiltered (left) and filtered (right) squared returns. Blue lines represent 95 % confidence band.

To have an idea of the bilateral dependence of the data in a non-extreme setting we consider Table 3 which shows the ranking of the 20 strongest non-extreme correlations. It is useful to compare the relative strength of bilateral dependencies in the extreme and non-extreme setting to see if relations change in the tail. This is done in subsection 6.1. For the non-extreme setting, we observe that especially oil-related commodities are strongly correlated. The reason for this has to do with crude oil being an input for gasoline and heating oil. Other commodities within the same group, such as food, and precious or industrial metals, are also strongly connected.

Table 3: Ranking of the 20 largest non-extreme correlations for 21 series of commodity futures returns

Connection	Correlation	Rank	Connection	Correlation	Rank
Heating oil–Crude oil	0.883	1	Aluminium–Copper	0.579	11
RBOB gasoline–Crude oil	0.791	2	Kerosene–RBOB gasoline	0.524	12
RBOB gasoline–Heating oil	0.728	3	Tin–Copper	0.498	13
Zinc–Copper	0.657	4	Corn–Ethanol	0.480	14
Kerosene–Crude oil	0.656	5	Zinc–Tin	0.456	15
Kerosene–Heating oil	0.641	6	Platinum–Copper	0.434	16
Platinum–Gold	0.636	7	Platinum–Aluminium	0.410	17
Soybeans–Corn	0.623	8	Tin–Aluminium	0.406	18
Wheat–Corn	0.593	9	Oats–Corn	0.404	19
Zinc–Aluminium	0.588	10	Wheat–Soybeans	0.397	20

6 Empirical application

The methods to find the extremal graphical structure and tail correlations, which are outlined in section 3 and analysed with a simulation study in section 4, are applied to the data set of filtered commodity futures returns described in section 5. Firstly, the economical findings based on the chosen threshold of $p = 0.85$ results are shown. After that, a sensitivity analysis on the threshold selection is performed, in order to verify the robustness of these results.

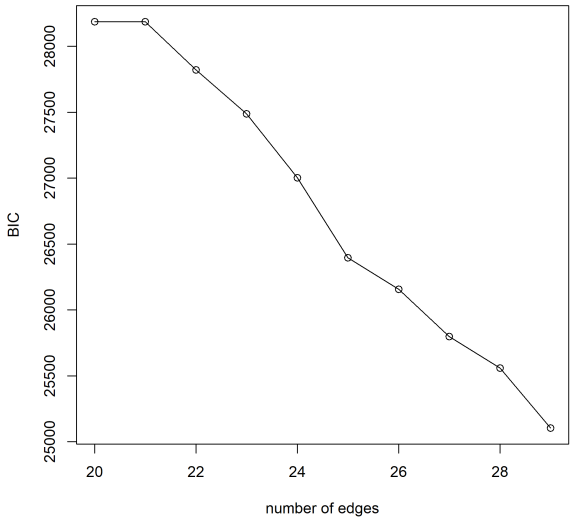
6.1 Empirical results

Table 4 shows the BIC and the number of edges in the graph for the model obtained by each of the three approaches. It shows that the greedy forward and clustering algorithm are further away from the true graph than the extremal graphical lasso, as a higher BIC is obtained with a larger number of edges. Figure 10 shows the BIC for different numbers of edges. The number of edges comes from the different steps in the greedy forward algorithm, from the ρ parameter in reglasso and from the resolution parameter γ in the clustering algorithm. For the greedy forward selection, we can conclude from its BIC plot in Figure 10a that the algorithm stops too early at 28 edges. The explanation for that is that the clique size is limited to three for simplicity reasons. Increasing this allows for more potential edges to be evaluated, but has the downside of increasing computational effort greatly.

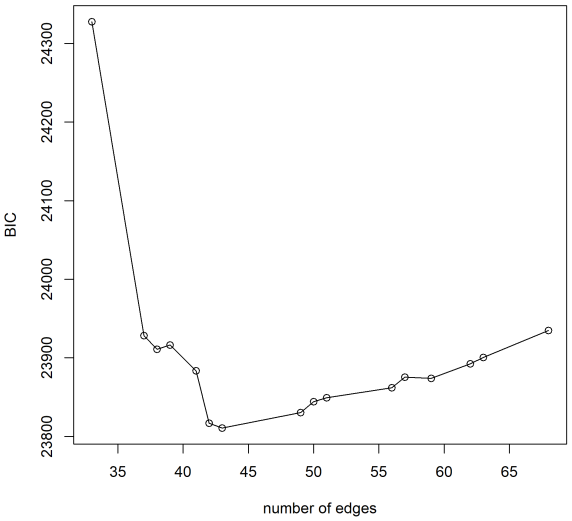
The clustering algorithm is able to improve slightly on the forward greedy but still comes short compared to the extremal graphical lasso. Lower resolution parameter settings have a larger number of edges in the graph. However, for these settings, the nodes tend to cluster together in a few large blobs and the rest is left as separate nodes. This creates large cliques that are difficult to estimate due to their high dimensionality.

Table 4: The number of edges and BIC for the models obtained for the three different structural identification methods. The full sample and threshold $p = 0.85$ are used.

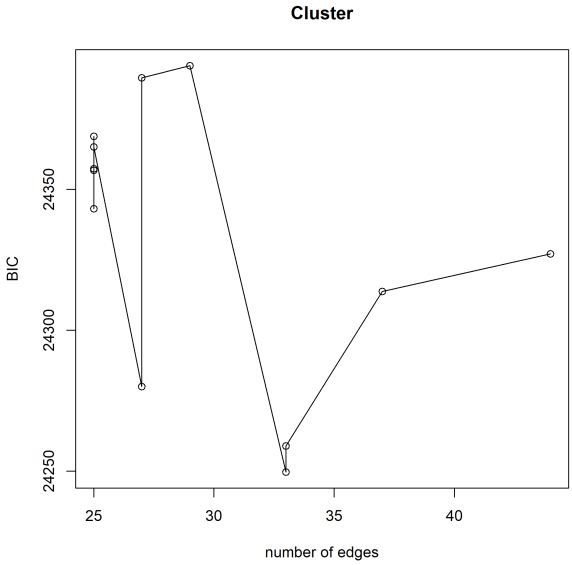
	$p = 0.85$		
	Greedy forward selection	Extremal graphical lasso	Clustering algorithm
Number of edges	28	43	33
BIC	25,494	23,811	24,238



(a) BIC versus edges for greedy forward algorithm.



(b) BIC versus edges for extremal graphical lasso.

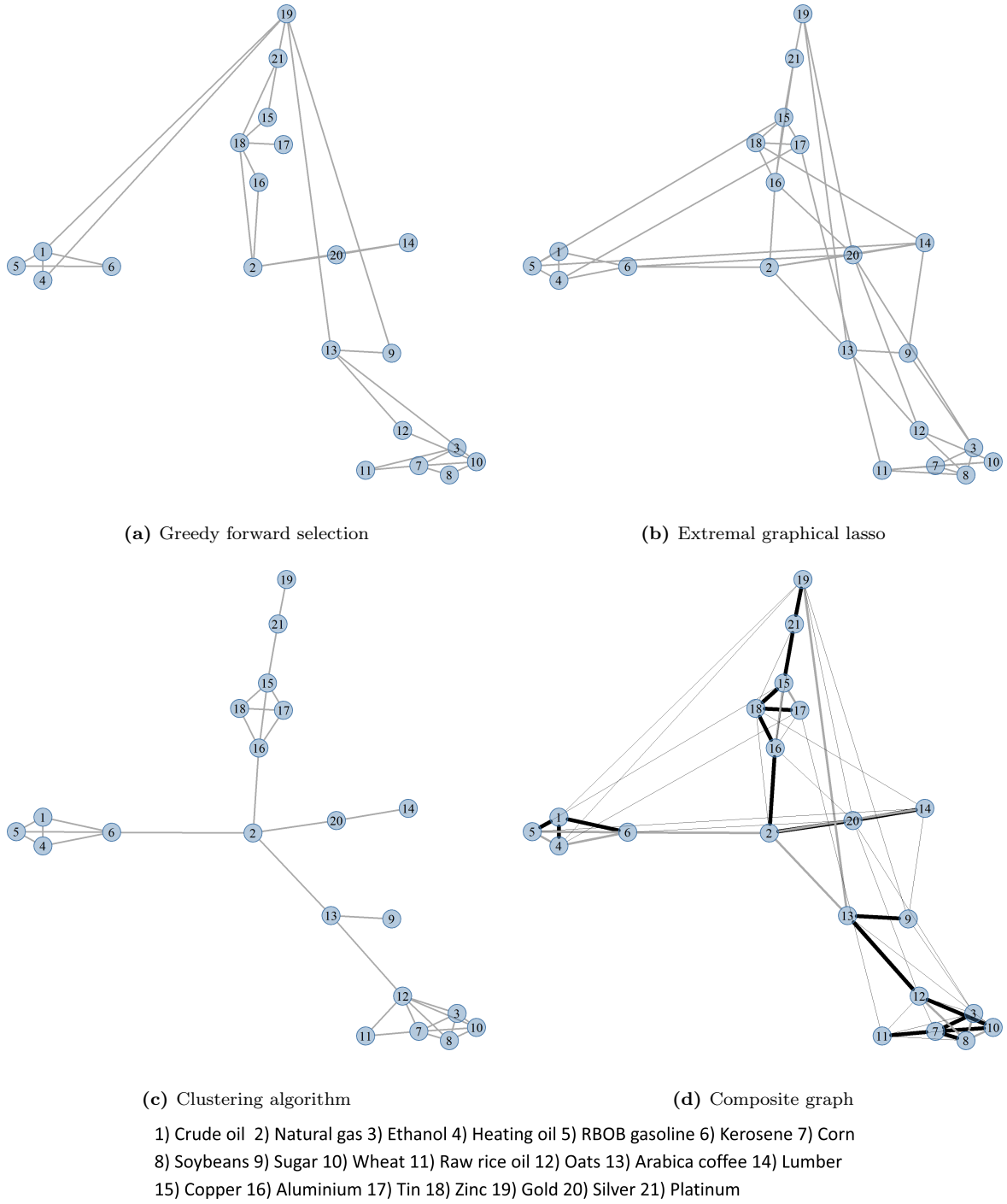


(c) BIC versus edges for clustering algorithm.

Figure 10: BIC versus edges for all three algorithms. Threshold $p = 0.85$ is used.

Figure 11 shows the graphs found by each of the three methods and a composite graph that shows how the models agree. For the composite graph, thick and black lines correspond to edges that are included in all approaches, grey and normal thickness in two and light grey and thin in one. The main advantage of the clustering algorithm is observed in this figure, namely its interpretability. We see three main clusters being formed. The first one consists of the oil commodities: crude oil, heating oil, gasoline and kerosene. This is a strong clique that is found by each of the three models, except for the greedy forward due to its limit of clique size 3. Another strongly connected group of commodities are essential food commodities, i.e. corn, soybeans, wheat rice, and oats, and ethanol. These crops serve as a crucial input for the world's food supply and can serve as substitutes for each other. Ethanol is part of this group due to its input

being corn. Sugar and coffee are not part of this cluster. Instead, they are connected together as a more discretionary food product. The third large cluster consists of industrial metals: copper, aluminium, tin and zinc. That group is separate from the precious metals gold and platinum. Silver is not in this group. Instead, it is weakly connected with lumber.



(a) Greedy forward selection

(b) Extremal graphical lasso

(c) Clustering algorithm

(d) Composite graph

1) Crude oil 2) Natural gas 3) Ethanol 4) Heating oil 5) RBOB gasoline 6) Kerosene 7) Corn 8) Soybeans 9) Sugar 10) Wheat 11) Raw rice oil 12) Oats 13) Arabica coffee 14) Lumber 15) Copper 16) Aluminium 17) Tin 18) Zinc 19) Gold 20) Silver 21) Platinum

Figure 11: Graphical structure for three different approaches and composite graph. Thick and black lines are included in all approaches, grey and normal in two and light grey and thin in one. The threshold $p = 0.85$ is used.

Table 5 shows the ten highest model-implied tail correlations, i.e. χ_{ij}, of the three approaches and

their rank for the corresponding approach. We observe that the models agree to a large extent and have a similar ranking as that of the non-extremal correlations given in Table 3. There is, however, one notable change in the tail correlations ranking compared to that of their non-extreme cousins. The link between wheat and corn is higher in the ranking for the high price increases. This is most likely caused by the fact that price jumps caused by disasters, such as draughts or hurricanes, affect both crops equally as they are often grown in the same regions, combined with the fact that they are substitutes.

The strongest links between commodity types are those with natural gas. This is seen most clearly in the graph for the clustering algorithm, but is also visible in the other plots. The link between industrial metals and natural gas is the strongest with tail correlations ranging around 0.4. The reason for this is that the metal industry is an energy-intensive industry that relies on natural gas as an energy input. The link between other commodity types such as liquid energy or food commodities is much weaker with tail correlations around 0.10 and 0.05 respectively. This further supports the neutrality hypothesis in the energy-food nexus. We also see weak tail correlations between ethanol and energy commodities. The non-extreme correlation between gasoline and ethanol has rank 40 of the 210 correlations. For the tail correlation, the rank has increased to around 180 for each of the three models. An explanation for this decrease in dependency is that this dependency is demand-pull driven. More demand for gasoline drives up its price. As gasoline is required to be mixed with ethanol in numerous countries such as the USA and Brazil, this also drives up the demand, and therefore the price, of ethanol. Gasoline price shocks, however, are often driven by supply-side issues for crude oil that does not have an effect on ethanol production.

Table 5: Ranking of the 20 largest tail correlations for 21 series of commodity futures returns for the three approaches. Threshold $p = 0.85$ is used.

	Greedy forward selection		Eglasso		Clustering	
Connection	Tail correlation	Rank	Tail correlation	Rank	Tail correlation	Rank
Heating oil–Crude oil	0.697	1	0.747	1	0.705	1
RBOB gasoline–Crude oil	0.597	2	0.631	2	0.589	2
RBOB gasoline–Heating oil	0.511	10	0.629	3	0.584	3
Zinc–Copper	0.541	4	0.621	4	0.558	5
Wheat–Corn	0.569	3	0.613	5	0.570	4
Kerosene–Crude oil	0.525	5	0.587	6	0.528	8
Platinum–Gold	0.512	8	0.580	7	0.490	12
Zinc–Aluminium	0.512	9	0.578	8	0.542	6
Kerosene–Heating oil	0.521	6	0.565	9	0.526	9
Aluminium–Copper	0.503	11	0.564	10	0.523	10

6.2 Threshold sensitivity

The threshold p is used to eliminate the non-extreme observations, i.e. the observations for which none of the elements exceeds the p^{th} quantile. This is an important parameter to set because a threshold that is set too low introduces bias, as we include observation that are not far enough in the tail, while a threshold that is set too high introduces parameter uncertainty due to the smaller number of observations.

The threshold is set to 0.85 in the main analysis using a visual inspection of several tail correlation plots. It is, however, important to consider the sensitivity to this parameter. For this reason, we consider a lower threshold of $p = 0.75$ and a higher threshold of $p = 0.95$. After standardisation, the original threshold has 672 observations. For the lower threshold, this increases to 744. The higher threshold has 415 observations.

Comparing Table 6 and Table 7 with Table 4, it is observed that the greedy algorithm remains stable in the number of edges. Yet again, this is caused by the limit on clique size. For the extremal graphical lasso and clustering algorithm, however, the story is different. Both methods directly rely on the empirical estimate of the variogram. For the lower threshold, we see the number of observations increase and therefore expect a better estimate of the parameter matrix. This translates in a better BIC in relation to that of the greedy forward method compared to the base case of $p = 0.85$. Per contra, if the number of observations is decreased, as is the case for the higher threshold, we see a sharp decline in model fit.

The ranking of the tail correlations does not differ to a large degree for the different thresholds. We do see higher correlations for lower thresholds. The correlations rankings for all three thresholds are found in Appendix C.

Table 6: The number of edges and BIC for the models obtained for the three different structural identification methods. The full sample and threshold $p = 0.75$ are used.

	$p = 0.75$		
	Greedy forward selection	Extremal graphical lasso	Clustering algorithm
Number of edges	29	57	29
BIC	39,609	35,243	36,956

Table 7: The number of edges and BIC for the models obtained for the three different structural identification methods. The full sample and threshold $p = 0.95$ are used.

	$p = 0.95$		
	Greedy forward selection	Extremal graphical lasso	Clustering algorithm
Number of edges	28	35	22
BIC	3,566	9,237	9,163

7 Conclusion

This research investigates the upper tail dependence structure of 21 commodity futures returns, spanning energy, agriculture and metal commodities. It does this by using the theory on extremal graphical models developed by Engelke and Hitz (2020). For identifying the graphical structure, three different approaches are taken. The first method uses the minimum spanning tree as a starting point and adds edges using a greedy forward search based on the BIC. The second method, called extremal graphical lasso, forces sparsity on the off-diagonal elements precision matrix $\Sigma^{(k)}$ for all k and uses it to construct the variogram. The third method, named the clustering method, finds connected nodes and connects them as cliques. The disconnected cliques are sewn together using the strongest single connection between cliques based on the empirical variogram.

A simulation study on these three methods shows that an empirical estimator does not suffice for the tail correlations and that model-implied tail correlations come closer to the true tail correlation matrix. The extremal graphical lasso is best able to recover the true edges. This translates in the most accurate tail correlations. However, it overestimates the number of edges in the graph, making it difficult to draw visual conclusions from the graphs. In a final conclusion, it is found that the clustering algorithm has a main weakness in that the accuracy of the estimated tail correlation is very sensitive to a mislabelling of the nodes. It is, however, still useful to understand the network structure of the data.

The theory on extremal graphical models and the three graph identification methods are applied to the filtered data set consisting of 21 futures series returns ranging from 2006 to 2021. It is found that the results of the three approaches are close to each other and that the strongest connections are found between nodes that are in the same group. In the connections between groups, we see that natural gas serves as the pivot. The strength of these connections is lower with respect to the non-extreme correlations. This supports the neutrality hypothesis.

There are multiple ways that this research can be extended. The clustering algorithm has room for improvement as its performance is quite sensitive to mislabelling nodes. Furthermore, connecting cliques happen with a single edge, instead of a single node, which is the case in block graphs. Grouping the nodes in different ways and overlapping these results to find the separator nodes can be a way to tackle both problems. Other extremal graph identification strategies can be tested as well. Engelke, Lalancette and Volgushev are also working on a neighbourhood selection method for extremes in parallel to the extremal graphical lasso. This method has shown some promising results, but no specific details on the implementation of this approach are released at the time of writing.

A different approach to studying the dependence structure of commodities is to study its causal structure. Bayesian networks are graphical models that are more concerned with cause and effect. Forming this kind of model would allow for having a more detailed look into which commodity price spike causes a spike in another commodity. To the best of the author's knowledge, there is no extremal version of this theory at this point in time.

An extension of the current research can be to translate the found results into a solution for this tail dependence by forming a hedging portfolio. In this portfolio, a balance has to be struck between non-extreme and extreme dependence. For example, the connection between corn and wheat is weaker in normal times than during large price hikes. This dependence is, therefore, less of concern most of the time. However, during those few extreme occasions, we may find that the hedge proves the most useful. A balance has therefore to be struck. Combining the extreme and non-extreme results in one portfolio should allow investors and producers to be more protected against large shocks in prices.

References

Albulescu, C. T., Tiwari, A. K., and Ji, Q. (2020). Copula-based local dependence among energy, agriculture and metal commodities markets. *Energy*, 202:117762.

Alem, Y. and Söderbom, M. (2012). Household-level consumption in urban ethiopia: the effects of a large food price shock. *World development*, 40(1):146–162.

Baumeister, C. and Kilian, L. (2014). Do oil price increases cause higher food prices? *Economic Policy*, 29(80):691–747.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1-2):5–59.

BP (2021). Statistical review of world energy 2021. Technical report, British Petroleum Company.

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

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

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

Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. *arXiv preprint arXiv:2012.06179*.

Fellmann, T., Hélaine, S., and Nekhay, O. (2014). Harvest failures, temporary export restrictions and global food security: the example of limited grain exports from russia, ukraine and kazakhstan. *Food Security*, 6(5):727–742.

Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Headey, D. (2011). Rethinking the global food crisis: The role of trade shocks. *Food Policy*, 36(2):136–146.

Headey, D. and Fan, S. (2008). Anatomy of a crisis: the causes and consequences of surging food prices. *Agricultural economics*, 39:375–391.

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

IEA (2021). Renewables 2021. Technical report, International Energy Agency.

Ji, Q., Bouri, E., Roubaud, D., and Shahzad, S. J. H. (2018). Risk spillover between energy and agricultural commodity markets: A dependence-switching covar-copula model. *Energy Economics*, 75:14–27.

Kristoufek, L., Janda, K., and Zilberman, D. (2012). Correlations between biofuels and related commodities before and during the food crisis: A taxonomy perspective. *Energy Economics*, 34(5):1380–1391.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

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

Marktanner, M. and Noiset, L. P. (2013). Food price crisis, poverty, and inequality. *The Developing Economies*, 51(3):303–320.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

Nazlioglu, S. (2011). World oil and agricultural commodity prices: Evidence from nonlinear causality. *Energy policy*, 39(5):2935–2943.

Nazlioglu, S. and Soytas, U. (2012). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics*, 34(4):1098–1104.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 435(7043):814–818.

Papastathopoulos, I. and Strokorb, K. (2016). Conditional independence among max-stable laws. *Statistics & Probability Letters*, 108:9–15.

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

Reboredo, J. C. (2012). Do food and oil prices co-move? *Energy policy*, 49:456–467.

Reboredo, J. C. and Ugolini, A. (2016). The impact of downward/upward oil price movements on metal prices. *Resources Policy*, 49:129–141.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

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

Traag, V. A., Van Dooren, P., and Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, 84(1):016114.

Traag, V. A., Waltman, L., and Van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12.

Wang, Y., Wu, C., and Yang, L. (2014). Oil price shocks and agricultural commodity prices. *Energy Economics*, 44:22–35.

Yahya, M., Oglend, A., and Dahl, R. E. (2019). Temporal and spectral dependence between crude oil and agricultural commodities: A wavelet-based copula approach. *Energy Economics*, 80:277–296.

Zhang, Z., Lohr, L., Escalante, C., and Wetzstein, M. (2010). Food versus fuel: What do prices tell us? *Energy policy*, 38(1):445–451.

A AIC for different GARCH tests

Table 8: Akaike Information Criterion for different lags of the ARMA(p,q)-GARCH(r,s) model. Each row corresponds to the univariate future return series of the 21 different commodities. Bold-faced numbers indicate the lowest values of the row.

AIC	ARMA(1,0)-GARCH(1,1)	ARMA(1,1)-GARCH(1,1)	ARMA(1,1)-GARCH(2,1)
Crude oil	**-3.494**	-3.491	-3.491
Natural gas	-2.635	**-2.638**	-2.637
Ethanol	**-3.472**	-3.470	-3.467
Heating oil	**-3.561**	-3.559	-3.557
RBOB gasoline	**-3.190**	-3.189	-3.187
Kerosene	-3.831	**-3.841**	-3.839
Corn	**-3.673**	-3.671	-3.669
Soybeans	**-4.089**	-4.087	-4.085
Sugar	-3.401	**-3.403**	-3.400
Wheat	**-3.499**	-3.497	-3.495
Raw rice	**-3.989**	-3.988	-3.986
Oats	**-3.325**	-3.324	-3.321
Arabica coffee	-3.578	**-3.588**	-3.574
Lumber	**-3.360**	-3.358	-3.357
Copper	-3.916	**-3.917**	-3.914
Aluminium	**-4.219**	-4.217	-4.215
Tin	**-3.900**	-3.897	-3.895
Zinc	-3.625	**-3.627**	-3.625
Gold	**-4.650**	-4.648	-4.646
Silver	**-3.877**	-3.874	-3.872
Platinum	**-4.037**	-4.032	-4.030

B Jarque-Bera and Ljung-Box test results

Table 9: Jarque-Bera test for filtered returns with normal distribution assumption

	Statistic	p-value		Statistic	p-value		Statistic	p-value
Crude oil	76.9	0.000	Soybeans	20.6	0.000	Copper	175.5	0.000
Natural gas	5.5	0.065	Sugar	77.8	0.000	Aluminium	22.3	0.000
Ethanol	160.1	0.000	Wheat	24.2	0.000	Tin	311.6	0.000
Heating oil	172.4	0.000	Raw rice	29.0	0.000	Zinc	1.1	0.567
RBOB gasoline	548.8	0.000	Oats	48.3	0.000	Gold	27.5	0.000
Kerosene	106.7	0.000	Arabica coffee	21.6	0.000	Silver	491.2	0.000
Corn	1702.2	0.000	Lumber	25.3	0.000	Platinum	88.8	0.000

Table 10: Ljung-Box tests for the unfiltered and filtered returns for 21 commodity futures.

	Unfiltered returns		Filtered returns		Unfiltered squared returns		Filtered squared returns	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Crude oil	35.2	0.019	27.5	0.093	594.1	0.000	21.8	0.295
Natural gas	29.9	0.071	22.0	0.283	76.1	0.000	13.5	0.811
Ethanol	29.5	0.078	21.2	0.325	81.6	0.000	9.0	0.973
Heating oil	21.0	0.398	20.6	0.359	172.4	0.000	33.3	0.022
RBOB gasoline	43.1	0.002	32.1	0.030	220.6	0.000	26.6	0.114
Kerosene	58.4	0.000	23.4	0.219	134.6	0.000	21.4	0.315
Corn	31.8	0.045	27.8	0.086	75.9	0.000	10.0	0.953
Soybeans	30.2	0.067	18.8	0.468	213.7	0.000	36.6	0.009
Sugar	26.8	0.142	21.0	0.339	98.0	0.000	9.7	0.960
Wheat	36.3	0.014	30.5	0.046	90.7	0.000	16.5	0.627
Raw rice	44.1	0.001	37.7	0.006	258.3	0.000	26.1	0.128
Oats	22.1	0.337	20.3	0.378	19.3	0.505	17.8	0.536
Arabica coffee	25.4	0.188	27.1	0.101	43.4	0.002	24.8	0.168
Lumber	31.4	0.050	29.8	0.055	120.8	0.000	9.3	0.969
Copper	53.5	0.000	31.2	0.039	276.9	0.000	25.6	0.141
Aluminium	22.0	0.340	14.5	0.754	231.2	0.000	24.6	0.175
Tin	69.7	0.000	23.6	0.214	499.5	0.000	11.3	0.913
Zinc	14.9	0.780	8.2	0.985	398.6	0.000	17.1	0.581
Gold	23.2	0.280	8.9	0.976	300.7	0.000	27.1	0.103
Silver	7.4	0.995	8.1	0.985	92.7	0.000	33.3	0.022
Platinum	16.9	0.658	11.0	0.925	266.5	0.000	18.6	0.482

C Tail correlation tables for different thresholds

Table 11: Ranking of the 20 largest tail correlations for 21 series of commodity futures returns for the three approaches. Threshold $p = 0.75$ is used.

	Greedy forward selection		Eglasso		Clustering	
Connection	Tail correlation	Rank	Tail correlation	Rank	Tail correlation	Rank
Heating oil–Crude oil	0.725	1	0.764	1	0.740	1
RBOB gasoline–Crude oil	0.615	2	0.648	2	0.633	2
RBOB gasoline–Heating oil	0.608	3	0.644	3	0.628	3
Zinc–Copper	0.597	5	0.630	4	0.611	5
Wheat–Corn	0.600	4	0.624	5	0.616	4
Kerosene–Crude oil	0.595	6	0.602	6	0.573	9
Platinum–Gold	0.565	9	0.592	7	0.580	7
Zinc–Aluminium	0.580	7	0.588	8	0.581	6
Kerosene–Heating oil	0.524	13	0.586	9	0.570	10
Corn–Ethanol	0.564	10	0.582	10	0.562	12

Table 12: Ranking of the 20 largest tail correlations for 21 series of commodity futures returns for the three approaches. Threshold $p = 0.85$ is used.

	Greedy forward selection		Eglasso		Clustering	
Connection	Tail correlation	Rank	Tail correlation	Rank	Tail correlation	Rank
Heating oil–Crude oil	0.697	1	0.747	1	0.705	1
RBOB gasoline–Crude oil	0.597	2	0.631	2	0.589	2
RBOB gasoline–Heating oil	0.511	10	0.629	3	0.584	3
Zinc–Copper	0.541	4	0.621	4	0.558	5
Wheat–Corn	0.569	3	0.613	5	0.570	4
Kerosene–Crude oil	0.525	5	0.587	6	0.528	8
Platinum–Gold	0.512	8	0.580	7	0.490	12
Zinc–Aluminium	0.512	9	0.578	8	0.542	6
Kerosene–Heating oil	0.521	6	0.565	9	0.526	9
Aluminium–Copper	0.503	11	0.564	10	0.523	10

Table 13: Ranking of the 20 largest tail correlations for 21 series of commodity futures returns for the three approaches. Threshold $p = 0.95$ is used.

	Greedy forward selection		Eglasso		Clustering	
Connection	Tail correlation	Rank	Tail correlation	Rank	Tail correlation	Rank
Heating oil–Crude oil	0.620	1	0.750	1	0.622	1
RBOB gasoline–Heating oil	0.499	2	0.624	2	0.471	2
RBOB gasoline–Crude oil	0.402	9	0.620	3	0.465	3
Zinc–Copper	0.463	4	0.600	4	0.460	4
Wheat–Corn	0.473	3	0.592	5	0.452	5
Platinum–Gold	0.382	12	0.575	6	0.382	11
Kerosene–Crude oil	0.409	7	0.563	7	0.407	8
Soybeans–Corn	0.424	5	0.561	8	0.406	9
Zinc–Aluminium	0.411	6	0.553	9	0.435	6
Zinc–Tin	0.383	11	0.535	10	0.362	12