Eötvös Loránd University
Corvinus University of Budapest

Szabolcs Majoros

Multivariate stable distributions and their application in modeling returns

Thesis

Actuarial and Financial Mathematics MSc
Quantitive Finance Specialization

Supervisor:

András Zempléni
Department of Probability Theory and Statistics

Budapest, 2018

Acknowledgment

I would like to thank András Zempléni not just for supervising the thesis, but for all the help and advice in the prior works of the subject through the past two years. Also, I would like to thank him the many opportunities he gave me to gain experience and to improve myself.

The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

Contents

1 Introduction

1.1 Motivation

For the sake of simplification, it was often assumed that the returns of financial assets are distributed normally, which seems to be a natural choice in practice, because of it's theoretical importance: the Central Limit Theorem. However, this supposition does not exactly hold in return modeling, due to their extremal behavior, which can been observed in the markets and for many different assets.

Some of the common behaviors of asset returns, known as stylized facts are:

- **Heavy tails**
  This is the most important property – at least in the case of the thesis. This means, that the tails of the characterizing distribution are decaying in power-law, which is a much slower order of decay than the exponentially decaying normal distribution's tails. The consequence of this is that much more probability will be concentrated on the tails, so the probability of extreme returns rises, compared to non-heavy tailed distributions.

- **Gain-Loss asymmetry**
  The second most important property. Mathematically, this asymmetry means that the characterizing distribution of the returns is skewed negatively, so the losses of the assets are usually bigger than the gains. Normal distribution is symmetrical so we are unable to model this phenomenon with it.

- **Difference in scale of time**
  Usually, if we observe returns from long frequency data, the returns will more likely be normally distributed, however when we are trying to model returns from short frequency data, heavy tails tend to appear.

To model returns more precisely we have some alternatives, such as Student t-distribution, power exponential distribution, Weibull distribution and other time independent and dependent models. More of these can be found e.g. in [12], but not all of them are theoretically established. Instead of these, I work with stable distributions in my thesis.

Stable distributions are a rich class of probability distributions with many practical properties, especially in finance or insurance. Paul Lévy was the one who first studied this distribution family and he proved the Generalized Central Limit Theorem, which is an important theorem in the case of stable distributions. The first attempt to use these distributions in finance was done by Benoit Mandelbrot, who was modeling cotton price

changes with stable distributions [3]. Since then many other studies were published in the subject. The most recent works were done by J. P. Nolan, who is also working on a comprehensive book [11] on stable distributions. Modeling stock returns with stable distributions was in attention for a while and most of the results were promising, however some studies stated that the distribution of stock returns must have finite second moments [14], which would exclude non-Gaussian stable distributions. The studies only observed stock returns and didn't take into consideration an interesting asset category for stable distributions: cryptocurrencies. Cryptocurrencies have much different characteristics than stocks, commodities or regular currencies, therefore stable distributions may be used to model their returns. Bitcoin, the most famous cryptocurrency had the one of the most interesting and volatile price movements in the last few years, which fluctuation can also be observed on other cryptocurrencies, presumably depending on each other. This is why stable distributions – univariate and multivariate – could be proper tools to model their returns.

At first I introduce univariate stable distributions and their parameter estimation methods in Section 1, since I build on it later on. After these in Section 2, I go trough multivariate stable distributions and their most important properties, then I introduce the bivariate parameter estimation method in Section 3 and the general procedure in Section 5, which I put in to practice on cryptocurrency data, first in two, and then in three dimensions too in Sections 4 and 6. For checking the fit of the distribution, I use two goodness-of-fit methods in univariate and multivariate cases, which are introduced in Section 3. I mostly follow Nolan's works, as he wrote many extensive articles about stable distributions. These articles are giving a good overview about the subject, building on mathematical results done in the past 40 years. Additionally, the base of the thesis is from [18] and [19], my prior works in the subject.

1.2 Univariate stable distributions

The following introduction to univariate stable distributions is based on [11] and [17]. By definition, a random variable X is stable, if to any positive $a, b \in \mathbb{R}$, there are positive $c \in \mathbb{R}$ and $d \in \mathbb{R}$, such that the distribution of $aX_1 + bX_2$ has the same distribution as $cX + d$, where the random variables X_1 and X_2 are independent, identically distributed to X. We can be familiar with this property from the normal distribution, as it is a member of the stable distribution family too.

In the univariate case, the distribution is described by four parameters: index $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma > 0$ and shift $\delta \in \mathbb{R}$ parameters. The usual notion for the distribution is $S(\alpha, \beta, \gamma, \delta)$. One serious drawback makes it harder to work with this

distribution family, that they have no closed form of density function, apart from a few special cases: the well known normal distribution, the Cauchy and Lévy distributions. Instead of density functions, they are described by their characteristic functions, which we can handle better now as the computers became more advanced and faster. By definition, the characteristic function of a X random variable is $\phi_X(t) = E[e^{itX}]$, which is computed for absolutely continuous distributions with $f(x)$ density function in the $\int_{-\infty}^{\infty} f(x)e^{itx}dx$ form. The characteristic function of stable distributions has the following form:

$$\varphi(t) = \begin{cases} \exp\left\{-\gamma^\alpha |t|^\alpha (1 + i\beta \tan\frac{\pi\alpha}{2} \cdot \operatorname{sign} t)(|\gamma t|^{1-\alpha} - 1) + i\delta t\right\} & \alpha \neq 1 \\ \exp\left\{ -\gamma|t| \left(1 + i\beta\frac{2}{\pi} \operatorname{sign} t \cdot \log\left(\gamma|t|\right)\right) + i\delta t\right\} & \alpha = 1. \end{cases}$$

Stable distribution are always absolutely continuous with any set of parameters and unimodal too.

There are a few different representations of the characteristic function, which differ in parametrization. The above form is called the S_0 representation, which is more useful in practical problems, because it is easier to compute.

From the characteristic function we can get back the special cases mentioned above: by taking $S(2, 0, \gamma, \delta)$, we get the normal characteristic function with $N(\delta, 2\gamma^2)$ parameters. The $S(1, 0, \gamma, \delta)$ parametrization is the Cauchy distribution (or the Student-t distribution with degree of freedom 1) and for $\alpha = 0.5$ and $\beta = 1$ we get the Lévy distribution. One of the most interesting properties of stable distributions is that only the $l < \alpha$ moments are finite, with the exception of $\alpha = 2$, where this means that the normal distribution is the only stable distribution with finite variance. It is interesting too, if $\beta \in \{-1, 1\}$, then the distribution is completely skewed respectively to left or right, and additionally, if $\alpha < 1$ is true too, then the support of the density is concentrated only to a half line. If $\beta = 0$, the distribution is symmetric to δ with any α. Also, if α is close to 2, then β doesn't have much impact on the skewness. This can be easily seen from the characteristic function, where if we substitute $\alpha = 2$, the value of the $\beta \tan\frac{\pi\alpha}{2}$ in the characteristic function will be 0, therefore β doesn't play any role.

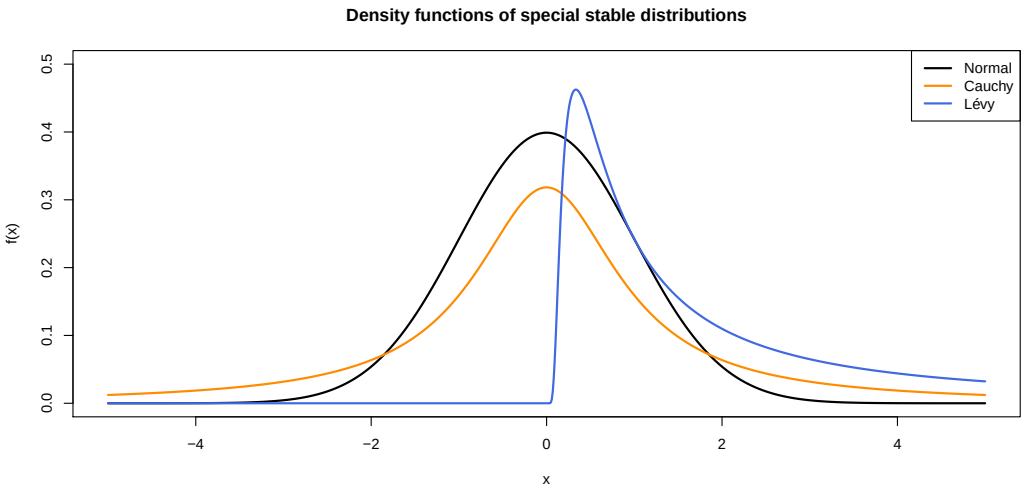


Figure 1: Special stable distributions

Let's take a simple comparison between the normal and the Cauchy distribution. There is significant difference between the variables $X \sim N(0, 2)$ and $Y \sim C(0, 1)$ (which are the same as $S(2, 0, 1, 0)$ and $S(1, 0, 1, 0)$, so the γ parameters are matching): for example $P(X > 5) = 0.00621$ and $P(Y > 5) = 0.06283$. This means the probability to get a value higher than 5, calculated from Cauchy distribution is 10 times larger than calculated from a normal distribution.

We can standardize the distribution the same way as we are used to at normal distributions. By dividing a stable $S(\alpha, \beta, \gamma, \delta)$ r.v. by γ and subtracting with δ, the distribution will be $S(\alpha, \beta, 1, 0)$, which we note by $S(\alpha, \beta)$. This makes the distribution family very flexible in practical use.

Probably the most important theoretical property of stable distributions is the Generalized Central Limit theorem [10].

1.1. Theorem. *The X r.v. is stable, where $0 < \alpha \leq 2$ if, and only if there are $X_1, X_2, \ldots, X_n$ non-degenerate, independent, identically distributed random variables and $a_n, b_n \in \mathbb{R}$ normalizing sequences, so that*

$$\frac{X_1 + X_2 + \ldots + X_n}{b_n} - a_n \xrightarrow{d} X.$$

The main difference between the classical Central Limit theorem and the theorem above is that it doesn't require X to have finite second moment. The consequence of the theorem is that the domain of attraction of stable distributions is not empty. An example from [10] is, that if the tails of X satisfy $x^\alpha P(X > x) \to c^+$ and $x^\alpha P(X < -x) \to c^-$ as $x \to \infty$,

with $c^+ + c^- > 0$ and $1 < \alpha < 2$ (so the expected value μ of X is surely finite), then

$$\frac{X_1 + X_2 + \ldots + X_n}{b_n} - a_n \xrightarrow{d} X \sim S(\alpha, \beta, 1, 0), \quad n \to \infty,$$

where $b_n = \left(\frac{2\Gamma(\alpha)\sin\left(\frac{\pi\alpha}{2}\right)}{\pi(c^+ + c^-)}\right)^{-\frac{1}{\alpha}} n^{\frac{1}{\alpha}}$, $a_n = n b_n^{-1} \mu$, $\beta = \frac{c^+ - c^-}{c^+ + c^-}$ and X represented with the S_1 parametrization (can be found in [10]).

1.3 Parameter estimation

Since stable distributions have no closed form of density function, the parameter estimation is a difficult task. The usual best choice, the maximum likelihood method can be used, but in our case it's based on inverting the characteristic function, which makes it a very slow, computationally heavy method for large samples. The method of moments estimation can't be used, because the moments may not exist. Fortunately, there is another way for estimation: the quantile method, proposed by McCulloch in [5]. This estimation procedure is based on the sample quantiles, which we can compute easily. It is good news, that this method is a consistent estimator for all of the parameters, although becomes biased when $\alpha < 0.5$. This should not be a problem in financial applications, it would be unusual to observe such an α on price fluctuations as it would result in infinite expected value of the returns.

Let x_p be the p-quantile of the distribution and define

$$\nu_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} \text{ and } \nu_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.5}}{x_{0.75} - x_{0.25}}.$$

Both expressions are functions of α and β, since the quantiles are simple linear transforms of each other, with given α and β. As a result of this property, γ and δ are both eliminated in ν_α and ν_β. If $x_p = \gamma z_p + \delta$, where the distribution of z_p is $S(\alpha, \beta)$, then

$$\nu_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} = \frac{\gamma z_{0.95} + \delta - (\gamma z_{0.05} + \delta)}{\gamma z_{0.75} + \delta - (\gamma z_{0.25} + \delta)} = \frac{\gamma(z_{0.95} - z_{0.05})}{\gamma(z_{0.75} - z_{0.25})} = \frac{z_{0.95} - z_{0.05}}{z_{0.75} - z_{0.25}}$$

$$\nu_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.5}}{x_{0.75} - x_{0.25}} = \frac{\gamma z_{0.95} + \delta + \gamma z_{0.05} + \delta - 2(\gamma z_{0.5} + \delta)}{\gamma z_{0.75} + \delta - (\gamma z_{0.25} + \delta)} = \frac{\gamma(z_{0.95} + z_{0.05} - 2z_{0.5})}{\gamma(z_{0.75} - z_{0.25})} =$$

$$= \frac{z_{0.95} + z_{0.05} - 2z_{0.5}}{z_{0.75} - z_{0.25}},$$

therefore both can be written in the from

$$\nu_\alpha = \Phi_1(\alpha, \beta), \quad \nu_\beta = \Phi_2(\alpha, \beta).$$

Having these functions inverted both parameters can be expressed as

$$\alpha = \psi_1(\nu_\alpha, \nu_\beta), \quad \beta = \psi_2(\nu_\alpha, \nu_\beta).$$

These values can be easily calculated for samples, however the inverted functions are lacking a closed form. In practice, both functions are usually evaluated from the distribution functions for different combination of parameters and are tabulated. When estimating, with the results of Φ_1 and Φ_2 we can obtain the values for the parameters by interpolating between the calculated values.

Estimation for γ and δ can be approached in a similar way, but there is a more simple method mentioned in [1], which utilizes the standardizing property of stable distributions. Let $X \sim S(\alpha, \beta, \gamma, \delta)$ and $Z \sim S(\alpha, \beta)$ and x_p, z_p be the p-quantile of X and Z. Then for every $0 < p_1, p_2 < 1$, where $p_1 \neq p_2$:

$$\gamma = \frac{x_{p_2} - x_{p_1}}{z_{p_2} - z_{p_1}} \text{ and } \delta = x_{p_1} - \gamma z_{p_1}.$$

Based on this, γ and δ can be expressed as

$$\gamma = \frac{x_{0.75} - x_{0.25}}{z_{0.75} - z_{0.25}}, \quad \delta = x_{0.5} - \gamma z_{0.5}.$$

2 Multivariate stable distributions

2.1 Definition and properties

Here I follow [7], [8] and [9], which are covering the most important properties of multivariate stable distributions and the possible estimation methods. First, let me define multivariate stable distributions.

2.1. Definition. A random variable $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is a d-dimensional stable vector, if to any $A, B \in \mathbb{R}$ there are $C, \mathbf{D} \in \mathbb{R}^d$ such that

$$A\mathbf{X_1} + B\mathbf{X_2} = C\mathbf{X} + \mathbf{D},$$

where $\mathbf{X_1}, \mathbf{X_2}$ are independent, identically distributed to $\mathbf{X}$ and are d-dimensional random variables.

Our only chance to describe the distribution is still to use characteristic functions apart from the usual few special cases (e.g. the multivariate Cauchy distribution), however the multivariate case is more abstract. In higher dimensions, for a vector variable $\mathbf{X}$ it's characteristic function is defined as $\varphi_{\mathbf{X}}(\mathbf{t}) = E[\exp\{i\mathbf{t}^\mathsf{T}\mathbf{X}\}]$. There are a few different representations of multivariate stable characteristic functions, but the following definition shows us their general form.

2.2. Definition. Let Λ be a finite measure on S_d, where $S_d = \left\{s \in \mathbb{R}^d \colon \|s\|_2 = 1\right\}$, the surface of the unit ball. This measure is called the spectral measure. The $\mathbf{X}$, d-dimensional variable is stable, denoted by $\mathbf{X} \sim S(\alpha, \Lambda, \boldsymbol{\delta})$, where $0 < \alpha \leq 2$ and $\boldsymbol{\delta} \in \mathbb{R}^d$, if it's characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\{-I_{\mathbf{X}}(\mathbf{t}) + i\mathbf{t}^{\mathsf{T}}\boldsymbol{\delta}\},$$

where

$$I_{\mathbf{X}}(\mathbf{t}) = \int_{S_d} \psi\left(\mathbf{t}^{\mathsf{T}}\mathbf{s}; \alpha\right) \Lambda(d\mathbf{s})$$

and

$$\psi\left(u; \alpha\right) = \begin{cases} |u|^{\alpha}\left(1 - i\tan\frac{\pi\alpha}{2} \cdot \operatorname{sign} u\right) & \alpha \neq 1 \\ |u|(1 + i\frac{2}{\pi}\operatorname{sign} u \cdot \log|u|) & \alpha = 1. \end{cases}$$

The $I_{\mathbf{X}}(\mathbf{t})$ determines the shape of the distribution and $\boldsymbol{\delta}$ is the location vector.

As we can see α and δ essentially remained the same, which is not true for β and γ. Instead, the measure Λ takes over their role. Additionally, this measure is what determines the dependence structure of the distribution, which makes the model fitting more complicated, since its non-parametric estimation is not feasible. Because of this, we fit a parametric model later on, where this Λ is discrete, more exactly that Λ is concentrated to a finite number of points. In this case, the measure can be written as

$$\Lambda(\cdot) = \sum_{i=1}^{n} \lambda_i \delta_{s_i}(\cdot)$$

where λ_i are the weights concentrated on $\delta_{\mathbf{s_i}}$ points of mass, $\mathbf{s_i} \in S_d$. With the discrete Λ, the characteristic function of $\mathbf{X}$ simplifies into the following form:

$$\varphi^*(\mathbf{t}) = \exp\left\{-\sum_{i=1}^{n} \psi(\mathbf{t}^{\mathsf{T}}\mathbf{s_i}; \alpha)\lambda_i + i\mathbf{t}^{\mathsf{T}}\boldsymbol{\delta}\right\}. \tag{1}$$

In this form, the characteristic function is easier to handle and can be understood more intuitively.

There is another important property of the distribution family, which we can be familiar with from the normal distribution's characteristics. With the help of the following proposition, we have another way to express multivariate stable distributions.

2.3. Proposition. *If* $\mathbf{X}$ *is* d*-dimensional stable with* $0 < \alpha \leq 2$*, then for every* $u \in \mathbb{R}^d$

$$\mathbf{u}^{\mathsf{T}}\mathbf{X} = u_1X_1 + \ldots + u_dX_d$$

is a univariate stable random variable, with the same α*.*

We note this univariate variable as $\mathbf{u}^\mathsf{T}\mathbf{X} \sim S(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}))$, where α is constant and the functions $\beta(\cdot)$, $\gamma(\cdot)$, $\delta(\cdot)$ completely determine $\mathbf{X}$, as it can be seen in the next theorem.

2.4. Theorem. *Let be* $\mathbf{u}^\mathsf{T}\mathbf{X} \sim S(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}))$. *Then the functions determining* $\mathbf{X}$ *can be written in the following forms:*

$$\gamma(\mathbf{u}) = \left(\int_{S_d} |\mathbf{u}^\mathsf{T}\mathbf{s}|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha} \tag{2}$$

$$\beta(\mathbf{u}) = \gamma(\mathbf{u})^{-\alpha} \int_{S_d} |\mathbf{u}^\mathsf{T}\mathbf{s}|^\alpha \operatorname{sign}(\mathbf{u}^\mathsf{T}\mathbf{s}) \Lambda(d\mathbf{s}) \tag{3}$$

$$\delta(\mathbf{u}) = \begin{cases} \mathbf{u}^\mathsf{T}\boldsymbol{\delta} & \alpha \neq 1 \\ \mathbf{u}^\mathsf{T}\boldsymbol{\delta} - \frac{2}{\pi} \int_{S_d} \mathbf{u}^\mathsf{T}\mathbf{s} \cdot \log\big(|\mathbf{u}^\mathsf{T}\mathbf{s}|\big) \Lambda(d\mathbf{s}) & \alpha = 1. \end{cases} \tag{4}$$

With these, $I_\mathbf{X}(\mathbf{t})$ *can be written as*

$$I_\mathbf{X}(\mathbf{t}) = \begin{cases} \gamma^\alpha(\mathbf{t})(1 - i\beta(\mathbf{t}) \tan \frac{\pi\alpha}{2}) & \alpha \neq 1 \\ \gamma(\mathbf{t})(1 - i\delta(\mathbf{t})) & \alpha = 1. \end{cases} \tag{5}$$

The connection between these properties gives us the opportunity, to be able to determine the multivariate distribution using the univariate projections and to perform calculations more easily. These are giving the base of the estimation procedure, which we can see in section 3.

2.2 Special cases

There are some special multivariate stable distributions worth mentioning, even if they will not be present explicitly in the estimation procedure.

2.5. Proposition. *If the components of* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $X_i \sim S(\alpha, \beta_i, \gamma_i, \delta_i)$ *are independent, then the characteristic function of* $\mathbf{X}$ *can be written as*

$$\varphi_\mathbf{X}(\mathbf{t}) = \exp\left\{ -\sum_{i=1}^{n} \omega(t_i; \alpha, \beta_i) \gamma_i^\alpha + i\mathbf{t}^\mathsf{T}\boldsymbol{\delta} \right\},$$

The independent case could be represented with a discrete Λ, where weights would be only on the intersection between the hypersphere S_d and the axises, so the margins are individually weighted.

2.6. Definition. Let $R \in \mathbb{R}^{d\times d}$ be a positive definite matrix. The $\mathbf{X}$ d-dimensional random variable is elliptically stable, if it's characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{-(\mathbf{t}R\mathbf{t})^{\alpha/2} + i\mathbf{t}^{\mathsf{T}}\boldsymbol{\delta}\right\}.$$

By taking $R = \gamma_0^2 I$, where $I \in \mathbb{R}^{d\times d}$ is the identity matrix, the distribution will be isotropic.

The elliptical stable distributions are similar to normal distributions in how their dependence structure look like, but they allow heavy tails too. The general representation with Λ gives the opportunity to model unusual, non-elliptical dependence structures as well.

2.7. Theorem. *Let $\mathbf{Z}$ be a d-dimensional normal vector with mean vector $\mathbf{0}$ and with covariance matrix $\Sigma \in \mathbb{R}^d$, $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$ and a univariate stable random variable W, with parameters $(\frac{\alpha}{2}, 1, (\cos\frac{\pi\alpha}{4})^{2/\alpha}, 0)$, independent to $\mathbf{Z}$. Then the vector*

$$\mathbf{X} = \delta + \sqrt{W}\mathbf{Z}$$

is also d-dimensional stable, with shift $\boldsymbol{\delta}$. In this case, the characteristic function simplifies to

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{-\left(\frac{1}{2}\mathbf{t}^{\mathsf{T}}\Sigma\mathbf{t}\right)^{\alpha/2} + i\mathbf{t}^{\mathsf{T}}\boldsymbol{\delta}\right\},$$

where $\Sigma_{ij} = \operatorname{cov}(\mathbf{Z}_i, \mathbf{Z}_j)$, $i, j = 1, \ldots, d$, the covariances between the components of $\mathbf{Z}$. The distribution of the projections $\mathbf{u}^{\mathsf{T}}\mathbf{X} \sim S(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}))$ can be described as

- $\beta(\mathbf{u}) = 0$
- $\gamma(\mathbf{u}) = \frac{1}{2}(\mathbf{u}^{\mathsf{T}}\Sigma\mathbf{u})^{1/2}$
- $\delta(\mathbf{u}) = \mathbf{u}^{\mathsf{T}}\boldsymbol{\delta}$,

for all $\mathbf{u} \in \mathbb{R}^d$.

2.3 Simulation

Generating samples from a univariate stable distribution can be done easily with Chambers method, shown in [6]. Let the distribution of W be exponential, with parameter $\lambda = 1$

and $U \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$. We can construct a symmetrical univariate stable random variable, with any $\alpha \in (0, 2]$ as

$$Z = \begin{cases} \dfrac{\sin \alpha U}{(\cos U)^{1/\alpha}} \left(\dfrac{\cos((\alpha - 1)U)}{W} \right)^{(1-\alpha)/\alpha} & \alpha \neq 1, \\ \tan U & \alpha = 1. \end{cases}$$

To construct non-symmetrical stable random variables, we need the constant c, defined from the desired parameters as $c = \dfrac{\arctan(\beta \tan(\frac{\pi\alpha}{2}))}{\alpha}$, $\alpha \neq 1$. Using these, by calculating

$$Z = \begin{cases} \dfrac{\sin \alpha(c + U)}{(\cos \alpha c \cos U)^{1/\alpha}} \left(\dfrac{\cos(\alpha c + (\alpha - 1)U)}{W} \right)^{(1-\alpha)/\alpha} & \alpha \neq 1 \\ \frac{2}{\pi} \left((\frac{\pi}{2} + \beta U) \tan U - \beta \log \dfrac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2} + \beta U} \right) & \alpha = 1, \end{cases}$$

the distribution of Z will be

$$Z \sim \begin{cases} S(\alpha, \beta, 1, \beta\gamma \tan \frac{\pi\alpha}{2}) & \alpha \neq 1 \\ S(\alpha, \beta, 1, 0) & \alpha = 1. \end{cases}$$

Fortunately, we can simulate samples from a multivariate stable distribution [7] with discrete Λ using only univariate stable distributions [9] with Chambers method and the measure itself.

2.8. Proposition. *Let $Z_1, \ldots, Z_n$ be independent, identically distributed, $Z_i \sim S(\alpha, 1, 1, 0)$ random variables. A stable vector $\mathbf{X}$, with discrete Λ, where γ_i are the weights corresponding to $\mathbf{s_i} \in S_d$ can be written as*

$$\mathbf{X} = \begin{cases} \sum_{i=1}^{n} \lambda_i^{1/\alpha} Z_i \mathbf{s_i}, & \alpha \neq 1 \\ \sum_{i=1}^{n} \lambda_i (Z_i + \frac{2}{\pi} \log \lambda_i) \mathbf{s_i}, & \alpha = 1. \end{cases}$$

In practice, this procedure is very fast, making simulation more preferable than actually calculating the density with help of the inversion formula from the characteristic function, which still requires a reasonable amount of time.

3 Paramater estimation, goodness of fit

I go through the bivariate estimation method proposed in [7], [8] and [9], which builds on the distribution's properties mentioned in Section 2. The method is based on the univariate projections and characteristic function of the distribution, where we assume that Λ is discrete. Using these, we get an equation system, whose solution will be the estimation of Λ.

3.1 Estimating procedure

Let $\mathbf{X_1}, \ldots, \mathbf{X_m}$ be our bivariate sample, where we assume that its distribution is bivarite stable. Additionally, we assume that Λ is discrete and concentrated exactly on n points.

Step 1

Firstly, we would like to eliminate the shift $\boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$ from the sample to reduce the characteristic function (1) into a simpler form. This correction makes our latter calculations easier. We do this correction by estimating δ_1 and δ_2 separately e.g. using the quantile method for both marginal distributions and then subtract $\hat{\delta}_1$ and $\hat{\delta}_2$ from the corresponding margin. We can do this, as shifts do not change the distribution nor the other parameters as we could see in Section 1.2. Now the characteristic function should look like

$$\varphi_0(\mathbf{t}) = \exp\left\{ -\sum_{i=1}^{n} \psi(\mathbf{t}^\top \mathbf{s_i}; \alpha)\lambda_i \right\}.$$

Step 2

In the next step, we need to take the points from S_2, where we assume the distribution is concentrated on and where we would like to estimate the weights. Our choice are the points $\mathbf{s_j} = \left(\cos\left(\frac{2\pi(j-1)}{n}\right), \sin\left(\frac{2\pi(j-1)}{n}\right)\right)$, $j = 1, \ldots, n$, which are an equal partition of points on the unit circle. Additionally, we need a grid for the characteristic function $\mathbf{t_1}, \ldots, \mathbf{t_n} \in S_2$, which also determines the $\langle \mathbf{t_j}, \mathbf{X_1} \rangle, \ldots, \langle \mathbf{t_j}, \mathbf{X_m} \rangle$ projections of the distribution. To make the calculation easier, we take these grid points same as the points from S_2, so that $\mathbf{t_j} = \mathbf{s_j}$, $j = 1, \ldots, n$.

Step 3

For every projection, we estimate the value of (2), (3), and if $\alpha = 1$, then the value of (4) too, although it is very unlikely for the estimate of α to be equal to 1. To be able to do

that, we need to use quantile or ML methods to estimate α on the constructed projections. Since α is constant, we take a pooled modification of the parameter as $\hat{\alpha}^* = \frac{1}{n}\sum_{j=1}^{n}\hat{\alpha}(\mathbf{t_j})$. We know that for every projection, α should be exactly the same, but with estimation, α will be slightly different on every projections due to the noise in the sample. By increasing the sample sizes, the estimated α parameters would stabilize for every projection since both univariate estimation methods are consistent. After evaluating (2), (3) and (4), we can calculate the estimated values of $I_{\mathbf{X}}(\mathbf{t})$ for every projection.

Step 4

Since Λ is discrete, $I_{\mathbf{X}}(\mathbf{t})$ can be written into the form: $I_{\mathbf{X}}(\mathbf{t}) = \sum_{j=1}^{n}\psi\left(\mathbf{t}^{\mathsf{T}}\mathbf{s_j};\hat{\alpha}^*\right)\lambda_j$. Based on that, define the $n \times n$ complex matrix $\mathbf{\Psi}$ as

$$\mathbf{\Psi}(\mathbf{t_1},\ldots,\mathbf{t_n};\mathbf{s_1},\ldots,\mathbf{s_n}) = \begin{bmatrix} \psi\left(\mathbf{t_1}^{\mathsf{T}}\mathbf{s_1};\hat{\alpha}^*\right) & \ldots & \psi\left(\mathbf{t_1}^{\mathsf{T}}\mathbf{s_n};\hat{\alpha}^*\right) \\ \vdots & \ddots & \vdots \\ \psi\left(\mathbf{t_n}^{\mathsf{T}}\mathbf{s_1};\hat{\alpha}^*\right) & \ldots & \psi\left(\mathbf{t_n}^{\mathsf{T}}\mathbf{s_n};\hat{\alpha}^*\right) \end{bmatrix} \tag{6}$$

and the $n \times 1$ unknown vector $\boldsymbol{\lambda} = \left[\lambda_1,\ldots,\lambda_n\right]$, what we are trying to find in the end. We define the vector $\mathbf{I_X}(\mathbf{t}^*) = \left[I_{\mathbf{X}}(\mathbf{t_1}),\ldots,I_{\mathbf{X}}(\mathbf{t_n})\right]$, with the calculated values from the previous step. With these, we can write the equation system

$$\mathbf{\Psi}\boldsymbol{\lambda} = \mathbf{I_X}. \tag{7}$$

By solving (7), we can determine the distribution with $\hat{\boldsymbol{\lambda}}$, however we run into some problems. First of all, $\hat{\boldsymbol{\lambda}}$ will most likely be a complex vector, which we can not really interpret from the perspective of the modeling and neither fortunate for describing the distribution itself. The second problem is, that if the size of the grid is even, then the system (7) is singular. That is because $\psi(-t;\alpha) = \overline{\psi(t;\alpha)}$ and $I_{\mathbf{X}}(-\mathbf{t}) = \overline{I_{\mathbf{X}}(\mathbf{t})}$. Fortunately, we can deal with these problems using some modifications.

Step 4/1

Let's assume that we are trying to find $\hat{\boldsymbol{\lambda}}$ on even number of points ($n = 2k$), where $\mathbf{t_j} = \mathbf{s_j}$ as before. We saw that picking an even number of points is causing singularity in the system, but we can turn the tide in our favor. This case $\mathbf{I_X}(\mathbf{t_i}) = \overline{\mathbf{I_X}(\mathbf{t_{i+k}})}$ and $\psi\left(\mathbf{t_i}^{\mathsf{T}}\mathbf{s_j};\alpha\right) = \overline{\psi\left(\mathbf{t_{i+k}}^{\mathsf{T}}\mathbf{s_j};\alpha\right)}$, as we could see before. By picking these pairs we can see, that

$$\operatorname{Re} I_i = \frac{I_i + I_{i+k}}{2} = \sum_{j=1}^{n}\operatorname{Re}\psi_{i,j}\lambda_j$$

and

$$\operatorname{Im} I_i = -\frac{I_i - I_{i+k}}{2} = \sum_{j=1}^{n} \operatorname{Im} \psi_{i,j} \lambda_j,$$

where $\mathbf{I_X}(\mathbf{t_i}) = I_i$ and $\psi_{i,j} = \psi\left(\mathbf{t_i}^\mathsf{T}\mathbf{s_j}; \hat{\alpha}^*\right)$. This is because every calculated value of $\mathbf{I_X}$ has it's conjugated part in the system as a result of the symmetry of the construction. We can now define a new $n \times 1$ vector with the real and imaginary parts of $I_{\mathbf{X}}(\mathbf{t})$ as

$$\boldsymbol{c} = \Big[\operatorname{Re} I_1, \operatorname{Re} I_2, \ldots, \operatorname{Re} I_k, \operatorname{Im} I_1, \operatorname{Im} I_2, \ldots, \operatorname{Im} I_k\Big]$$

and a new $n \times n$ matrix $\boldsymbol{A}$ as

$$a_{i,j} = \begin{cases} \operatorname{Re} \psi_{i,j}, & i = 1, \ldots, k \\ \operatorname{Im} \psi_{i,j}, & i = k+1, \ldots, n \end{cases}$$

Fortunately, the system $\boldsymbol{A\lambda} = \boldsymbol{c}$ is now non-singular and the solution will be a real vector, however still not usable. The problem is, that the solution may contain negative weights, which we cannot interpret.

Step 4/2

To get non-negative weights we must modify the system once more. To be able to guarantee non-negativity, we redefine the problem as a quadratic programming problem as

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{c} - \boldsymbol{A\lambda}\|^2 = \min_{\boldsymbol{\lambda}} (\boldsymbol{c} - \boldsymbol{A\lambda})^\mathsf{T} (\boldsymbol{c} - \boldsymbol{A\lambda}), \quad \boldsymbol{\lambda} \geq 0.$$

With this last step, we constructed the estimation procedure, which now can be implemented and be used to solve practical problems. Later in the applications, I will estimate Λ with this approach, implemented in R programming langauge.

3.2 Properties

Before going on, we need to note some important facts and properties about the estimation procedure.

- We can approximate the real spectral measure with the discrete Λ theoretically too. Byczkowski, Nolan and Rajput showed in [20], that for a stable vector $\boldsymbol{X}$, with Λ spectral measure, where $0 < \alpha < 2$ exists a discrete Λ^*, so that

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} |p(\boldsymbol{x}) - p^*(\boldsymbol{x})| \leq \epsilon,$$

where $p(\boldsymbol{x})$ is the theoretical density, $p^*(\boldsymbol{x})$ is the corresponding density to Λ^*, $\epsilon > 0$. The proof is long and difficult, therefore I don't include it in here, but can be found in [20]. Additionally, since both the quantile and the maximum likelihood methods are consistent and asymptotically unbiased, using these methods to estimate parameters of the projections give consistent results in the multivariate estimation procedure.

- The number of points of S_2 has to be an even number, $n = 2k$, where $k \in \mathbb{N} \setminus \{1\}$. We saw in step 4/1, that we can perform the necessary transformations only with this restriction. Also $k \neq 1$ is needed too, as it would result a simple estimation on the first marginal distribution with the selected set of points of S_2. Apart from these, n is a free parameter, but picking n as a power of two is the most preferable.

- Picking the appropriate number of points is not trivial. If the chosen n is not large enough, the fitted distribution's dependence structure will not match the sample's. However, if n is too large, the distribution can be overfitted, although theoretically it would give us the best results.

3.3 Goodness of fit

First, we have to check if the individual marginals can be accepted as being stable at all. The most ideal test statistic in our case is the Anderson-Darling method.

3.3.1 Anderson-Darling (AD) test

The AD test is distribution function based: the test measures the quadratic distance between the empirical and theoretical distribution functions, putting more weight on the tails [13]. This is where we want the better fit, because we would like to model the occurrence of extremal events as well as we can. The test statistic is defined as

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x),$$

where $F_n(x)$ is the empirical, $F(x)$ is the theoretical distribution function. For stable distributions, when the parameters of the distribution have to be estimated, the limit distribution of the test statistic is not available, so we must use Monte-Carlo simulation to be able to perform the test. In practice, we calculate the test statistic as

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i - 1}{n} \left(\log z_i + \log(1 - z_{n+1-i}) \right),$$

where $z_i = F(X_i)$ and X_i is the i-th element of the n long ordered sample.

3.3.2 Multivariate testing

Testing a multivariate distribution's fit is a more difficult task. One possible way is to use Kendall functions. I define the function in two dimensions, which can be analogously introduced in higher dimensions.

3.1. Definition. Given two random variables X, Y with joint distribution function F, define the random variable $Z = F(X, Y)$. The Kendall function of X and Y is then defined as

$$K(t) = P\left(Z < t\right).$$

The function transforms the bivariate distribution into one dimension, while keeping the information about the dependence structure. With this transformation, testing goodness of fit is a much easier task, as we can apply Cramér–von Mises type of tests to the functions. When we compute Kendall functions for samples, we have to calculate the corresponding empirical (discrete) functions. First, on a sample with n elements, we have to calculate

$$M_i = \frac{1}{n} \sum_{j \neq i} \mathbf{1}(X_j < X_i, Y_j < Y_i), \quad i = 1 \dots n.$$

In words, for every $i = 1, \dots, n$ we compute M_i as the relative frequency of points in the lower quadrant of (X_i, Y_i). Next, we calculate the empirical cumulative distribution function of the M_i values, which gives us the empirical Kendall function of the distribution [16]. We can use these in testing, if we compute the Kendall function for both the sample and the fitted distribution. Since it has no closed form for multivariate stable distributions, it is approximated by a sample from the distribution.

With the help of the Kendall functions, the test statistics can be defined as the quadratic distance between the two Kendall functions:

$$D = \sum_i (K_n(M_i) - K_n^*(M_i))^2, \tag{8}$$

where K_n and K_n^* are the empirical Kendall functions of the sample and the fitted distribution. Formally, our hypothesises are

H_0 : The distribution of the sample is stable

H_1 : The distribution of the sample is not stable.

Unfortunately, the limit distribution of D is unknown too, so we must use Monte-Carlo sampling in order to be able to perform the test. Once we have done that, the test can be evaluated on the given α significance level. If $X \in \mathcal{X}_0 = \{\mathbf{X} \colon \mathbf{D}(\mathbf{X}) < L_{1-\alpha}\}$, where $L_{1-\alpha}$ is the $(1-\alpha)$-quantile of the test statistics estimated distribution, then we can't reject H_0.

3.3.3 Block bootstrap

Not exactly a goodness of fit method, but fits here the best. The estimation procedure, with high number of points from S_d can result to an overfitted model. One way to check it, is to repeat the same estimation process on bootstrapped samples. There are many different ways to construct these samples for the given sample. Since we deal with financial data, where dependence is usually present, we should use the circular block bootstrap method. The substance of the procedure is to make the original time series $\mathbf{X}$ circular, i.e. create a new time series like $\mathbf{Y}(t^*) = \mathbf{X}(t)_{[t^* \equiv t \mod n]}$, so after the last element the beginning of the time series is attached. Next, we have to split this new series into overlapping blocks, with a given block length b [2]. The optimal block length should be chosen the way as Politis and White suggested in [4]. With the optimal b and the constructed blocks, the new bootstrap samples will be sampled with replacement from the blocks, until the length of the bootstrapped sample will be at least as long as the original sample. This way, the new samples will still have similar dependence structure as the original sample.

To be able to check if our model is overfitted, we have to construct many of these bootstrap samples, then repeat the same estimation process on them. If we have the sufficient number of results, we can construct a confidence interval for every λ_i value and we will be able to compare them to the original results to see if the variance of the originally estimated $\boldsymbol{\lambda}$ is small enough.

4 Application in 2 dimensions

In the applications, I fit multivariate stable distributions to cryptocurrency daily logreturns with large market capitalizations: Bitcoin, Ripple and Litecoin. At first, I think it is necessary to give a brief overview of cryptocurrencies in general, although I will not go into the technological details much.

4.1 Cryptocurrencies

Cryptocurrencies drew a lot attention in the past few years as an alternative investing asset. They are digital assets with way different mechanics compared to stocks or regular currencies. The very first of it's kind was Bitcoin, founded by Satoshi Nakamoto in 2009 as an open-source software. It is still unclear who worked under the name of Satoshi, as nobody ever met him/them and has vanished from every platforms they were present since 2011. With the creation of the currency, he also created the block-chain technology, the base of the currency's mechanics. Since the birth of the Bitcoin, many other cryp-

tocurrencies have been developed. By the time I am writing this, there are more than 3900 different cryptocurrencies, with different mechanics and purposes. Also, the total market capitalization of cryptocurrencies have grown in a rapid rate over the past years and at the moment it is over 420 billion USD (according to `www.coinmarketcap.com`), which is around 2.2% of the total market capitalization of the largest stock exchange of the world, the New York Stock Exchange (19.6 trillion USD).

There are many unique properties and interesting questions around these currencies. Most of them don't have any reserve behind them and have decentralized control, however for example, the cryptocurrency Petro is backed up by Venezuela's reserves of oil and is centralized by the government of Venezuela. That brings up an important question: why do they have any value and what influences the price? It's hard to answer this question directly, because many restrictions are setting back these currencies to be used as a regular currency, therefore they cannot be viewed as money based on it's functions. The Bitcoin, the largest cryptocurrency is not accepted as a medium of exchange in many countries and fully banned from certain nations. It is hard to use it as a measure of value or store of value, because it is so volatile, that it's value from 2013 to 2017 grew more than $+10\,000\%$ with many massive ups and downs. Also, there are built in limitations in Bitcoin's and in other cryptocurrencies mechanism. Firstly, in case of Bitcoin, new coins arise through mining, which is an important part of blockchain technology. Mining is the activity, which verifies new transactions in the system, keeping the blockchain complete with the addition of the new blocks. Whenever someone successfully verifies a transaction, they are rewarded with a given amount of bitcoin. This amount in the early times was 50 bitcoin (BTC) and after every 210 000 verified transactions this amount gets halved (at the moment, 12.5 BTC per transactions is awarded). Also, the maximum supply limit of BTC is 21 million, which is expected to be reached by 2140. After this, no new BTC will be rewarded, so the growth is limited, therefore it cannot be paired to a growing economy, which would require more and more amount of money for the growing number of transactions. Another problem is that the verification of the transactions are requiring more processing power and time as more BTC is rewarded and the difficulty of the verification (solving the cryptographic hash) rises. All these properties are paying roles in how the price changes, but the most important drivers in the price at the moment are the speculative interests. The speculation with Bitcoin and other cryptocurrencies makes so heavy fluctuation in the prices, that it may need to be modeled with the help of new tools and not with the classic models. That's where stable distributions could play a role.

4.2 Data and the univariate estimation, goodness of fit

The data is from April of 2013 to February of 2018. I split the whole data into 10 overlapping, $\approx$ 4 months wide windows with $\approx$ 1000 observations. The first period is the oldest, and the 10th period is the most recent in the applications. At first, I fit stable distributions to Bitcoin and Litecoin daily logreturns. I take these two at first to see how the bivariate fit looks like and to see how the dependence structure between the two asset changes over time. The prices of the three assets can be seen on the figure below.

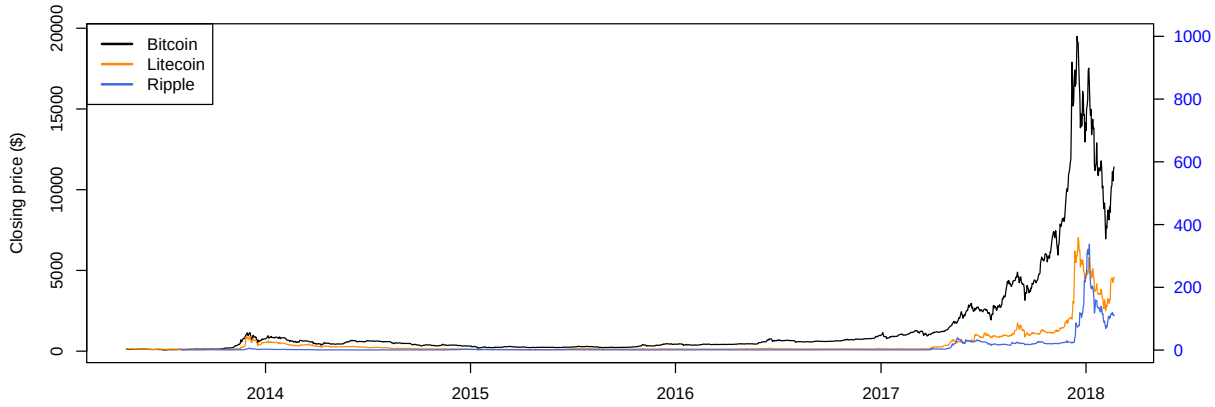


Figure 2: Prices of the selected cryptocurrencies in US dollar. Axis for Bitcoin is on the left, axis for Litecoin and Ripple on the right with different scale. The price of Ripple is multiplied by 100 for better visualization.

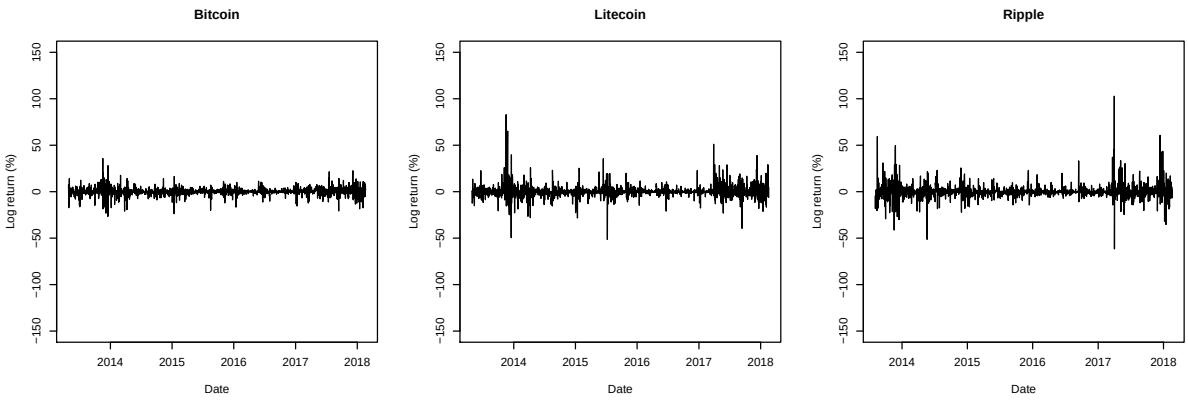


Figure 3: Daily logreturns of the selected cryptocurrencies

To be able to apply multivariate stable distributions to the logreturns, we have to check whether the marginal distributions of them can be accepted as being stable at all.

I estimated the parameters using MLE for α and quantile method for β, γ and δ. The reason to use both methods was necessary, but more on that later. I follow the same logic at the multivariate estimation procedure.

Bitcoin parameters			
α	β	γ	δ
1.319	-0.014	1.768	0.064
1.292	0.043	1.636	0.033
1.254	-0.014	1.418	0.072
1.291	-0.067	1.243	0.069
1.3	0.014	1.131	0.08
1.295	-0.059	1.144	0.163
1.27	-0.005	1.136	0.217
1.273	0.001	1.167	0.244
1.225	-0.054	1.223	0.283
1.184	-0.07	1.381	0.361

Litecoin parameters			
α	β	γ	δ
1.211	0.015	2.058	-0.246
1.15	0.064	1.824	-0.24
1.117	0.016	1.607	-0.121
1.15	-0.041	1.426	0.021
1.144	-0.035	1.21	0.016
1.107	-0.027	1.169	0.012
1.063	0.046	1.175	-0.02
1.058	0.115	1.262	-0.054
1.043	0.149	1.395	-0.078
1.051	0.096	1.5	-0.054

Table 1: The estimated parameters for Bitcoin and Litecoin daily logreturns for every period (from oldest (1.) to the newest (10.)). Note that with time, α decreases, while γ slightly grows after a rapid fall. The effects of these parameter changes are that more probability is getting concentrated on the tails. Also, in the case of Litecoin, a slight positive skewness appears in the last periods.

I performed Anderson-Darling test on Bitcoin and Litecoin logreturns for all periods. I have simulated 1000 test statistic values for both to get critical values, however there were some issues during the calculation. Problem is, that some of the values from the simulation will always be non-interpretable, because when we calculate the test statistic on a sample from a stable distribution having so small α, some values will be so small/large that R won't be able to calculate precisely the distribution function at that value and will truncate the value to 0 or 1. That means, the logarithm in the test statistic can easily get a value of infinity. Fortunately, the vast majority of the simulated values are usable, therefore I could perform the tests. The whole process took about 7 hours, because calculating the probability distribution function for stable distributions is very slow, as it can only be calculated by the help of inverting the characteristic function.

Bitcoin	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Critical Value	2.428	2.387	2.399	2.664	2.343	2.426	2.55	2.659	2.57	2.491
Test statistic	1.709	1.123	0.824	1.232	1.269	1.582	1.777	2.473	3.906	4.054

Litecoin	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Critical Value	2.625	2.591	2.334	2.577	2.358	2.366	2.594	2.548	2.172	1.967
Test statistic	0.747	0.787	1.041	1.807	1.963	1.069	1.011	1.293	1.397	1.663

Table 2: Results of the performed Anderson-Darling tests for every period. Null hypothesis (univariate stability) rejections happened only at Bitcoin's last two tested periods.

The results from the test statistics are promising, nevertheless I got two rejections. This may be caused by the rapid change of volatility over the last years, which is contained in the last two windows. The price fluctuated from a few hundred to over 10 thousand with very different gain/loss asymmetry in these years. In these two cases I still fit multivariate stable distributions as it may show interesting changes in dependence structure between the two currencies.

4.3 Multivariate estimation and goodness of fit

Now we are ready for carrying out the multivariate fit. The estimation is performed with using both ML and quantile method: the former is used for α at the two marginal distributions, and the quantile method for the other parameters on every projection, similarly as we can see in Table 1. This is why the univariate estimation and testing was done that way. The pooled α^* is now calculated only from the estimated α parameters at the marginals. The main reason behind this, is that ML method usually performs better at estimating the tails and the quantile method tends to overestimate them. This can be critical, since a small difference in the value of α means significant change in the probability of extremal events. Also, the other problem is, if we would like to calculate α with ML for all the projections, we should definitely be using the quantile method, because even for one dataset, with ≈ 1000 elements ML runs for at least 5 minutes. This amount of time is acceptable for two marginals, but it is not practical if we would like to use it for every projection. Computing the necessary values for the multivariate parameter estimation with these changes gives us the best results with the least consumed time.

I estimated the measure on 8, 16 and 32 points to be able to compare the different results, as I expect better fit by increasing the number of points. The estimation itself, without re-estimating α for all period run in a few seconds, because quantile method is computationally very light. However, performing the Kendall function based test statistic

with Monte-Carlo simulation took about 33 hours. This is because, I simulated 1000 test statistics for all periods and for the three different approaches. In ideal case, the number of Monte-Carlo simulations should be increased, but even this experimentation took a very long time.

	8 points		16 points		32 points	
Window	Test statistic	Critical value	Test statistic	Critical value	Test statistic	Critical value
1.	1.257	0.355	0.081	0.213	0.074	0.610
2.	1.390	2.342	0.169	1.115	0.049	0.221
3.	1.785	0.376	0.929	0.302	0.351	0.635
4.	2.434	0.563	0.276	1.066	0.525	0.682
5.	2.832	0.684	0.368	0.278	0.441	0.684
6.	2.518	0.495	0.481	0.495	0.423	0.444
7.	2.121	4.093	0.455	0.399	0.286	0.230
8.	0.472	0.986	0.171	0.323	0.306	0.873
9.	2.540	5.315	0.361	1.141	0.214	2.810
10.	6.072	0.457	1.266	0.840	1.280	0.335

Table 3: Results of the performed Cramér–von Mises type test statistics based on Kendall functions for the 3 different estimation approaches. Critical values were chosen at the 95% significance level. The values with red background are the tests, where the null hypothesis has to be rejected.

For the last period, none of the fitted multivariate stable distributions are acceptable. This is not that surprising, as we could see, that for the last period I couldn't fit an univariate stable distribution to Bitcoin's logreturns. For the second, eighth and ninth period, we didn't get rejection, however the Anderson-Darling test for Bitcoin at the ninth period resulted in a rejection. For the rest of the periods, apart from the seventh period, by increasing the number of points on the circle, we usually get acceptable fit. Also, it is important that some of the estimated critical values were extremal compared to the others. This could have been mitigated, if the number of simulations was higher, although most of these extremal values can be seen for one period (9.), so this may be partially caused by the underlying distribution.

To be able to visualize and understand better the estimated spectral measures, it can be useful to look at the figures below. These plots are giving an idea about how the density may look like too, because the spectral measure characterizes the shape of the density.

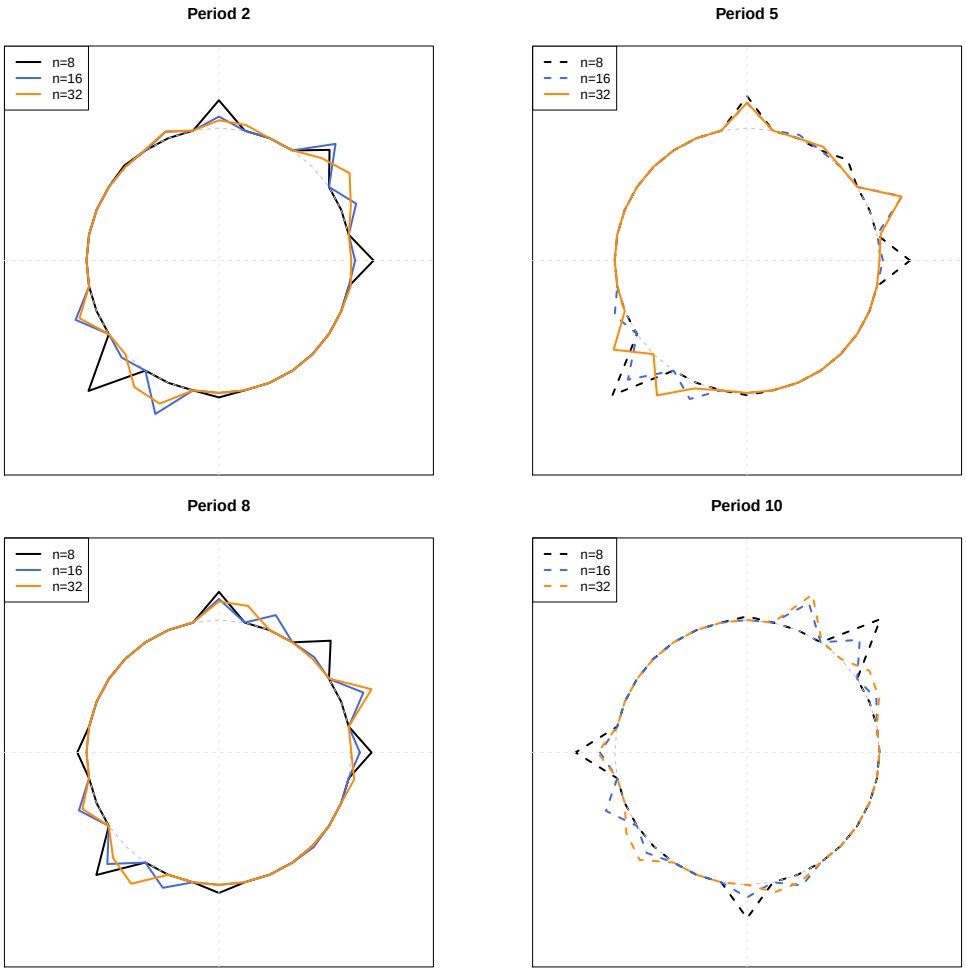
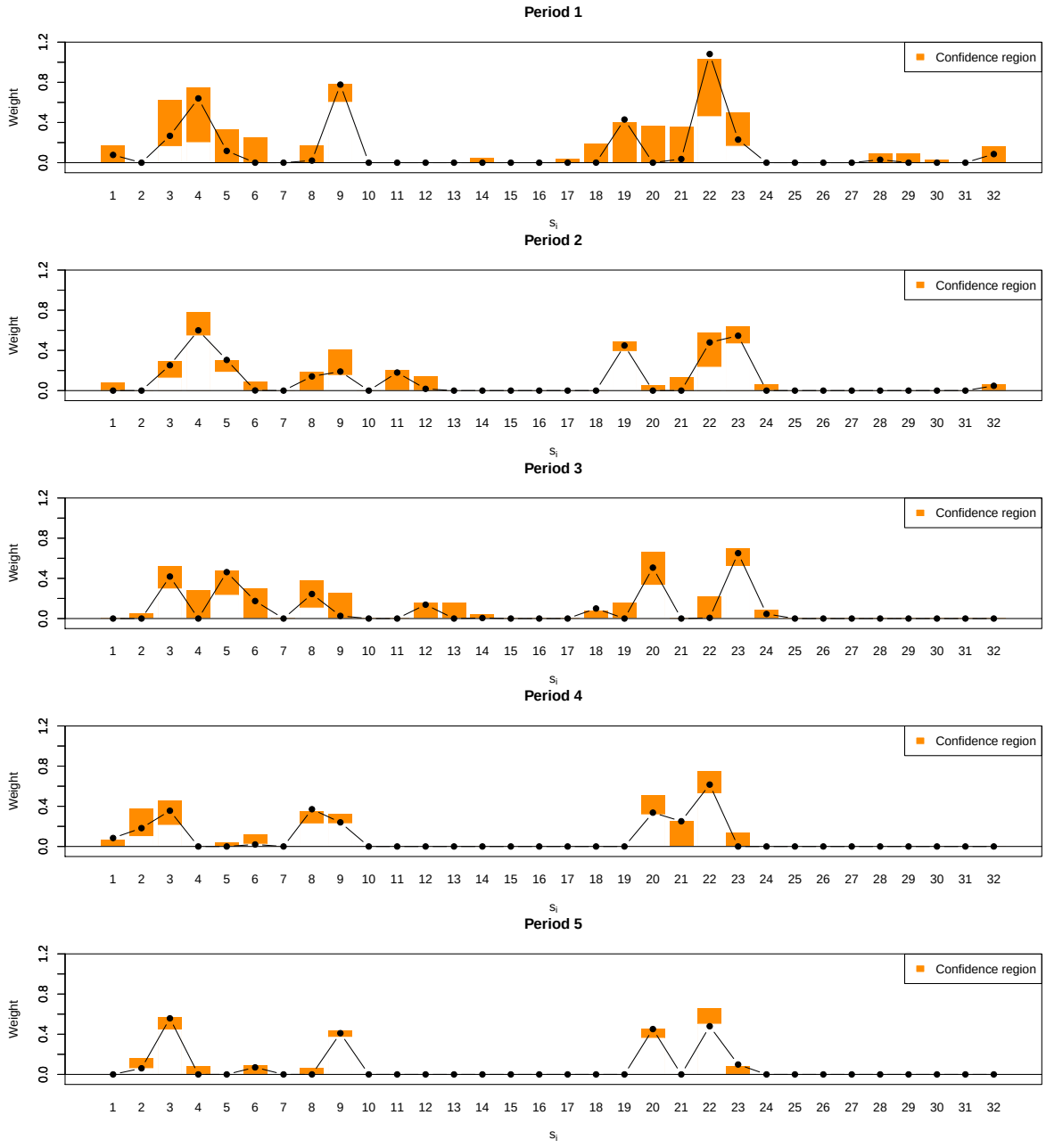


Figure 4: Visualization of the estimated spectral measures for the periods 2-5-8-10, with different number of points. Dashed line means that the fit was not acceptable (parallel to Table 3). Figures for the remaining periods can be found in the Appendix.

The different number of points are giving significantly different estimated spectral measures. With less points from S_2, the measure is concentrated on fewer points and results in a simpler dependence structure. Best example is the measure at period 2, where all the fits are acceptable. It is visible, that the weights in the lower left quadrant are getting spread to more and more points by increasing the number of points. Another good example, is period 5, where with 8 points, the procedure found weight on the very first point s_1, but with 32 points, the method couldn't find any.

I got the best results, when the estimation was done on 32 points, which is expected based on theoretical property showed in [20]. However, in applications, especially in fi-

nancial applications with such number of points we can get overfitted models. This is absolutely true for cryptocurrencies, where the dependence structure can change really fast, due to their unpredictable nature. I checked the fitted distributions (32 points) if they are overfitted by repeating the same estimation procedure, with the same calibration for all the period on 500-500 bootstrapped samples, as I mentioned in section 3.3.3. Unfortunately, I couldn't go over 500 iterations as my computer couldn't handle more then that within a reasonable amount of time. The results can be seen on the figures below.



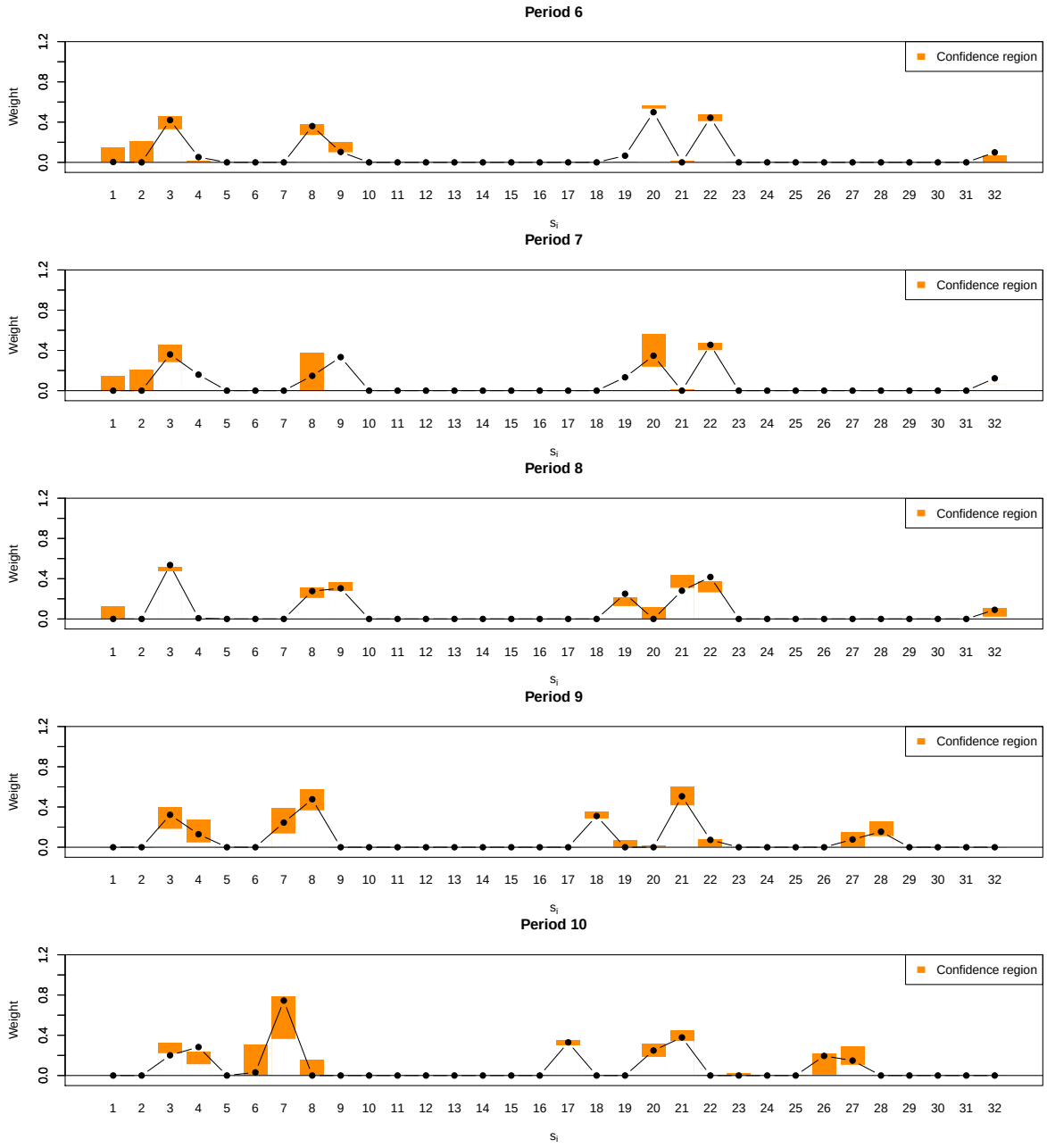
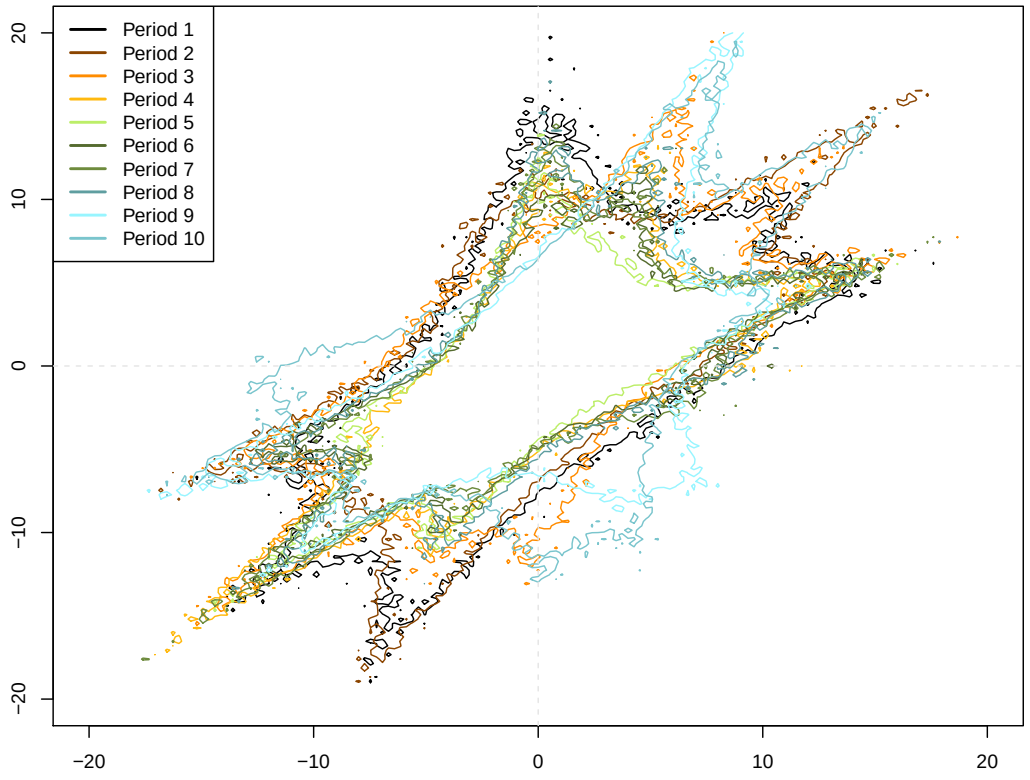


Figure 5: The 95% confidence region for every estimated weight with 32 points from S_2, evaluated on bootstrapped samples. If the width of the confidence region is 0, no line or region is displayed behind the given weight.

The results are showing that the fitted distributions are slightly overfitted. The original weights are sometimes out of the confidence region, however, the method always found the absolute 0 weights. These are mostly the points $s_{10}, \ldots, s_{18}$ and $s_{24}, \ldots, s_{32}$, where

the negative dependence would be present. It is promising, that the method always finds the big weights too, even if they are out of the confidence region. The reason behind this, is that if we increase the number of points during the estimation procedure, the bigger weights are getting split around the adjacent points. A good example is the first period and the point s_{22}, where we can see that with the bootstrapped results, the adjacent points are significantly weighted, but the original result of the estimation only puts bigger weight on s_{22}. My suggestion is that it is unnecessary to go beyond 32 points. Based on the results, any even number of points between 16 and 32 are ideal choices, with some trade-offs. One can lower the points to 16 to get a more stable result, but statistically it may not be acceptable.

By looking at the spectral measures, we can say that the dependence structure changed over time, but the best way to tell this is to look at the densities. Although, it would take a lot of time to compute the actual density function, we can simulate a larger sample from the given distribution and run a simple density estimation on it. The results of these for every period can be seen on the figures below.



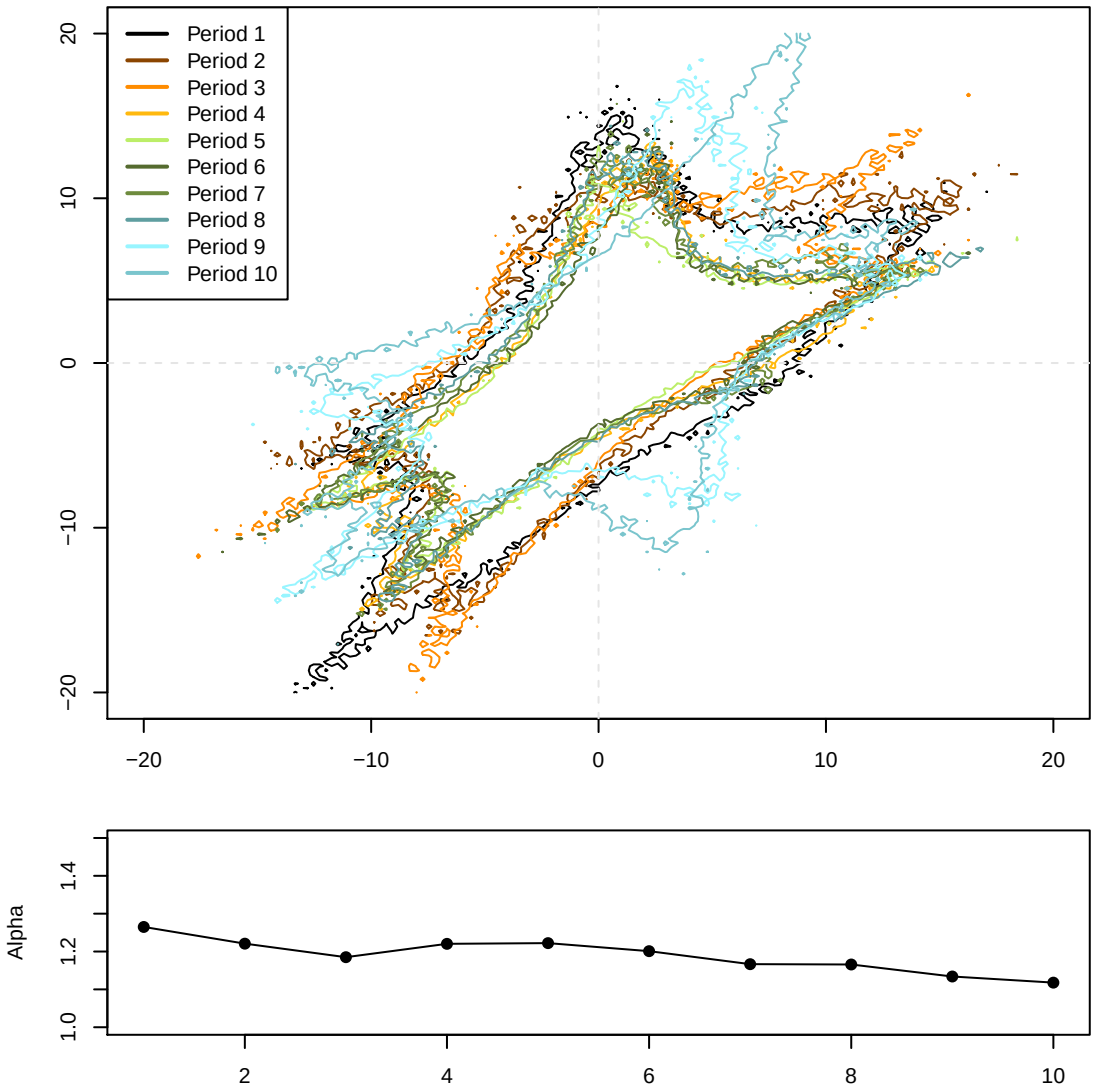


Figure 6: The 95% probability covering regions based on the density estimations with the estimations on 16 points (first) and 32 points (second) from S_2, and the estimated pooled α^* for all periods. Distribution of Bitcoin and Litecoin logreturns on the horizontal and vertical axises in order.

With time, the dependence structure visibly changes with both calibrations. The two results are similar, the angles with bigger weights are mostly present at both. The two dominant angles at the upper right quadrant are shifting with time, at first far from each other then in the end closer together, while moving back and forth in the interim periods. The lower left quadrant also has two dominant angles, both moving back and forth with time. In the last period both changes, having shifted these angles closer to the horizontal

axis and a new angle appears in the lower right quadrant, giving more probability to opposite movement in price changes. The contour lines around the covering region are nothing like the classic elliptic contour lines we are used to, e.g. the normal distribution. This is partially caused by the low estimated α, giving heavy tails to the distribution and making the already dominant angles more dominant and spreading the covering region into a larger area.

We can easily calculate probabilities or risk measures with the distribution. I calculate probabilities, because a simple VaR should be calculated from the sum of the variables and the goal is to take into consideration the dependence structure directly. For comparison, I'm calculating the conditional probability

$$P(\text{Litecoin logreturn} < -10\% \mid \text{Bitcoin logreturn} < -10\%)$$

based on the estimated stable distributions on 32 points and fitted bivariate normal distributions.

Period	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Normal	0.88%	0.31%	0.17%	0.25%	0.29%	0.23%	0.61%	2.95%	3.19%	3.51%
Stable	46.5%	59.02%	76.81%	73.68%	72.91%	80%	89.84%	61.96%	66.98%	71.76%

Table 4: Calculated conditional probabilities. The probabilities from the stable distribution are calculated from samples with the given parameters.

The results are very illustrative. The conditional probabilities calculated from the fitted normal distributions are always around a few percent, but the probabilities from the fitted stable distributions are huge compared to them. These results are in line with the extremal (tail) independence and dependence of the normal and stable distributions. We say that a distribution has extremal independence or dependence, if the probabilities

$$\theta_l = \lim_{q \to 0} P(Y < F_2^{-1}(q) | X < F_1^{-1}(q)),$$
$$\theta_u = \lim_{q \to 1} P(Y \geq F_2^{-1}(q) | X \geq F_1^{-1}(q))$$

tend to zero or nonzero, where F_1 and F_2 are the two distribution functions of X and Y [15]. Although -10 isn't an extreme quantile, these properties can already be observed from the calculations.

5 Parameter estimation in higher dimensions

In $d > 2$ dimensions the estimation procedure gets a bit more difficult, because Λ is concentrated on a sphere. The main difficulty is to select a proper set of points from

the surface of the sphere, where we can repeat the same modifications as in Section 3. I haven't found any sources, where the higher dimensional estimation was studied in any way, therefore I show a simple generalization of the estimation, built on the previously seen bivariate method, which has a fast running time.

The key of estimating parameters for a $d > 2$ dimensional stable distribution is to select the points of S_d pairwise from the marginals. My suggestion is to pick the points from the circular cross section of the sphere, where one coordinate is always 0 for a given circle. This means, by picking n points pairwise, we will perform the estimation, based on $\binom{d}{2} \cdot n$ points.

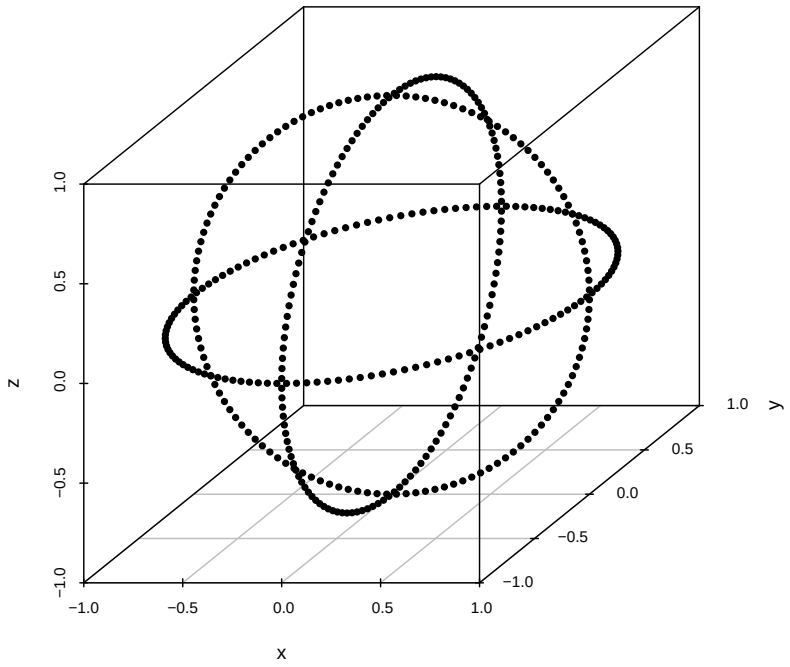


Figure 7: The suggested set of points from S_3.

Assume that the distribution of the $\mathbf{X_1}, \ldots, \mathbf{X_m}$ d-dimensional sample is stable. Also, we assume that Λ is discrete and concentrated on $\binom{d}{2} \cdot n$ points.

Step 1

The first step can be analogously done based on the bivariate estimation method. This case, we have to estimate the vector $\boldsymbol{\delta} = \left[\delta_1, \ldots, \delta_d\right]^\mathsf{T}$ by components with e.g. quantile

method, which we have to subtract from the original sample to have the sample shifted to the origin.

Step 2

We saw in Section 3, that the number of points from a circle had to be even. We still need this assumption, but for every individual circular cross section of the sphere. Additionally, we can't pick the same points from every circular cross section, because then we would be having duplicated points at the intersections of them and it would give us uninterpretable results. Therefore we pick the points rotated as

$$\mathbf{s_j^{l,k}} = \left(0, \ldots, 0, \underbrace{\cos\left(\frac{2\pi(j-1)}{n} + \frac{\pi}{n}\right)}_{\text{l-th coordinate}}, 0, \ldots, 0, \underbrace{\sin\left(\frac{2\pi(j-1)}{n} + \frac{\pi}{n}\right)}_{\text{k-th coordinate}}, 0, \ldots, 0 \right),$$

where $\mathbf{s_j^{l,k}}$ is the circular cross section from the sphere constructed for the l-th and k-th marginals, $l \neq k$, $j = 1, \ldots, n$. We pick the grid points as $\mathbf{t_j^{l,k}} = \mathbf{s_j^{l,k}}$ to be able to compute the projections $\langle \mathbf{t_j^{l,k}}, \mathbf{X_1} \rangle, \ldots, \langle \mathbf{t_j^{l,k}}, \mathbf{X_m} \rangle$.

Step 3

We have to calculate (2), (3) and (4) for every projection as before. The pooled α remains essentially the same, the only real difference is that it is calculated from more projections as $\hat{\alpha}^* = \frac{1}{\binom{d}{2} \cdot n} \sum_{l \neq k} \sum_{j=1}^{n} \hat{\alpha}(\mathbf{t_j^{l,k}})$. After these, we can compute every

$$\mathbf{I_{l,k}}(\mathbf{t_j}) = \sum_{j=1}^{n} \psi\left(\mathbf{t}^\top \mathbf{s_j}; \hat{\alpha}^*\right) \lambda_j^{l,k}$$

values, where $l \neq k$ and $j = 1, \ldots, n$. However, the equation system we solved in Section 3 needs to be modified.

Step 4

The modified system is

$$\mathbf{\Psi^*} = \begin{bmatrix} \mathbf{\Psi_{1,2}} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{\Psi_{d-1,d}} \end{bmatrix}, \quad \mathbf{I^*} = \begin{bmatrix} \mathbf{I_{1,2}} \\ \mathbf{I_{1,3}} \\ \vdots \\ \mathbf{I_{d-1,d}} \end{bmatrix},$$

where $\mathbf{\Psi^*} \in \mathbb{R}^{\binom{d}{2}\cdot n \times \binom{d}{2}\cdot n}$, contains every calculated $\mathbf{\Psi_{l,k}}$ matrices, which are the same as (6), but calculated from the l-th and k-th marginals. The $\mathbf{\Psi^*}$ matrix has the $\mathbf{\Psi_{l,k}}$ matrices in it's diagonal, while it's other elements are zero. The vector $\mathbf{I}^* \in \mathbb{R}^{\binom{d}{2}\cdot n}$ is modified with the same logic as $\mathbf{\Psi^*}$, so it contains every $\mathbf{I_{l,k}}$ vectors combined together. So now, we are ready to write the equation system

$$\mathbf{\Psi^* \lambda^*} = \mathbf{I^*}_{\mathbf{X}}, \tag{9}$$

which also has to be modified, because it is singular too due to the symmetrical construction of the points.

Step 4/1

We have to restrict the method to even number of points ($n = 2r$) as before. Now, it is true, that $\mathbf{I_X}(\mathbf{t_i^{l,k}}) = \overline{\mathbf{I_X}(\mathbf{t_{i+r}^{l,k}})}$ and $\psi\left(\mathbf{t_i}^\mathsf{T}\mathbf{s_j}; \alpha\right) = \overline{\psi\left((\mathbf{t_{i+r}}^{\mathbf{l,k}})^\mathsf{T}\mathbf{s_j}; \alpha\right)}$. We have to do the same transformation on the system as before in Section 3, so we calculate the vectors

$$\operatorname{Re} I_i^{l,k} = \frac{I_i^{l,k} + I_{i+r}^{l,k}}{2} = \sum_{j=1}^{n} \operatorname{Re} \psi_{i,j}^{l,k} \lambda_j$$

$$\operatorname{Im} I_i^{l,k} = -\frac{I_i^{l,k} - I_{i+r}^{l,k}}{2} = \sum_{j=1}^{n} \operatorname{Im} \psi_{i,j}^{l,k} \lambda_j,$$

where $\mathbf{I_X}(\mathbf{t_i^{l,k}}) = I_i^{l,k}$ and $\psi_{i,j}^{l,k} = \psi\left((\mathbf{t_i^{l,k}})^\mathsf{T}\mathbf{s_j}; \hat{\alpha}^*\right)$. We now define the new $\binom{d}{2} \cdot n \times 1$ vector with the real and imaginary parts of $I_{\mathbf{X}}(\mathbf{t^{l,k}})$ as

$$\boldsymbol{c^*} = \left[\operatorname{Re} I_1^{1,2}, \operatorname{Im} I_1^{1,2}, \operatorname{Re} I_2^{1,2}, \operatorname{Im} I_2^{1,2}, \ldots, \operatorname{Re} I_r^{1,2}, \operatorname{Im} I_r^{1,2}, \operatorname{Re} I_1^{1,3}, \operatorname{Im} I_1^{1,3}, \ldots, \operatorname{Re} I_r^{n-1,n}, \operatorname{Im} I_r^{n-1,n}\right]$$

and the new $\binom{d}{2} \cdot n \times \binom{d}{2} \cdot n$ matrix $\boldsymbol{A^*}$ as

$$a_{i,j}^* = \begin{cases} \operatorname{Re} \psi_{i,j}^{1,2}, & i,j = 1, \ldots, r \\ \operatorname{Im} \psi_{i,j}^{1,2}, & i,j = r+1, \ldots, n \\ \operatorname{Re} \psi_{i,j}^{1,3}, & i,j = n+1, \ldots, n+r \\ \operatorname{Im} \psi_{i,j}^{1,3}, & i,j = n+r+1, \ldots, 2n \\ \vdots & \\ \operatorname{Re} \psi_{i,j}^{d-1,d}, & i,j = \binom{d}{2}(n-1)+1, \ldots, \binom{d}{2}(n-1)+r \\ \operatorname{Im} \psi_{i,j}^{d-1,d}, & i,j = \binom{d}{2}(n-1)+r+1, \ldots, \binom{d}{2}n \end{cases}$$

Now the system $\boldsymbol{A^*\lambda^*} = \boldsymbol{c^*}$ is non-singular and real, so we can perform the last modification step, to get positive weights.

Step 4/2

We can analogously redefine the problem as a quadratic programming problem with $\mathbf{A}^*$ and $\mathbf{c}^*$:

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{c}^* - \boldsymbol{A}^*\boldsymbol{\lambda}^*\|^2 = \min_{\boldsymbol{\lambda}^*}(\boldsymbol{c}^* - \boldsymbol{A}^*\boldsymbol{\lambda}^*)^\mathsf{T}(\boldsymbol{c}^* - \boldsymbol{A}^*\boldsymbol{\lambda}^*), \quad \lambda \geq 0.$$

The solution gives us the desired results for $\boldsymbol{\lambda}$.

6 Application in 3 dimensions

Now we are able to look into fitting 3 dimensional stable distributions to the data. I fit the distribution to all the three cryptocurrency's logreturns that I showed in section 4.2. For that, I must perform the preliminary estimation and testing.

In this case, I do the fitting only on the last three periods, which are the equivalent to the 8.,9. and 10. periods in section 4.2. We have to keep in mind, that the AD test resulted in rejection for Bitcoin's last two period and the Kendall function based test completely rejected the last period for the three calibration. Despite these problems, these periods are the most interesting for us, because the drastic changes in the prices happened in these periods. The estimation and testing is done with the same logic as before, therefore I am estimating α with MLE, β, γ and δ with quantile method.

Ripple parameters			
α	β	γ	δ
1.168	0.212	1.565	-0.479
1.168	0.223	1.619	-0.485
1.114	0.177	1.882	-0.473

Table 5: The estimated parameters of the fitted stable distribution calculated from Ripple logreturns for the three periods.

The univariate estimations for the periods gave similar results to Bitcoin's and Litecoin's. The α is decreasing, while γ is rising as time passes. The β shows a significant skewness to the right, however the shift δ is always negative.

Ripple	1.	2.	3.
Critical Value	2.818	2.154	2.269
Test statistic	1.020	1.255	1.209

Table 6: The results of the Anderson-Darling tests for the distribution of Ripple logreturns.

The AD test didn't reject the null-hypothesis for any of the three periods, all the value of the test statistics are under the critical values chosen on 95% confidence level. Unfortunately, many of the simulated test statistic values were non-interpretable again, but enough remained usable to evaluate the tests.

I performed the multivariate goodness of fit tests too with Kendall functions. In 3 dimensions, the empirical Kendall functions are calculated from the values

$$M_i = \frac{1}{n} \sum_{j \neq i} \mathbf{1}(X_j < X_i, Y_j < Y_i, Z_j < Z_i), \quad i = 1 \dots n.$$

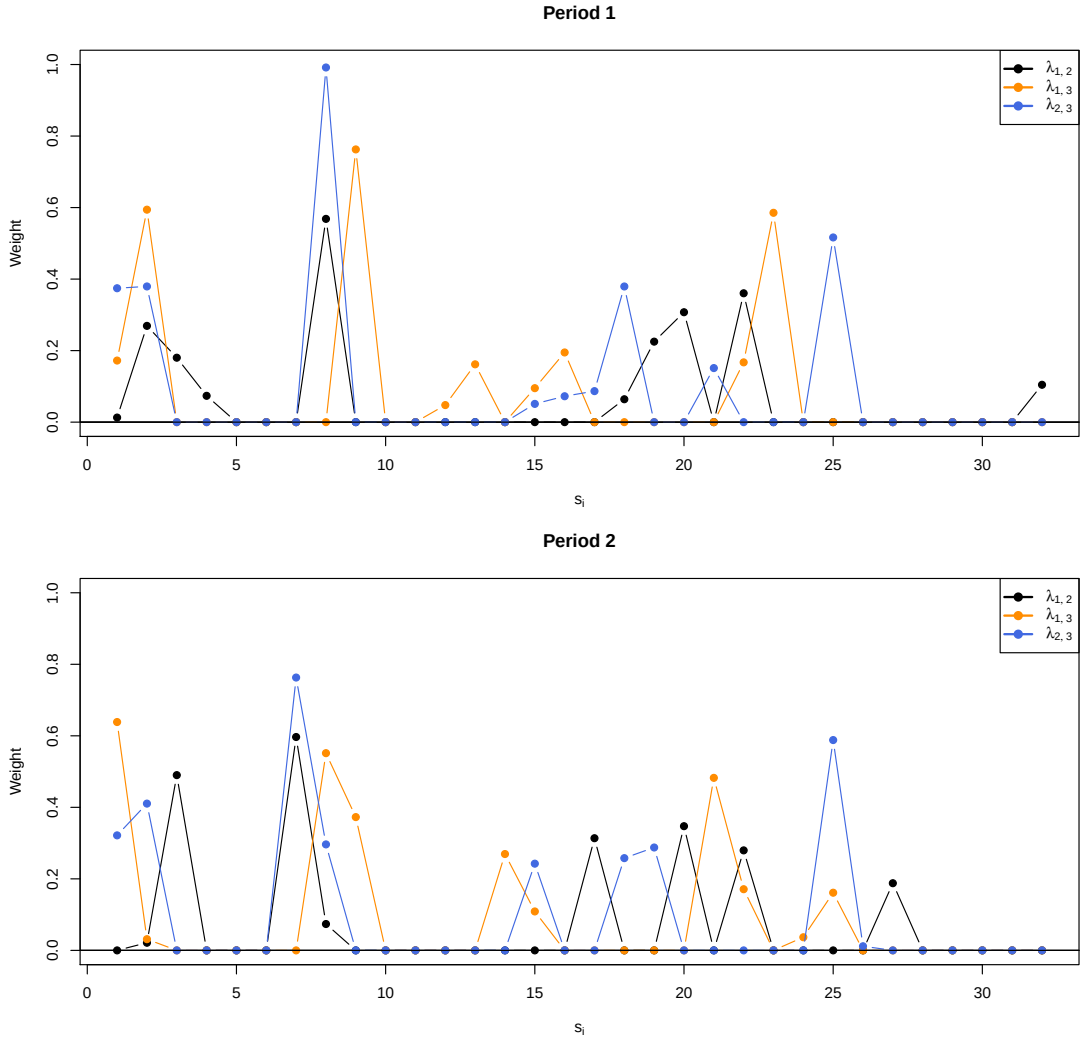
In words, for a given point triplet we have to count how many points fall under it within all three coordinates, then from the calculated values, we construct the empirical Kendall function of the distribution and we are ready for testing and for the simulation. I performed the fitting of the distributions with three different point calibrations again, with 8, 16 and 32 points, but now from the earlier mentioned circular cross section of the sphere. In 3 dimensions this means I did the fitting on $\binom{3}{2} \cdot 8 = 24$, $\binom{3}{2} \cdot 16 = 48$ and $\binom{3}{2} \cdot 32 = 96$ points in total, which is a significant raise in the number of parameters. Since I only selected three periods now, the total running time of the Cramér–von Mises tests were way lower than before.

	$\binom{3}{2} \cdot 8$ points		$\binom{3}{2} \cdot 16$ points		$\binom{3}{2} \cdot 32$ points	
Window	Test statistic	Critical value	Test statistic	Critical value	Test statistic	Critical value
1.	2.142	3.501	4.802	1.633	1.453	3.388
2.	1.052	4.173	2.077	0.659	0.460	2.076
3.	0.996	3.661	1.278	4.303	0.890	0.490

Table 7: Results of the performed Cramér–von Mises type test statistics with three different calibrations. Critical values were chosen based on 95% significance level as before. The values with red background are the tests, where the null hypothesis had to be rejected.

The results are interesting, however they are not really parallel to the result seen in Table 3. The estimations based on 16 points per circular cross sections of the sphere were the worst of all, two periods got absolutely rejected. The third period is not rejected, but it can be generally said that the values of the test statistic are all higher with 16 points than the other two approaches. Based on the results, estimation on 8 points was the best overall. The testing for the third period with 32 points is rejected too, but these results may have changed, if the simulation were done using more repetitions.

Since the spectral measure is now concentrated on the surface of a sphere, visualizations gets more difficult. Density plots are not feasible, so I show the spectral measures, but on simplified figures, showing every $\boldsymbol{\lambda}^{l,k}, l \neq k$, when the estimation was done on 32 points per circles.



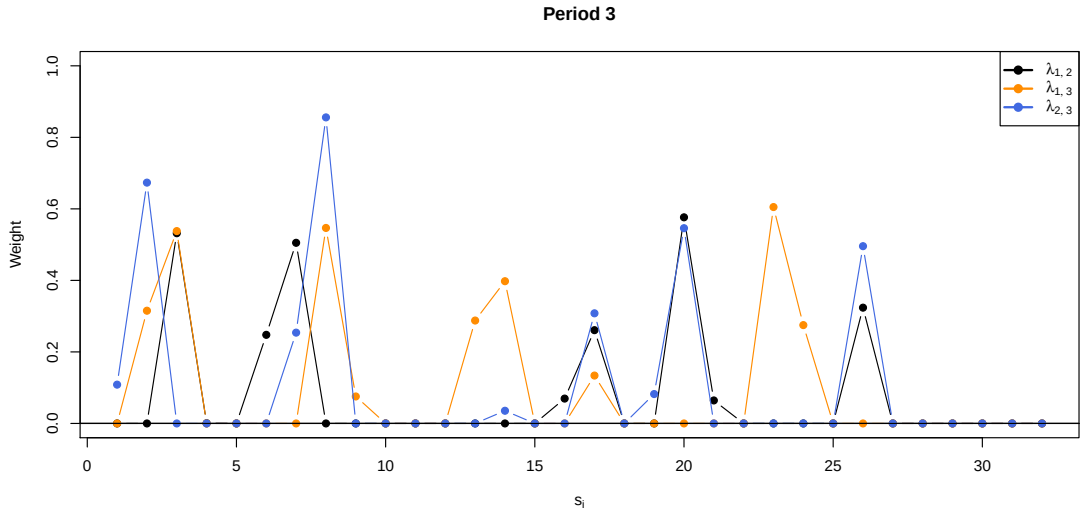


Figure 8: Estimated spectral measures (32 points per circles). Notion for the margins: 1–Bitcoin, 2–Litecoin, 3–Ripple.

It is visible, that every possible pair of cryptocurrency logreturns have overall similar dependence structures, so the three asset's price seems to react in a similar way to each other and to new information on the market. The change in the dependence structure is similar to what we could see on Figure 6: two dominant angles are present in the positive region of $\mathbb{R}^3$, caused by the weights $\boldsymbol{\lambda}_1^{l,k}, \ldots, \boldsymbol{\lambda}_{10}^{l,k}$, which are getting closer to each other as time passes. The weights in the negative region of $\mathbb{R}^3$ are showing some realignment, focusing more onto a fewer density points, creating more dominant angles.

Despite the difficulties in visualizing the density, it is easy to calculate probabilities. Here, I take all three cryptocurrencies into consideration and estimate the probability

$$P(\text{Ripple logreturn} < -10\% \mid \text{Bitcoin logreturn} < -10\%, \text{Litecoin logreturn} < -10\%)$$

from the fitted normal and earlier fitted stable distributions. The results are a bit different though, as the probability calculated from the normal distributions are higher now, unlike in Table 4.

Period	1.	2.	3.
Normal	30.243%	22.018%	20.21%
Stable	87.903%	86.385%	63.481%

Table 8: Calculated conditional probabilities from the 3 dimensional normal and stable distributions, fitted to the logreturns.

The probabilities based on the stable distributions are still significantly higher than the ones from the normal. It is interesting that for both, the probabilities decrease parallel to each other, but with different intensity.

7 Summary

The thesis started off with a general overview of univariate stable distributions, stressing the most important and interesting properties along with the often used parameter estimating methods. After these, multivariate stable distributions were introduced, where several key properties were presented. With the help of these features and the univariate stable distributions, we could construct a general bivariate estimation method.

In the applications, after a brief overview of cryptocurrencies we could see that the univariate stable distributions, apart from a few cases always had a good fit on the given set of cryptocurrency logreturn data based on the AD tests. Also, the changes of the estimated parameters nicely reflected the changes in the characteristics of the returns. At the bivariate estimations, we could see 3 different point calibrations. The results of the goodness of fit tests clearly showed that by increasing the number of points, the fit becomes better, although becomes slightly overfitted. With the estimated spectral measures, we could observe how the dependence structure changed over time through the visualized spectral measures and the estimated densities. It was easy to detect the changes, as there were several angles, where the density was highly concentrated on.

After the application of the bivariate procedure, a higher dimensional parameter estimation procedure was proposed, which heavily builds on the earlier seen bivariate method. The application of this new approach was successfully used in 3 dimensions, although results were harder to be visualized in this case, but we could see the essence of the results. We can say, that fitting stable distributions to logreturns of cryptocurrencies in the tested dimensions can be used very well, nevertheless the distribution family was rejected by a few authors, when modeled stock returns. Based on the results, stable distributions could be used for modeling the price changes of cryptocurrencies.

Let me bring attention to a few properties that I didn't mention before. Despite that I had promising results, there are a few problems with stable distributions in applications. First, when we are simulating from stable distributions with such low α that we could see before, there will be unusually big or small values in our samples. This is why it is necessary to use ML method for estimating α, which usually doesn't underestimate α. Second, we need sufficient amount of elements in the samples to be able to perform the univariate estimation as best we can. Low sample sizes will cause the estimations to result in false parameters α. Third, for the multivariate estimation, if the estimated α parameters of the marginal distributions differ too much, we shouldn't try to fit multivariate stable distribution. In this case, the pooled α^* would give absolutely false results, because every individual α should be close to each other. This is a consequence of the Proposition 2.3. Also, e.g. if we have estimations of the α parameters with values 0.9 and 1.1, the

corresponding first marginal would have infinite expected value, while the second distributions expected value is finite. The pooled α^* would suppose a finite expected value, therefore we can't fit a multivariate distribution, even though the two estimations are not that far from each other.

The cryptocurrency data are from `www.kaggle.com`. These data and many more cryptocurrencies soon will be available in the `crypto` package of R. The calculations of probabilities and sampling from univariate stable distribution were done with the help of [22] package, while the univariate parameter estimations were performed with [23]. I used both of them for my own codes in the multivariate estimation and sampling, along with the package [24], which solves the QP problem. Anderson-Darling tests were done with [21], the determination of optimal block lengths for bootstrapping with [26] and the density estimations with [25].

References

Papers:

[1] A. W. Barker, *Estimation of Stable Distribution Parameters from a Dependent Sample*, Sydney, 2014

[2] András Zempléni, *Price fluctuations lecture notes*, 2018

[3] Benoit Mandelbrot, *The variation of certain speculative prices*, The Journal of Business, Vol. 36, No. 4, pp. 394-419, Oct., 1963

[4] D. Politis, H. White, *Automatic Block-Length Selection for the Dependent Bootstrap*, 2003

[5] J. Huston McCulloch, *Simple consistent estimators of stable distribution parameters*, Marcel Dekker, 1986

[6] J. M. Chambers, C. L. Mallows, B. W. Stuck, *A method for simulating stable random variables*, 1976

[7] J. P. Nolan, *An overview of multivariate stable distributions*, 2008

[8] J. P. Nolan, A. K. Panorska, J. H. McCulloch, *Estimation of stable spectral measures*, 2001

[9] J. P. Nolan, *Modeling financial data with stable distributions*, 2005

[10] J. P. Nolan, *Stable Distributions*, 2014

[11] J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser, 2018, Boston, Chapter 1

[12] Juuso Töyli, *Essays on asset return distributions*, Helsinki University of Technology Laboratory of Computational Engineering Publications, 2002

[13] Stephens, M. A., *EDF Statistics for Goodness of Fit and Some Comparisons*, Journal of the American Statistical Association, 69, pp. 730-737., 1974

[14] Q. M. Shao, *Do Stock Returns Follow a Finite Variance Distribution?*, Singapore Management University, 2001

[15] Rafael Schmidt, *Tail dependence of elliptically contoured distributions*, Math Meth Oper Res, 2002

[16] Roger B. Nelsena; Jose Juan Quesada-Molina, Jose Antonio Rodriguez-Lallena, Manuel Ubeda-Flores, *Kendall distribution functions*, 2002

[17] Szabolcs Majoros, *Stabilis eloszlások és alkalmazásuk a pénzügyekben, BSc thesis*, 2016

[18] Szabolcs Majoros, *Parameter estimation of multivariate stable distribution (in Hungarian)*, TDK paper (Scientific Students' Associations), II. prize, 2016

[19] Szabolcs Majoros, András Zempléni, *Applying bivariate stable distributions to daily logreturns of stocks*, conference poster presented at the 8th Annual Financial Market Liquidity Conference, 2017

[20] Tomasz Byczkowski, J. P. Nolan, Balram Rajput, *Approximation of multidimensional stable densities*, 1993

Packages:

[21] Carlos J. Gil Bellost, *ADGofTest*: Anderson-Darling GoF test. R package version 0.3. http://CRAN.R-project.org/package=ADGofTest, 2011

[22] Diethelm Wuertz, Martin Maechler and Rmetrics core team members, *stabledist*: Stable Distribution Functions. R package version 0.7-0. http://CRAN.R-project.org/package=stabledist, 2015

[23] Rmetrics Core Team, Diethelm Wuertz, Tobias Setz and Yohan Chalabi, *fBasics*: Rmetrics - Markets and Basic Statistics. R package version 3011.87. http://CRAN.R-project.org/package=fBasics, 2014

[24] S original by Berwin A. Turlach R port by Andreas Weingessel <Andreas.Weingessel@ci.tuwien.ac.at>, *quadprog*: Functions to solve Quadratic Programming Problems.. R package version 1.5-5. http://CRAN.R-project.org/package=quadprog, 2013

[25] Tarn Duong, *ks*: Kernel Smoothing. R package version 1.10.3. http://CRAN.R-project.org/package=ks, 2016

[26] Tristen Hayfield and Jeffrey S. Racine, Nonparametric Econometrics: The *np* Package. Journal of Statistical Software 27(5). URL http://www.jstatsoft.org/v27/i05/., 2008

8 Appendix

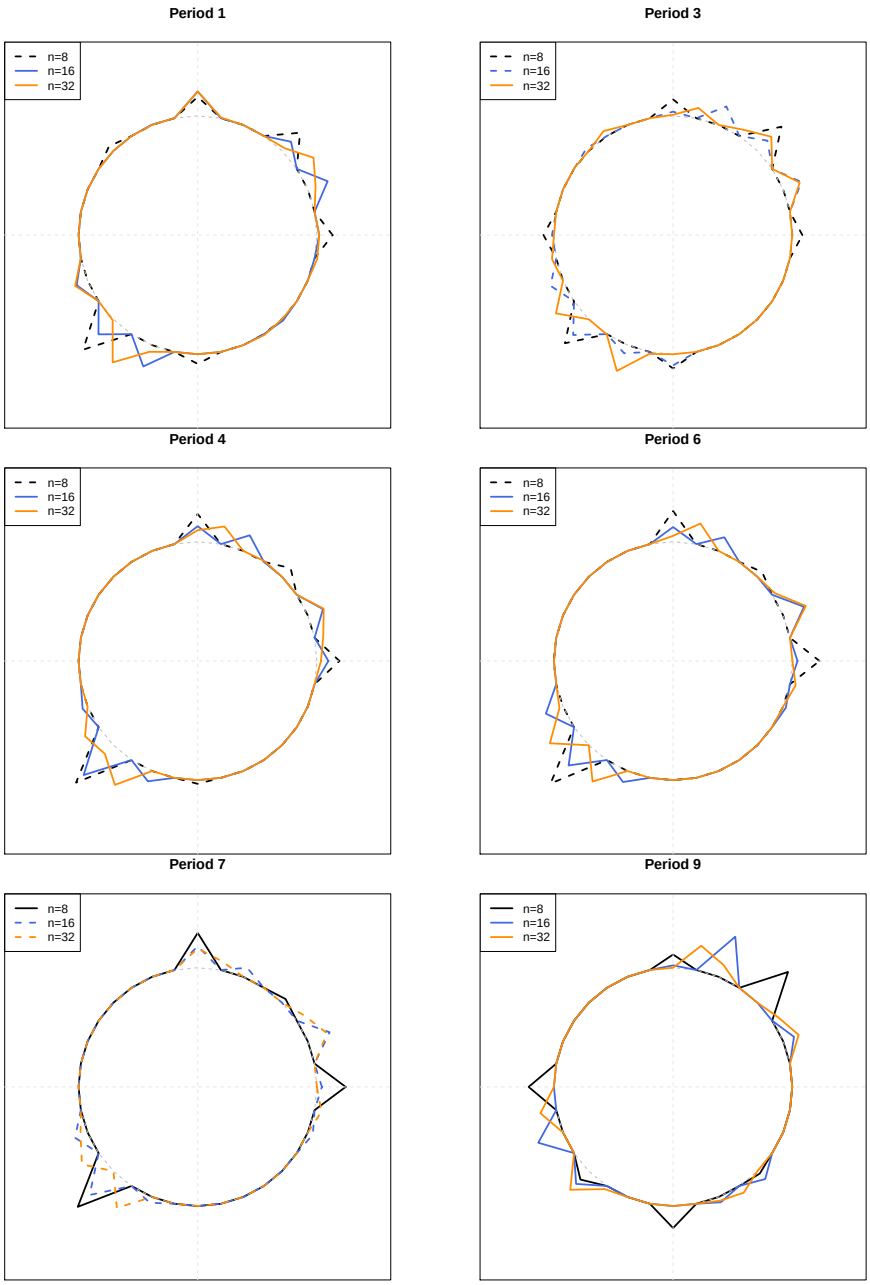


Figure 9: Visualization of the estimated spectral measures for the periods 1-3-4-6-7-9, with different number of points. Dashed line means that the fit was not acceptable (parallel to Table 3).