Hedging Cryptos with Bitcoin Futures

Francis Liu*

Meng-Jou Lu †

Natalie Packham‡

Wolfgang Karl Härdle§¶

This version: November 25, 2021

Abstract

The introduction of derivatives on Bitcion enables investors to hedge risk exposures in cryptocurrencies. We investigate different methods of determining the optimal hedge ratio when hedging various cryptocurrencies and crypto-portfolios with Bitcoin futures. Because of volatility swings and jumps in cryptocurrency prices, the traditional variance-based approach to obtain hedge ratios is infeasible. As a consequence, we consider two extensions of the traditional approach: first, different dependence structures are modelled by different copulae, such as the Gaussian, Student-t, Normal Inverse Gaussian and Archimedean copulae; second, different risk measures, such as value-at-risk, expected shortfall and spectral risk measures, are employed to find the optimal hedge ratio. Various measures of hedge effectiveness in out-of-sample tests give insights in the practice of hedging Bitcoin, Ethereum, Cardano, the CRIX index and a number of crypto-portfolios in the time period December 2017 until May 2021. We find that ... *[needs to be amended.]*

JEL classification: C38, C53, F34, G11, G17

Keywords: Cryptocurrencies, risk management, hedging, copulas (delete: Portfolio Selection, Spectral Risk Measurement,l Coherent Risk)

*Department of Business and Economics, Berlin School of Economics and Law, Badensche Str. 52, 10825 Berlin, Germany. Blockchain Research Center, Humboldt-Universität zu Berlin, Germany. International Research Training Group 1792, Humboldt-Universität zu Berlin, Germany. E-mail: `Francis.Liu@hwr-berlin.de`.

†Department of Finance, Asia University, 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan Department of Finance, Asia University, 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan E-mail: `mangrou@gmail.com`.

‡Department of Business and Economics, Berlin School of Economics and Law, Badensche Str. 52, 10825 Berlin, Germany. International Research Training Group 1792, Humboldt-Universität zu Berlin, Germany. E-mail: `packham@hwr-berlin.de`.

§Blockchain Research Center, Humboldt-Universität zu Berlin, Germany. Wang Yanan Institute for Studies in Economics, Xiamen University, China. Sim Kee Boon Institute for Financial Economics, Singapore Management University, Singapore. Faculty of Mathematics and Physics, Charles University, Czech Republic. National Yang Ming Chiao Tung University, Taiwan. E-mail: `haerdle@wiwi.hu-berlin.de`.

¶Financial support of the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Deutsche Forschungsgemeinschaft's IRTG 1792 grant, the Yushan Scholar Program of Taiwan the Czech Science Foundation's grant no. 19-28231X / CAS: XDA 23020303, as well as support by Ansar Aynetdinov (`ansar.aynetdinov@hu-berlin.de`) are greatly acknowledged.

Contents

1 Introduction

Cryptocurrencies (CCs) are a growing asset class. Many more CCs are now available on the market since the first cryptocurrency Bitcoin (BTC) surfaced (Nakamoto, 2009). In response to the rapid development of the CC market, the CME Group launched exchange-traded BTC futures contracts in December 2017. Trading volume in BTC futures surpassed \$ 2 trillion in 2020 (CryptoCompare, 2020). *[CryptoCompare not in references; possibly add as footnote (if it's a website, not an academic reference).]*

By April 2021, the market value of outstanding coins had risen to \$ 2.3 trillion, more than 6% of the world's narrow money supply and almost 3% of the world GDP.*[Is this open interest in futures? Then a comparision with money supply and GDP is tricky. Or is 2.3 trillion the USD value of mined coins?]* The price of BTC even surged to \$ 64,500 in mid-April 2021 up by 460% from from \$ 11,500 six months earlier in October 2020 and up by 850% from a year earlier. Just a month later, by mid-May 2021, the price had fallen to \$ 50,000, a one-month return of -22.5%. More individual and institutional investors are adding CCs and CC derivatives into their portfolios, creating the need to understand downside risks and find suitable ways to hedge against extreme risks. Fom a risk management perspective, the roller-coaster ride of crypto prices creates significant basis risk, even when using simple hedges involving crypto portfolios and BTC futures. This requires analysing the dependence structure of cryptos and futures beyond linear correlation.

In this paper, we analyse static hedges of crypto portfolios with Bitcoin futures. Owing to the asymmetry of crypto returns as well as the occurrence of extreme events, we consider different dependence structures via a variety of copula models and we optimise the hedge ratio using different risk measures. A similar study was conducted by (Barbi and Romagnoli, 2014) for equity and FX portfolios.

The hedge ratio is the appropriate amount of futures contracts to be held in order to eliminate risk exposure in the underlying security. The determination of the optimal hedge ratio relies primarily on the dependence between BTC and futures prices. Copulae provide the flexibility to model multivariate random variables separately by their margins and dependence structure. The concept of copulae was originally developed (but not under this name) by Wassily Hoeffding (Hoeffding, 1940a) and later popularised by the work of Abe Sklar (Sklar, 1959).

Different risk measures account for investors' risk attitudes. They serve as loss functions in the searching process of the optimal hedge ratio. Of the vast literature discussed the relationship between risk measures and investor's risk attitude, we refer readers to Artzner et al. (1999) for an axiomatic, economic reasoning approach of risk measure construction; Embrechts et al. (2002) for reasoning of using Expected Shortfall (ES) and Spectral Risk Measures (SRM) in addition to VaR; Acerbi (2002) for direct linkage between risk measures and investor's risk attitude using the concept of a "risk aversion function".

Financial asset returns have long known to be non-Gaussian, see e.g. (Fama, 1963; **?**). Specifically, Gaussian models cannot produce the heavy tails and the asymmetry observed in asset returns, which in turn implies a consistent underestimation of financial risks. Therefore, to minimize down risk, one cannot solely rely on second-order moment calculations. Moreover, variance as a risk measure does not account for the variety of investors' utility functions. In particular, it is known that investors are tail-risk averse, see **?**, Bollerslev et al. (2015) find that the jump tail risk is more closely associated with changes in risk-aversion. *[Unclear. Do investors constantly change their risk aversion?]* It is important to link the investor utility's functions as hedging the tail risk. *[Careful. We do not do this* *in our paper, so maybe tone down.]* As such, significant tail risks lead to the need to investigate even static hedges with more refined methods than minimising the variance assuming normally distributed asset returns(Ederington and Salas, 2008).

In order to capture a variety of risk preferences, in addition to variance, we include the risk measures value-at-risk (VaR), expected shortfall (ES), and spectral risk measures (SRM). VaR is widely used by the finance industry and easy to understand. ES and SRM are chosen because of their coherence property, in particular, they recognize diversification benefits. SRM can also be directly related to an individual's utility function. Examples are the exponential SRM and power SRM introduced by Dowd et al. (2008).

[The paragraph below largely repeats what has been said earlier. I suggest to take what is new and add it to the earlier paragraph. There is no need to introduce formal notation at this stage.] This paper considers hedging BTC using its future. i.e. to find an optimal hedge ratio h^* such that the risk of a hedged portfolio $r^h = r^S - h^* r^F$ has minimal risk. Here r^S as the log return of BTC spot price, r^F the log return of BTC future. The leptokurtic properties mentioned above leads us to deploy a comprehensive way of modelling dependency namely copulae together with various risk measures as loss function to find optimal hedge ratio. We first calibrate the log returns of BTC and CME futuresby copulae, then find the optimal quantity of assets in the hedged portfolio according to a range of risk measures. Barbi and Romagnoli (2014) use the C-convolution operator introduced by Cherubini et al. (2011) to derive the distribution of linear combination of margins with copula as their dependence structure. *[The terminology C-convolution operator does not appear again in the paper. Either remove or denote where this is defined.]* We slightly amend their lemma and come up with a formula for the linear combination of random variables for our purpose.

This paper is organized as follows. Section 2 introduces the notion of optimal hedge ratio; section 3 decribes the method of estimation of copulae; section 4 provides the empirical result; section 5 concludes.All calculations in this work can be reproduced. The results are reproducible with data and codes available on www.quantlet.com.

2 Optimal hedge ratio

[Please note it is futures contract, not future contract.]

We form a portfolio with two assets, a spot asset and a futures contract, for example Bitcoin spot and a CME Bitcoin futures contract. Our objective is to minimize the risk of the exposure in the spot. To keep a simple portfolio setting, we go long one unit of the spot and short h units of the future, $h \geq 0$. Letting r^S and r^F be the log returns of the spot and futures price, the log return of the portfolio is *[Where does this formula come from? Log returns are not additive across assets, therefore the formula is wrong. This would hold for discrete returns. Otherwise, write that the portolio return is approximated by this formula.]*

$$R^h = R^S - hR^F.$$

[Capitalised returns as they are random variables.] If the portfolio reduces the risk of the spot position, then we call this a hedge portfolio. (was: We call this portfolio a hedged portfolio: the price movement of spot is hedged by the price movement of future.)

To measure risk, we define a risk measures ρ to be a mapping from a financial position, such as R^h, to a real number, which is often interpreted as the amount of money to make the position acceptable (e.g. to a regulator), see e.g. (?). (was: Risk is measured by a risk measure. Assume the payoff r^h of a hedge portfolio lives in a probability space, $r^h \in L(\Omega, \mathcal{F}, \mathbb{P})$, and there is a risk measure on r^h $\rho : r^h \mapsto \mathbb{R}$.) Hedging refers to finding the optimal hedge ratio (OHR) h^* that minimizes the risk,

$$h^* = \underset{h}{\operatorname{argmin}}\, \rho(R^h).$$

(delete, as redundant with what was said above: Most risk measures are defined as functionals of the portfolio loss distribution F_{r^h}, i.e. $\rho : F_{r^h} \mapsto \mathbb{R}$.) For example, Value-at-Risk (VaR) at the confidence level α is the absolute value of the $1-\alpha$-quantile of R^h, i.e., $\mathrm{VaR}_{1-\alpha} = -F_{R^h}^{(-1)}(1-\alpha) = -\inf\{x \in \mathbb{R} : 1-\alpha \leq F_{R^h}(x)\}$, where F_{R^h} is the distribution function of R^h. (delete: We need the knowledge of F_{R^h} in order to measure risk.) The density f_{R^H} of R^h is obtained from the joint density of R^S and $-h\,R^F$ by convolution, i.e., $f_{R^h}(z) = \int_{-\infty}^{\infty} f_{R^S,-h\,r^F}(x, z-x)dx$, see e.g. (Härdle and Simar, 2019). (was: By convolution of random variables (Härdle and Simar, 2019), $f_{r^h}(z) = \int_{-\infty}^{\infty} f_{r^S,-hr^F}(x, z-x)dx$, where $f_{r^S,-hr^F}$ is the joint pdf of r^S and $-hr^F$. Obviously the cdf of r^h and the risk measure depend on the joint distribution of r^S and $-hr^F$.)

Optimising h according to $f_{r^S,-hr^F}$ is unfavorable in the sense that one would need to calibrate the joint pdf $f_{r^S,-hr^F}$ whenever updating h. This is not only time-consuming, but also unnecessary, as we show below. (was: This is too time consuming and unnecessary.) Another problem of using the joint pdf is that one lacks the flexibility to model the margins separately from the dependence structure. (delete: A joint pdf completely determines the form of its marginals, for example, margins of a bivariate t-distribution are univariate t-distributions.) To overcome both of these problems, we use copulae. The benefit of using copulae is two fold. First, copulae allow us to model the margins and dependence structure separately, see Sklar's Theorem (Sklar, 1959). Second, copulae are invariant under strictly monotone increasing function (Schweizer et al., 1981), see the Lemma below. See e.g. (Nelsen, 1999; Joe, 1997; McNeil et al., 2005) for Sklar's Theorem:

Theorem 1 (Hoeffding Sklar Theorem) *Let F be a joint distribution function with margins F_X, F_Y. Then, there exists a copula $C : [0,1]^2 \mapsto [0,1]$ such that, for all $x, y \in \mathbb{R}$*

$$F(x, y) = C\{F_X(x), F_Y(y)\}. \tag{1}$$

If the margins are continuous, then C is unique; otherwise C is unique on the range of the margins.

Conversely, if C is a copula and F_X, F_Y are univariate distribution functions, then the function F defined by (27) is a joint distribution function with margins F_X, F_Y.

Indeed, many basic results about copulae can be traced back to early works of Wassily Hoeffding (Hoeffding, 1940b, 1941). The works aimed to derive a measure of relationship of variables, which is invariant under change of scale. See also Fisher and Sen (2012) for English translations of the original papers written in German. The following Lemma is not hard to prove. *[Give a source of the Lemma. Is it in the papers above? Or give a proof.]*

Lemma 1

$$C_{X,hY}\{F_X(s), F_{hY}(t)\} = C_{X,Y}\{F_X(s), F_Y(t/h)\}. \tag{2}$$

Leveraging these two features of copulae, Barbi and Romagnoli (2014) introduce the distribution of linear combinations of random variables using copulae. We slightly edit the Corollary 2.1 of their work and yield the following expression of the distribution.

Proposition 2 *Let X and Y be two real-valued continuous random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with absolutely continuous copula $C_{X,Y}$ and marginal distribution functions F_X and F_Y. Then, the distribution function of Z is given by*

$$F_Z(z) = 1 - \int_0^1 D_1 C_{X,Y}\left[u, F_Y\left\{\frac{F_X^{(-1)}(u) - z}{h}\right\}\right] du. \tag{3}$$

Here, $F^{(-1)}$ denotes the inverse of F, i.e., the quantile function.

Here $D_1 C(u,v) = \dfrac{\partial}{\partial u} C(u,v)$ and, see e.g. Equation (5.15) of (McNeil et al., 2005),

$$D_1 C_{X,Y}\{F_X(x), F_Y(y)\} = \mathbf{P}(Y \le y | X = x). \tag{4}$$

Proof. *[Use $\mathbb{E}$ instead of E.]*

Using the identity (4) gives

$$F_Z(z) = \mathbf{P}(X - hY \le z) = \mathsf{E}\left\{\mathbf{P}\left(Y \ge \frac{X - z}{h}\Big| X\right)\right\}$$

$$= 1 - \mathsf{E}\left\{\mathbf{P}\left(Y \le \frac{X - z}{h}\Big| X\right)\right\} = 1 - \int_0^1 D_1 C_{X,Y}\left[u, F_Y\left\{\frac{F_X^{(-1)}(u) - z}{h}\right\}\right] du.$$

■

Corollary 1 *Given the formulation of the above portfolio* [Restate in terms of random variables not portfolio], *the pdf of Z can be written as*

$$f_Z(z) = \left|\frac{1}{h}\right| \int_0^1 c_{X,Y}\left[F_Y\left\{\frac{F_X^{(-1)}(u) - z}{h}\right\}, u\right] \cdot f_Y\left\{\frac{F_X^{(-1)}(u) - z}{h}\right\} du, \tag{5}$$

or

$$f_Z(z) = \int_0^1 c_{X,Y}\left[F_X\left\{z + hF_Y^{(-1)}(u)\right\}, u\right] \cdot f_X\left\{z + hF_Y^{(-1)}(u)\right\} du. \tag{6}$$

The two expressions are equivalent. Note that the pdf of Z in the above proposition is readily accessible, as long as we have the copula density and the marginal densities. The proof and a generic expression can be found in the appendix.

3 Empirical Procedure

[The title is too unspecific. How about "Methodology to determine the optimal hedge ratio"?]

We introduce the empirical procedure to obtain the optimal hedge ratio (OHR). First, we split the time series of spot and futures into sets of training and testing data. The training data makes up the first 300 observations and its corresponding testing data consists of the consecutive five observations. We then roll five observations forward to obtain the next training and test data sets and repeat this until the end of the time series. Note that the testing data are non-overlapping.

Next, we obtain the OHR as follows:

1. **Construct Univariate Kernel Density Function (KDE).** From the training data we calibrate the spot and futures' univariate kernel density functions using the Gaussian kernel with bandwidth determined by the refined plug-in method (Härdle et al., 2004, section 3.3.3).

2. **Calibrate Copulae.** We then calibrate the copulae outlined in 3.1 via the method of moments described in 3.3.1.

3. **Select Copula.** We compute the Akaike Information Criterion. The copula with the best (i.e., lowest) AIC is used for the next step. A discussion of this step is found in 3.5.

4. **Determine OHR.** We determine the OHRs numerically using different risk measures as the loss function by drawing samples from the selected copula and KDEs. The risk measures used as risk reduction objectives are outlined in 3.6

5. **Obtain testing log-return of hedged portfolio.** Finally, we apply the OHRs to the test data $r_h = r_s - h^* r_f$.

3.1 Copulae

As seen in the last section, the risk measures are all functionals of the joint distribution of R^S and R^F. To capture different aspects of the dependence structure, we therefore consider a number of different copulas, which are layed out in details below: (was: The following copulae are considered:) Gaussian-, t-, Frank-, Gumbel-, Clayton-, Plackett-, mixture, and factor copula. As this hedging exercise concerns only portfolios with two assets, we focus on the bivariate version of copulae and some important features of a copula, including Kendall's τ,

formula here,

Spearman's ρ,

formula here,

upper tail dependence

$$\lambda_U \stackrel{\text{def}}{=} \lim_{q \to 1^-} \mathbf{P}\{X > F_X^{(-1)}(q) | Y > F_Y^{(-1)}(q)\}$$

and lower tail dependence

$$\lambda_L \stackrel{\text{def}}{=} \lim_{q \to 0^+} \mathbf{P}\{X \leq F_X^{(-1)}(q) | Y \leq F_Y^{(-1)}(q)\}.$$

Furthermore, we denote the Fréchet-Hoeffding lower bound by $\boldsymbol{W}$, the product copula by $\boldsymbol{\Pi}$, and the Fréchet-Hoeffding upper bound by $\boldsymbol{M}$. They represent cases of perfect negative dependence, independence, and perfect positive dependence, respectively. For further details, we refer readers to Joe (1997) and Nelsen (1999); see also Härdle and Okhrin (2010).

3.1.1 Elliptical Copulae

[As no definition of elliptical copulas is given (or needed), I suggest to call the section "Gaussian and t-copula" and just mention that they belong to the wider class of elliptical copulas. Alternatively, have one setion for Gaussian and one for t-copula.]

Elliptical copulae are dependence structures derived from elliptical distributions. A special case is the (bivariate) Gaussian copula, defined as

$$
\begin{aligned}
\boldsymbol{C}(u,v) &= \Phi_{2,\rho}\{\Phi^{-1}(u),\Phi^{-1}(v)\} \\
&= \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right\} \mathrm{d}s\,\mathrm{d}t, \qquad (7)
\end{aligned}
$$

[Note the difference between a variable d and the differential operator d.*]* where $\Phi_{2,\rho}$ is the cdf of bivariate Normal distribution with zero mean, unit variance, and correlation coefficient ρ, and Φ^{-1} is the quantile function univariate standard normal distribution. Note that we use ρ to denote the correlation parameter as well as a $\rho(\cdot)$ to denote a risk measure. (was: Please note that we use ρ here to represent the correlation parameter in a Gaussian copula only for traditional purposes. In other sections, $\rho(\cdot)$ is a risk measure.) The Gaussian copula is fully specified by the correlation parameter ρ. Like all elliptical copulas, it is symmetric. It has no tail dependence, which, in a finance context, implies that it often underestimates tail risk.

The Gaussian copula density is

$$
\begin{aligned}
\boldsymbol{c}_\rho(u,v) &= \frac{\boldsymbol{\varphi}_{2,\rho}\{\Phi^{-1}(u),\Phi^{-1}(v)\}}{\varphi\{\Phi^{-1}(u)\}\cdot\varphi\{\Phi^{-1}(v)\}} \\
l &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\}, \qquad (8)
\end{aligned}
$$

where $\boldsymbol{\varphi}_{2,\rho}(\cdot)$ is the pdf corresponding to $\Phi_{2,\rho}$, and $\varphi(\cdot)$ the standard normal distribution pdf.

Kendall's τ_K and Spearman's ρ_S of a bivariate Gaussian *[Use a consistent notation, either τ_K or τ; likewise ρ_S or ρ.]* Copula are

$$
\tau_K(\rho) = \frac{2}{\pi}\arcsin\rho \qquad (9)
$$

$$
\rho_S(\rho) = \frac{6}{\pi}\arcsin\frac{\rho}{2}. \qquad (10)
$$

The Student t-copula has the form

$$
\begin{aligned}
\boldsymbol{C}(u,v) &= \boldsymbol{T}_{2,\rho,\nu}\{T_\nu^{-1}(u),T_\nu^{-1}(v)\} \\
&= \int_{-\infty}^{T_\nu^{-1}(u)}\int_{-\infty}^{T_\nu^{-1}(v)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi\nu\sqrt{1-\rho^2}} \qquad (11) \\
&\left(1+\frac{s^2-2st\rho+t^2}{\nu}\right)^{-\frac{\nu+2}{2}} dsdt, \qquad (12)
\end{aligned}
$$

where $\boldsymbol{T}_{2,\rho,\nu}$ denotes the cdf of bivariate t distribution with scale parameter ρ *[ρ specifies the dependence, so why is it a scale parameter? Are you sure?]* and degrees of freedom parameter ν, and where $T_\nu^{-1}(\cdot)$ is the quantile function of a standard t distribution with degree of freedom ν. Contrary to the Gaussian copula, the t-copula has a non-zero tail dependence coefficient, which makes it more appropriate for dependence modelling in finance. The Gaussian copula arises as $\nu \to \infty$.

[Please make sure to use equation numbers only if the formulas are referenced.] The copula density

is

$$\boldsymbol{c}(u,v) = \frac{\boldsymbol{t}_{2,\rho,\nu}\{T_\nu^{-1}(u), T_\nu^{-1}(v)\}}{t_\nu\{T_\nu^{-1}(u)\} \cdot t_\nu\{T_\nu^{-1}(v)\}}, \tag{13}$$

where $\boldsymbol{t}_{2,\rho,\nu}$ is the pdf of $\boldsymbol{T}_{2,\rho,\nu}$ and t_ν the density of standard t distribution.

Like all the other elliptical copulae, the t-copula's Kendall's τ is identical to that of the Gaussian copula (see Demarta and McNeil, 2005, and references therein).

[until here]

3.1.2 Archimedean Copulae

The family of Archimedean copulae forms a large class of copulae with many convenient features. Contrary to elliptical copulas, which are derived from elliptical distributions. Archimedean copulas are determined via a simple parametric form of the dependence structure. A promiment feature is the ability to model asymmetric dependence structures. In general, they take a form

$$\boldsymbol{C}(u,v) = \psi^{-1}\{\psi(u), \psi(v)\}, \tag{14}$$

where $\psi : [0,1] \to [0,\infty)$ is a continuous, strictly decreasing and convex function such that $\psi(1) = 0$ for any permissible dependence parameter θ. ψ is also called generator. ψ^{-1} is the inverse of the generator.

[Remove the Frank copula? Or are we still using it? Then start with Clayton and Gumbel.]

The Frank copula (B3 in Joe (1997)) is a radial symmetric copula and cannot produce any tail dependence. It takes the form

$$\boldsymbol{C}_\theta(u,v) = \frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right\} \tag{15}$$

where $\theta \in [0,\infty]$ is the dependency parameter. $\boldsymbol{C}_{-\infty} = \boldsymbol{M}$, $\boldsymbol{C}_1 = \boldsymbol{\Pi}$, and $\boldsymbol{C}_\infty = \boldsymbol{W}$.

The Copula density is

$$\boldsymbol{c}_\theta(u,v) = \frac{\theta e^{\theta(u+v)(e^\theta - 1)}}{\left\{e^\theta - e^{\theta u} - e^{\theta v} + e^{\theta(u+v)}\right\}^2} \tag{16}$$

Frank copula has Kendall's τ and Spearman's ρ as follow:

$$\tau_K(\theta) = 1 - 4\frac{D_1\{-\log(\theta)\}}{\log(\theta)}, \tag{17}$$

and

$$\rho_S(\theta) = 1 - 12\frac{D_2\{-\log(\theta)\} - D_1\{\log(\theta)\}}{\log(\theta)}, \tag{18}$$

where D_1 and D_2 are the Debye function of order 1 and 2. Debye function is $D_n = \frac{n}{x^n}\int_0^x \frac{t^n}{e^t-1}dt$.

The Gumbel copula (B6 in Joe (1997)) has upper tail dependence with the dependence parameter

$\lambda^U = 2 - 2^{\frac{1}{\theta}}$ and displays no lower tail dependence.

$$\boldsymbol{C}_\theta(u, v) = \exp -\{(-\log(u))^\theta + (-\log(v))^\theta\}^{\frac{1}{\theta}},$$

where $\theta \in [1, \infty)$ is the dependence parameter.

While the Gumbel copula cannot model perfect counter-dependence (Nelsen, 2002), $\boldsymbol{C}_1 = \boldsymbol{\Pi}$ models the independence, and $\lim_\theta^\infty \boldsymbol{C}_\theta = \boldsymbol{W}$ models the perfect dependence. The copula density takes the form

$$\tau_K(\theta) = \frac{\theta - 1}{\theta}.$$

The Clayton copula, by contrast to Gumbel copula, generates lower tail dependence of the form $\lambda^L = 2^{-\frac{1}{\theta}}$, but cannot generate upper tail dependence. The Clayton copula takes the form

$$\boldsymbol{C}_\theta(u, v) = \left\{\max(u^{-\theta} + v^{-\theta} - 1, 0)\right\}^{-\frac{1}{\theta}},$$

where $\theta \in (-\infty, \infty)$ is the dependence parameter. Moreover, $\lim_\theta^{-\infty} \boldsymbol{C}_\theta = \boldsymbol{M}$, $\boldsymbol{C}_0 = \boldsymbol{\Pi}$, and $\lim_\theta^\infty \boldsymbol{C}_\theta = \boldsymbol{W}$. Kendall's τ of the Clayton copula is given by

$$\tau_K(\theta) = \frac{\theta}{\theta + 2}. \tag{19}$$

3.1.3 Mixture Copula

The mixture copula is a linear combination of copulae. For a 2-dimensional random variable $\boldsymbol{X} = (X_1, X_2)^\top$, its distribution can be written as linear combination of K copulae

$$\mathbf{P}(X_1 \leq x_1, X_2 \leq x_2) = \sum_{k=1}^K p^{(k)} \cdot \boldsymbol{C}^{(k)}\{F_{X_1}^{(k)}(x_1; \boldsymbol{\gamma}_1^{(k)}), F_{X_2}^{(k)}(x_2; \boldsymbol{\gamma}_2^{(k)}); \boldsymbol{\theta}^{\boldsymbol{(k)}}\} \tag{20}$$

where $p^{(k)} \in [0, 1]$ is the weight of each component, $\boldsymbol{\gamma}^{(k)}$ are the parameters (was: is the parameter) of the marginal distribution in the k-th component, and $\boldsymbol{\theta}^{\boldsymbol{(k)}}$ are the (was: is the) dependence parameters of the copula of the k-th component. The weights add up to one $\sum_{k=1}^K p^{(k)} = 1$.

We deploy a simplified version of the above representation by assuming the margins of $\boldsymbol{X}$ are not a mixture. *[Is this a precise formulation? The margins are not mixtures anyway, just specified for each copula component. Perhaps write: ... by specifying the same margins for each copula component.]*

[Check notation of quantile function throughout. I think we should use $F^{(-1)}$ instead of F^{-1}, as the latter can be mistaken for $1/F$].

By Sklar's theorem one may write *[Only for the special case where the margins are fixed, right? Mention this.]*

$$\boldsymbol{C}(u, v) = \sum_{k=1}^K p^{(k)} \cdot \boldsymbol{C}^{(k)}\{F_{X_1}^{-1}(u), F_{X_2}^{-1}(v); \boldsymbol{\theta}^{\boldsymbol{(k)}}\}.$$

The copula density is again a linear combination of copula densities

$$\boldsymbol{c}(u, v) = \sum_{k=1}^K p^{(k)} \cdot \boldsymbol{c}^{(k)}\{F_{X_1}^{-1}(u), F_{X_2}^{-1}(v); \boldsymbol{\theta}^{\boldsymbol{(k)}}\}. \tag{21}$$

While Kendall's τ of mixture copula is not known in closed form, Spearman's ρ is specified by the following statement.

Proposition 3 *In the setting of* (20) [Please check if this is correct! Also, please check if the copula must be continuous.], *let* $\rho_S^{(k)}$ *be Spearman's* ρ *of the* k*-th component (delete: and* $\sum_{k=1}^{K} p^{(k)} = 1$ *holds,). Spearman's* ρ *of the mixture copula is given by*

$$\rho_S = \sum_{k=1}^{K} p^{(k)} \cdot \rho_S^{(k)} \tag{22}$$

Proof. Since Spearman's ρ is defined as (Nelsen, 1999)

$$\rho_S = 12 \int_{\mathbb{I}^2} \boldsymbol{C}(s,t) ds dt - 3,$$

Spearman's ρ of the the mixture copula is given by summation of the components

$$\rho_S = 12 \int_{\mathbb{I}^2} \sum_{k=1}^{K} p^{(k)} \cdot \boldsymbol{C}^{(k)}(s,t) ds dt - 3. \tag{23}$$

■

[If the Fr'echet class is not used below, then I suggest to remove the example, and replace it by one sentence wiht a reference, i.e.: An example of a mixture copula is the Fr'echet class of copulas, which are given by convex combinations of $\boldsymbol{W}$*,* $\boldsymbol{\Pi}$*, and* $\boldsymbol{M}$ *(Nelsen, 1999).]*

Example 4 *The Fréchet class can be seen as an example of mixture copula. It is a convex combinations of* $\boldsymbol{W}$*,* $\boldsymbol{\Pi}$*, and* $\boldsymbol{M}$ *(Nelsen, 1999)*

$$\boldsymbol{C}_{\alpha,\beta}(u,v) = \alpha \boldsymbol{M}(u,v) + (1-\alpha-\beta)\boldsymbol{\Pi}(u,v) + \beta \boldsymbol{W}(u,v), \tag{24}$$

where α *and* β *are the dependence parameters, with* $\alpha, \beta \geq 0$ *and* $\alpha + \beta \leq 1$*. Its Kendall's* τ *and Spearman's* ρ *are*

$$\tau_K(\alpha,\beta) = \frac{(\alpha-\beta)(\alpha+\beta+2)}{3} \tag{25}$$

, and

$$\rho_S(\alpha,\beta) = \alpha - \beta \tag{26}$$

We use a mixture of Gaussian and independent copulas in our analysis, i.e.,

$$\boldsymbol{C}(u,v) = p \cdot \boldsymbol{C}^{\text{Gaussian}}(u,v) + (1-p)(uv),$$

with corresponding density is

$$\boldsymbol{c}(u,v) = p \cdot \boldsymbol{c}^{\text{Gaussian}}(u,v) + (1-p).$$

This mixture models the amount of "random noise" that appears in the dependence structure. In the hedging exercise, (delete: the structure of) the "random noise" adds an unhedgable component to the two-asset portfolio, whose weight $(1-p)$ is calibrated from market data (was: is not of our concern nor we can hedge the noise by a two-asset portfolio.) (delete: However, the proportion of the "random noise" does affect the distribution of r^h, so as the optimal hedging ratio h^*.) *[I think this can* *be deleted, but maybe not?]* One can consider the mixture copula as a handy tool for stress testing. Similar to this Gaussian mix Independent copula, t copula is also a two parameter copula allow us to model the noise, but its interpretation of parameters is not as intuitive as that of a mixture. The mixing variable p is the proportion of a manageable (hedgable) Gaussian copula, while the remaining proportion $1-p$ cannot be managed. *[Not sure I understand the comparison with the t copula. I think you might be thinking of the case where the scaling variable of the t-copula is large and the correlation is moderate, which produces some observations along the negative diagonal. However, this needs to be carefully explained – or left out.]*

3.1.4 NIG factor copula

The *normal inverse Gaussian (NIG)* distribution, introduced by (?), has density function

$$g(x;\alpha,\beta,\mu,\delta)=\frac{\alpha}{\pi}\mathbf{e}^{\delta\sqrt{\alpha^2-\beta^2}-\beta\mu}\frac{1}{q((x-\mu)/\delta)}K_1\left[\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right]\mathbf{e}^{\beta x},\quad x>0,$$

where $q(x)=\sqrt{1+x^2}$ and where K_1 is the modified Bessel function of third order and index 1. The parameters satisfy $0\leq|\beta|\leq\alpha$, $\mu\in\mathbb{R}$ and $\delta>0$. The parameters have the following interpretation: μ and δ are location and scale parameters, respectively, α determines the heaviness of the tails and β determines the degree of asymmetry. If $\beta=0$, then the distribution is symmetric around μ.

The moment-generating function of the NIG distribution is given by

$$M(u;\alpha,\beta,\mu,\delta)=\exp\left(\delta\left(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+u)^2}\right)+\mu u\right).$$

As a direct consequence, moments are easily calculated with the expectation and variance of the NIG distribution being

$$\mathbb{E}X=\mu+\frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}\tag{27}$$

$$\mathrm{Var}(X)=\frac{\alpha^2\delta}{(\alpha^2-\beta^2)^{3/2}}.\tag{28}$$

The $\mathrm{NIG}(\alpha,\beta,\mu\,\delta)$ distribution belongs to the class of so-called *normal variance-mean mixture*, (see Section 3.2 of (McNeil et al., 2005)): X follows an $\mathrm{NIG}(\alpha,\beta,\mu,\delta)$ distribution if X conditional on W follows a normal distribution with mean $\mu+\beta W$ and variance W, i.e.,

$$X|W\overset{\mathcal{L}}{\sim}\mathrm{N}(\mu+\beta W,W),$$

where W follows an *inverse Gaussian distribution*, denoted by $\mathrm{IG}(\delta,\sqrt{\alpha^2-\beta^2})$.

It is easily seen from the moment-generating function that the NIG distribution is preserved under linear combinations, provided the variables share the parameters α and β. For this and other reasons, the NIG distribution is popular in many areas of financial modelling; for example, it gives rise to the normal inverse Gaussian Lévy process, which may be represented as a Brownian motion with a time change.

In the setting here, we consider the *NIG factor copula*. This is not directly derived from the multivariate NIG distribution, but determined through a factor structure instead. The factor structure, which was applied e.g. in (?) for calibrating CDO's, gives additionaly flexibility as it does not force the components to have a mixing variable W. The following proposition introduces the NIG factor model and some of its properties.

Proposition 5 *Let $Z \sim NIG(\alpha, \beta, \mu, \delta)$ and $Z_i \sim NIG(\alpha, \beta, \mu_i, \delta_i)$, $i = 1, \ldots, n$ be independent NIG-distributed random variables. Then:*

1. $X_i = Z + Z_i \sim NIG(\alpha, \beta, \mu + \mu_i, \delta + \delta_i)$,
2. *and*

$$\begin{aligned} Cov(X_i, X_j) &= Var(Z), \\ Corr(X_i, X_j) &= \frac{\delta}{\sqrt{(\delta + \delta_i)(\delta + \delta_j)}}. \end{aligned} \tag{29}$$

Proof.

1. This follows directly from the moment-generating function.
2. For the covariance,

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \mathbb{E}[(Z + Z_i)(Z + Z_j)] - \mathbb{E}[Z + Z_i]\mathbb{E}[Z + Z_j] \\ &= \mathbb{E}[Z^2] - (\mathbb{E}Z)^2. \end{aligned}$$

The correlation is determined directly from (28).

■

The NIG factor copula is obtained by transforming the margins to uniforms (see Sklar's Theorem), giving (e.g. (**?**)):

$$C_{r^S, r^F}(F_{r^S}(r^S), F_{r^F}(r^F)) = \int_{\mathbb{R}} F_{Z_1}(F_{X_1}^{-1} \circ F_{r^S}(r^S) - z) \cdot F_{Z_2}(F_{X_2}^{-1} \circ F_{r^F}(r^F) - z) \cdot f_Z(z) dz$$

If the margins are continuous, then Spearman's rho of NIG factor copula is

$$\rho_S = 12 \int\int\int_{\mathbb{R}^3} F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot f_{Z_1}(x_1 - z) \cdot f_{Z_2}(x_2 - z) \cdot f_Z(z) dx_1 dx_2 dz - \frac{1}{48}.$$

3.2 Other Copula

[Why is there a separate subsection instead of 3.1.4?]

[Also, the reason to include the Packett copula needs to be made more clear; maybe with some evidence of what we see in the data? Or with explaining that other copulas do not have the property, and what it means that they do not have the property.]

The Plackett copula has an expression

$$\boldsymbol{C}_\theta(u, v) = \frac{1 + (\theta - 1)(u + v)}{2(\theta - 1)} - \frac{\sqrt{\{1 + (\theta - 1)(u + v)\}^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)} \tag{30}$$

$$\rho_S(\theta) = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \log \theta}{(\theta - 1)^2} \tag{31}$$

We include Placket copula in our analysis as it possesses a special property, the cross-product ratio is equal to the dependence parameter

$$\begin{aligned}
&\frac{\mathbf{P}(U \le u, V \le v) \cdot \mathbf{P}(U > u, V > v)}{\mathbf{P}(U \le u, V > v) \cdot \mathbf{P}(U > u, V \le v)} \\
&= \frac{\boldsymbol{C}_\theta(u,v)\{1-u-v+\boldsymbol{C}_\theta(u,v)\}}{\{u-\boldsymbol{C}_\theta(u,v)\}\{v-\boldsymbol{C}_\theta(u,v)} \\
&= \theta.
\end{aligned} \tag{32}$$

That is, the dependence parameter is equal to the ratio between number of concordence pairs and number of discordence pairs of a bivariate random variable.

3.3 Estimation

[Add a brief intro of which calibration methods are used and why. Also check if the title of the section should be "Calibration" instead of "Estimation".]

3.3.1 Simulated Method of Moments

Do we really use *simulated* method of moments throughout? I suggest to introduce this as *method of moments*.

To calibrate the various copulas, we use the *method of moments* calibration method (was: This method is) suggested by (Oh and Patton, 2013). It is targeted at copula-invariant properties such as Spearman's ρ, Kendall's τ and so-called *quantile dependence* measures, denoted by λ_q for quantile level q. (delete: In our setting, rank correlation e.g. Spearman's ρ or Kendall's τ, and quantile dependence measures at different levels λ_q are calibrated against their empirical counterparts.)

Spearman's rho, Kendall's tau, and quantile dependence of the (delete: a pair (X, Y) with) copula C are defined as *[Suggest to use $\mathbb{E}$ for expectation.]*

$$\rho_S = 12 \int\int_{I^2} C_{\boldsymbol{\theta}}(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3 \tag{33}$$

$$\tau_K = 4\,\mathsf{E}[C_{\boldsymbol{\theta}}\{F_X(x), F_Y(y)\}] - 1, \tag{34}$$

$$\lambda_q = \begin{cases} \mathbf{P}(F_X(X) \le q | F_Y(Y) \le q) = \dfrac{C_{\boldsymbol{\theta}}(q,q)}{q}, & \text{if } q \in (0, 0.5], \\ \mathbf{P}(F_X(X) > q | F_Y(Y) > q) = \dfrac{1-2q+C_{\boldsymbol{\theta}}(q,q)}{1-q}, & \text{if } q \in (0.5, 1). \end{cases} \tag{35}$$

The empirical counterparts are

$$
\hat{\rho}_S = \frac{12}{n}\sum_{k=1}^{n}\hat{F}_X(x_k)\hat{F}_Y(y_k) - 3,
$$

$$
\hat{\tau}_K = \frac{4}{n}\sum_{k=1}^{n}\hat{C}\{\hat{F}_X(x_i), \hat{F}_X(y_i)\} - 1,
$$

$$
\hat{\lambda}_q = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\dfrac{\mathbf{1}_{\{\hat{F}_X(x_k)\leq q, \hat{F}_Y(y_k)\leq q\}}}{q}, & \text{if } q \in (0, 0.5], \\ \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\dfrac{\mathbf{1}_{\{\hat{F}_X(x_k)> q, \hat{F}_Y(y_k)> q\}}}{1-q}, & \text{if } q \in (0.5, 1), \end{cases}
$$

where $\hat{F}(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{x_i \leq x\}}$ and $\hat{C}(u,v) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{u_i \leq u, v_i \leq v\}}$.

Denote by $\boldsymbol{m}(\boldsymbol{\theta})$ the m-dimensional vector of dependence measures according the dependence parameters $\boldsymbol{\theta}$,and let $\hat{\boldsymbol{m}}$ be the corresponding empirical counterpart. The difference between dependence measures and their counterpart is denoted by

$$
\boldsymbol{g}(\boldsymbol{\theta}) = \hat{\boldsymbol{m}} - \boldsymbol{m}(\boldsymbol{\theta}).
$$

The SMM *[MM? Also introduce the abbreviation]* estimator is

$$
\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\, \boldsymbol{g}(\boldsymbol{\theta})^\top \hat{\boldsymbol{W}} \boldsymbol{g}(\boldsymbol{\theta}),
$$

where $\hat{W}$ is some positive definite weight matrix. Here, we use $\boldsymbol{m}(\boldsymbol{\theta}) = (\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95})^\top$ for calibration of *[update this:]* Bitcoin price and CME Bitcoin future. *[How is $\hat{W}$ defined in our setting? Why the hat?]*

3.4 Maximum Likelihood Estimation

By the Hoeffding-Sklar theorem, the joint density of a d-dimensional random variable $\boldsymbol{X}$ with sample size n can be written as

$$
\boldsymbol{f_X}(x_1, ..., x_d) = \boldsymbol{c}\{F_{X_1}(x_1), ..., F_{X_d}(x_d)\}\prod_{j=1}^{d} f_{X_i}(x_i).
$$

We follow the treatment of MLE documented in section 10.1 of Joe (1997), namely the *inference functions for margins (IFM) method*. The log-likelihood $\sum_{i=1}^{n} \boldsymbol{f_X}(X_{i,1}, ..., X_{i,d})$ can be decomposed into a dependence part and a marginal part,

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\boldsymbol{c}\{F_{X_1}(x_{i,1};\boldsymbol{\delta}_1), ..., F_{X_d}(x_{i,d};\boldsymbol{\delta}_d);\boldsymbol{\gamma}\} + \sum_{i=1}^{n}\sum_{j=1}^{d} f_{X_j}(x_{i,j};\boldsymbol{\delta}_j) \qquad (36)\\
&= L_C(\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_d, \boldsymbol{\gamma}) + \sum_{j=1}^{d} L_j(\boldsymbol{\delta}_j) \qquad (37)
\end{aligned}
$$

where $\boldsymbol{\delta}_j$ are the parameters of the j-th margin, $\boldsymbol{\gamma}$ is the parameter of the parametric copula, and $\boldsymbol{\theta} = (\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_d, \boldsymbol{\gamma})$. Instead of searching the $\boldsymbol{\theta}$ in a high dimensional space, Joe (1997) suggests to search for $\hat{\boldsymbol{\delta}}_1, ..., \hat{\boldsymbol{\delta}}_d$ that maximize $L_1(\boldsymbol{\delta}_1), ..., L_d(\boldsymbol{\delta}_d)$, then search for $\hat{\boldsymbol{\gamma}}$ that maximize $L_C(\hat{\boldsymbol{\delta}}_1, ..., \hat{\boldsymbol{\delta}}_d, \boldsymbol{\gamma})$.

[I suggest to delete the next part, as the regularity conditions are unclear, and it is just a first-

order condition, which is a-priori not clear to hold in a two-step procedure.] That is, under regularity conditions, $(\hat{\boldsymbol{\delta}}_1, ..., \hat{\boldsymbol{\delta}}_d, \hat{\boldsymbol{\gamma}})$ is the solution of

$$\left(\frac{\partial L_1}{\partial \boldsymbol{\delta}_1}, ..., \frac{\partial L_d}{\partial \boldsymbol{\delta}_d}, \frac{\partial L_C}{\partial \boldsymbol{\gamma}}\right) = \mathbf{0}. \tag{38}$$

However, the IFM requires making assumption on the distribution of the margins.*[delete until here.]*

We follow Genest et al. (1995) who suggest (was: Genest et al. (1995) suggests) to replace the estimation of marginals parameters estimation by non-parametic estimation. Given non-parametric estimator $\hat{F}_i$ of the margins F_i, the estimator of the dependence parameters $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}} \sum_{i=1}^{n} \boldsymbol{c}\{\hat{F}_{X_1}(x_{i,1}), ..., \hat{F}_{X_d}(x_{i,d}); \boldsymbol{\gamma}\}.$$

3.4.1 Comparison

[Make this subsection; don't start subsubsections that do not have a number 2.]

[If the MLE method is not used at all, then it is OK to just very briefly mention it and highlight the comparison. No extra subsection needed.]

Both the simulated method of moments and the maximum likelihood estimation are unbiased. The question though which procedure is more suitable for hedging.

Figure 1 shows the empirical quantile dependence of Bitcoin and CME future and the copula implied quantile dependence of the MLE and MM calibration procedures. Although the MLE is a better fit to a range of quantile dependence in the middle, it fails to address the situation in the tails. Our data empirically has weaker quantile dependence in the ends *[weaker than what?]*, and those points generate PnL to the hedged portfolio. *[If you absolutetly want to address is this way, then it's better to write "we find that...".]* MM is preferred visually as it produces a better fit to the dependence structure in the two extremes. *[Visually preferred is very subjective; suggest to delete. You could argue that the MM method is targeted at copula properties, and allow to focus on tail behaviour, contrary to MLE.]* Therefore, we deploy the method of moments throughout the analysis. *[Then cut MLE part above short.]* We choose the 5^{th}-, 10^{th}-, 90^{th}-, 95^{th}-quantile, and Spearman's ρ as the moments.

3.5 Copula Selection

[I suggest: 3.3 Calibration and copula selection; 3.3.1 Method of moments; 3.3.2 Comparison with MLE; 3.3.3 Copula selection]

[Please avoid the word dependency. In probability theory, it is dependence.]

The dependence structure of price data changes across time, in which both the dependency parameters and the type of dependence, dependencies between cryptos and the BTCF are no exception. *[?]* For this reason, we allow for a flexible choice of the best-fitting copula, by re-calibrating periodically and re-evaluating performance of the various copulas. (was: In this hedging exercise, we find a best fitting copula to model the dependency for every set of training data.) We select the best-fitting copula, characterised by the lowest *Akaike Information Criterion (AIC)*,

$$\text{AIC} = 2k - 2\log(L),$$

where k is the number of estimated parameteres and L is the likelihood (**?**).

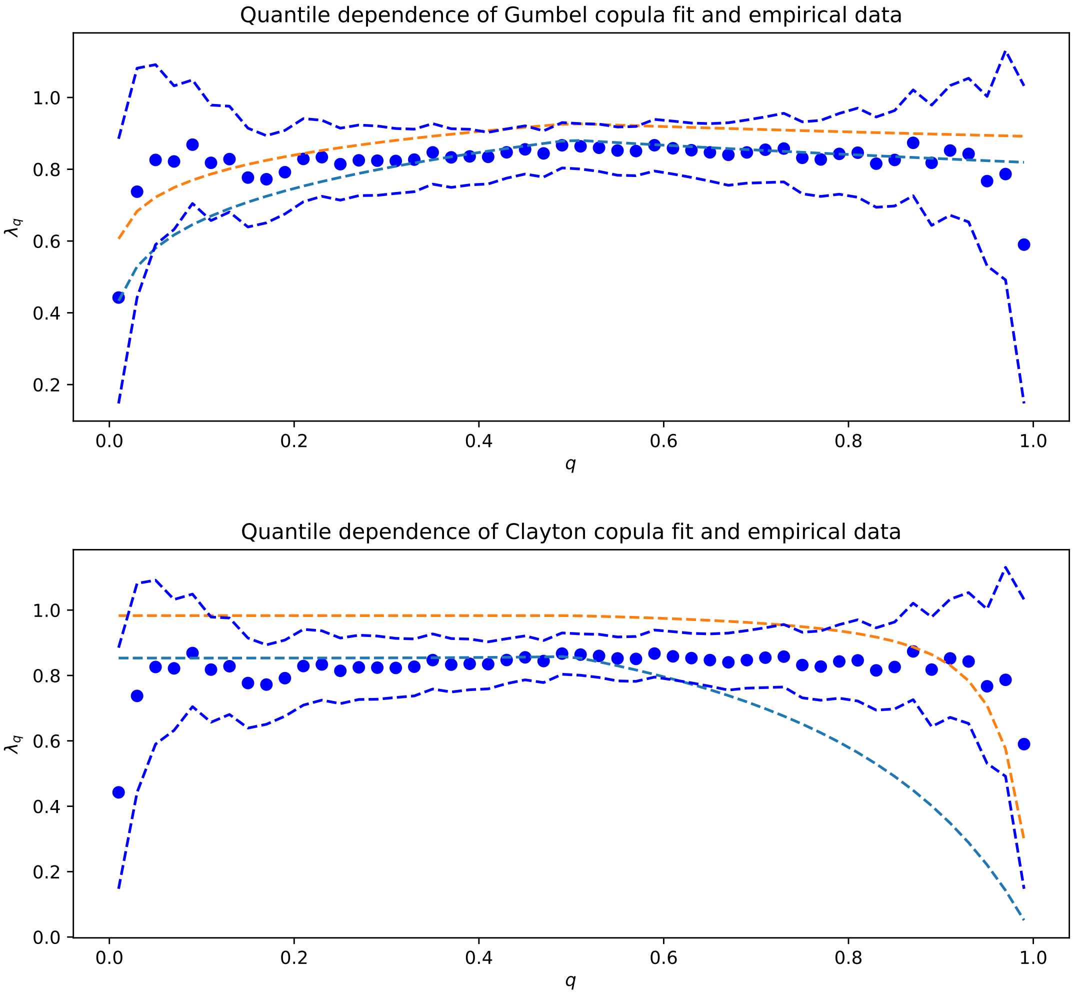

Figure 1: Quantile dependences of Gumbel, and Clayton Copula. The blue circle dots are the quantile dependence estimates of Bitcoin and CME future, blue dotted lines are the estimates' 90% confidence interval. Orange dotted line is the copula implied quantile dependence by MM estimation. Light blue dotted line is the copula implied quantile dependence by MLE estimation.

Other model selection criteria, such as the (was: Notice that there are other model selection procedure and criteria, e.g.) TIC or likelihood ratio test could be used instead. For a survey of model selection and inference, see Anderson et al. (1998). Among various copula selection procedures, AIC is a popular choice for its applicability, see e.g. Breymann et al. (2003) (was: for example Breymann et al. (2003) use the AIC to select best fitting copulae). In our case, the AICs are calculated only with dependency likelihood since the marginals are modelled via kernel density estimators. The selected copula will then be enter the calculation of the optimal hedge ratio.

3.6 Risk Measures

The optimal hedge ratio is determined for the following (was: We consider a) variety of risk measures: variance, Value-at-Risk (VaR), Expected Shortfall (ES), and Exponential Risk Measure (ERM). A summary of risk measures being used in portfolio selection problem can be found in Härdle et al. (2008). The risk measures are defined as follows. Let Z be a random variable with distribution function F_Z.

1. Variance: $\mathrm{Var}(Z) = \mathbb{E}[(Z - \mathbb{E}Z)^2]$.

2. VaR at confidence level α: $\text{VaR}_\alpha(Z) = -F_Z^{(-1)}(1-\alpha)$

3. ES at confidence level α: $\text{ES}(F_Z) = -\frac{1}{1-\alpha}\int_0^{1-\alpha} F_Z^{(-1)}(p)dp$

4. ERM with Arrow-Pratt coefficient of absolute risk aversion k:

$$\text{ERM}_k(F_Z) = \int_0^{1-\alpha} \phi(p) F_Z^{(-1)}(p)dp,$$

where ϕ is a weight function described in (39) below.

VaR, ES, and ERM fall into the class of spectral risk measures (SRM), which have the from (Acerbi, 2002)

$$\rho_\phi(r^h) = -\int_0^1 \phi(p) F_Z^{(-1)}(p)dp,$$

where p is the loss quantile and $\phi(p)$ is a user-defined weighting function defined on $[0,1]$. We consider only so-called admissible risk spectra $\phi(p)$, i.e., fulfilling

(i) ϕ is positive,

(ii) ϕ is decreasing,

(iii) and $\int \phi = 1$.

The VaR's $\phi(p)$ gives all its weight on the $1-\alpha$ quantile of Z and zero elsewhere, i.e. the weighting function is a Dirac delta function, and hence it violates the (ii) property of admissible risk spectra. The ES' $\phi(p)$ gives all tail quantiles the same weight of $\dfrac{1}{1-\alpha}$ and non-tail quantiles zero weight. The ERM assumes investors' risk preference are in the form of an exponential utility function $U(x) = 1 - e^{kx}$, so its corresponding risk spectrum is defined as *[Please double-check. All I could find was that the ERM is in the spirit of investors' risk preferences not that it matches investors preferences. Please also look in the notes.]*

$$\phi(p) = \frac{ke^{-k(1-p)}}{1-e^{-k}}, \tag{39}$$

where k is the Arrow-Pratt coefficient of absolute risk aversion. The parameter k has an economic interpretation as being the ratio between the second derivative and first derivative of investor's utility function on an risky asset,

$$k = -\frac{U''(x)}{U'(x)},$$

for x in all possible outcomes. In case of the exponential utility, k is the the constant absolute risk aversion (CARA).

[Also note that there is a one-to-one correspondence between coherent risk measures and ERM's with admissible risk spectra if I remember well. This is one of the main factors to use ERM's.]

4 Empirical Results

4.1 Data

In the empirical analysis, we consider the risk reduction capability of CME Bitcoin Futures (BTCF) on five cryptos, namely Bitcoin (BTC), Ethereum (ETH), Cardano (ADA), Litecoin (LTC) and Ripple (XRP), as well as five crypto indexes, namely BITX, BITW100, CRIX, BITW20, and BITW70.

ETH, ADA, LTC, and XRP are popular cryptos tradable in various exchanges and have large market capitalization. BITX, BITW100, and CRIX are market-cap weighted crypto indexes with BTC as constituent. BITX and BITW100 tracks the total return of the 10 and 100 cryptos with largest market-cap respectively. CRIX decides the number of constituents by AIC and tracks that number of cryptos with largest market-cap. In our case, the number of constituents in CRIX is 5. BITW20 is also a market-cap weighted crypto index but with 20 largest market-cap cryptos outside the constituents of BITX. BITW70 has the same construction as BITW20 but with 70 largest market-cap cryptos outside BITX and BITW20. Therefore, BTC is excluded as constituent in BITW20 and BITW70.

For each of the 10 hedging portfolios, a crypto or index is considered as the spot and held in a unit size long position, while the front BTCF is held in a short position with units corresponding to the OHR in order to reduce the risk of the spot. Except for the hedge of BTC, all hedging portfolios are considered cross-asset hedges.

We collect the spots' and BTCF's daily price at 15:00 US Central Time (CT). The reason of choosing this particular time is that the CME group determines the daily settlements for BTCF's based on the trading activities on CME Globex between 14:59 and 15:00 CT. This is also the reporting time of the daily closing price by Bloomberg. The crypto spot data is collected from the data provider called Tiingo (https://www.tiingo.com/). *[thanks somewhere.]* Tiingo aggregates crypto OHLC (open, high, low, and close) prices fed by APIs from various exchanges. It covers major exchanges, such as Binance, Gemini, Poloniex etc., so Tiingo's aggregated OHLC price is a good representation a tradable market price. For each crypto, we match the opening price at 15:00 CT from Tiingo with the daily BTCF closing price from Bloomberg. Since CRIX is not available at 15:00 CT, we recalculated an hourly CRIX using the monthly constituents weights and the hourly OHLC price data collected from Tiingo. BITX, BITW20, BITW70, and BITW100 are collected from the official website of their publisher Bitwise.com. The daily reporting time of the Bitwise indexes is 15:00 CT.

At the time of writing, the CRIX is undergoing the listing process on the S&P Dow Jones Indices, the official CRIX data will then be calculated with Lukka Prime Data and available to the public via S&P.

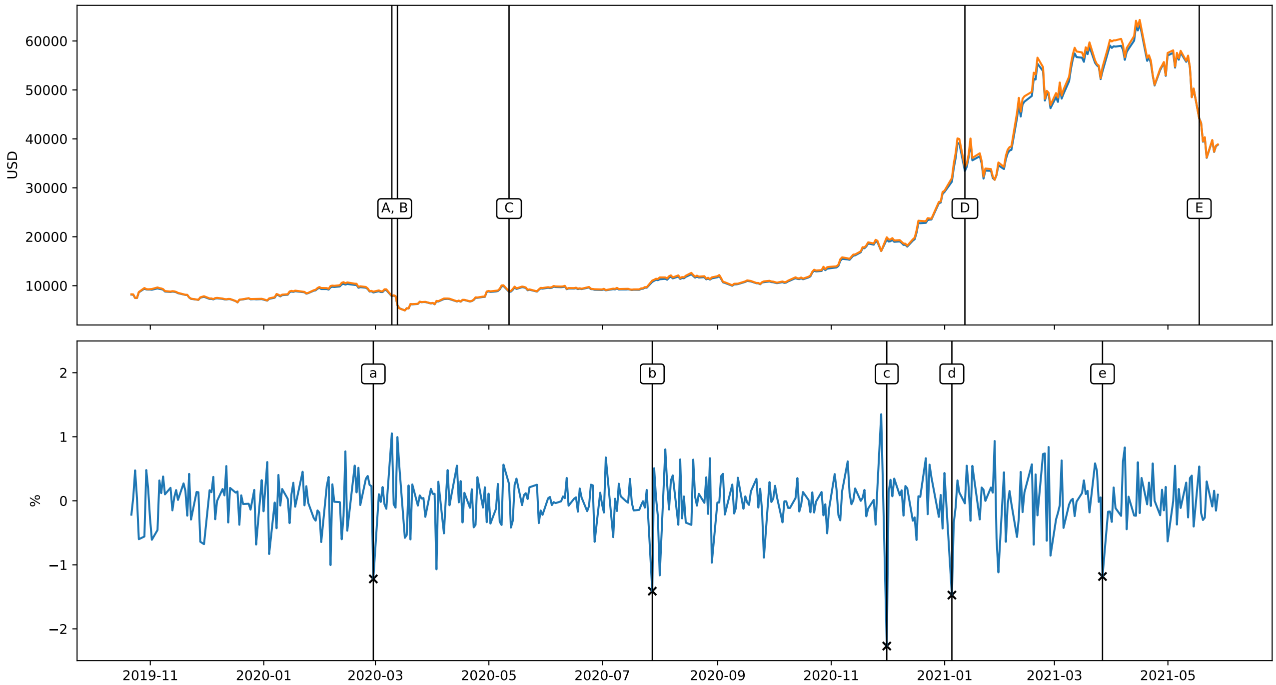

Figure 2: Out-of-sample BTC and BTCF price. The first panel presents the price of BTC in blue line and that of BTCF in orange line. The black vertical lines with capital letter labels indicate the five most negative daily return of BTC in the out-of-sample data. The second panel presents the difference between the daily log return of BTC and BTCF $r_s - r_f$. The black vertical lines with lowercase letter label indicate the five most negative difference. The crosses helps us better locate the level the difference.

4.2 An overview of the hedged portfolios without the copula selection step

Since different copulae may suggest different OHRs, we analyse the result of hedged portfolios without the copula selection step in order to get a better understanding of how a copula affect the hedged portfolio with various risk minimization objectives. To do so, we inspect the performance of copula in hedging by the mean square error and lowersemi variance. Mean square error is the distance between a perfect hedge and the hedged portfolio returns MSE $= \frac{1}{n}\sum_{i=1}^{n}(r_i^h)^2$. Lower semi variance is LSV $= \frac{1}{n}\sum_{i=1}^{n}\{\mathrm{E}(r^h) - r_i^h\}^2$. All results here are out-of-sample results obtained without the copula selection step in order to compare the performances across copulae.

Figure 5 and 6 are the mean square error and lower semivariance of BTC-BTCF, we can see the Frank copula is the worst performing copula: the resulting hedged portfolio returns is far away from a perfect hedge. In figure 7 and 8, the phenomena of Frank copula being inferior to its counterparts can be observed from the results of the CRIX, BITX, BITW100, and BITW20-BTCF portfolios. Interestingly, the spot in those portfolios usually have a strong dependency to the BTCF. In contrast, the inferiority of the Frank copula is less prominent in the BITW70, ADA, ETH, LTC and XRP-BTCF portfolios. We suspect that the Frank copula is not a choice to model assets with strong dependency.

We can also observe from figure 7 and 8 that Gumbel copula is not performing as good as other copula in the ETH, LTC, and XRP-BTCF portfolios. The reason is the Gumbel copula has only the upper tail dependence, while the ETH, LTC, and XRP exhibit lower tail dependency with BTCF. We will discuss this in the following section.

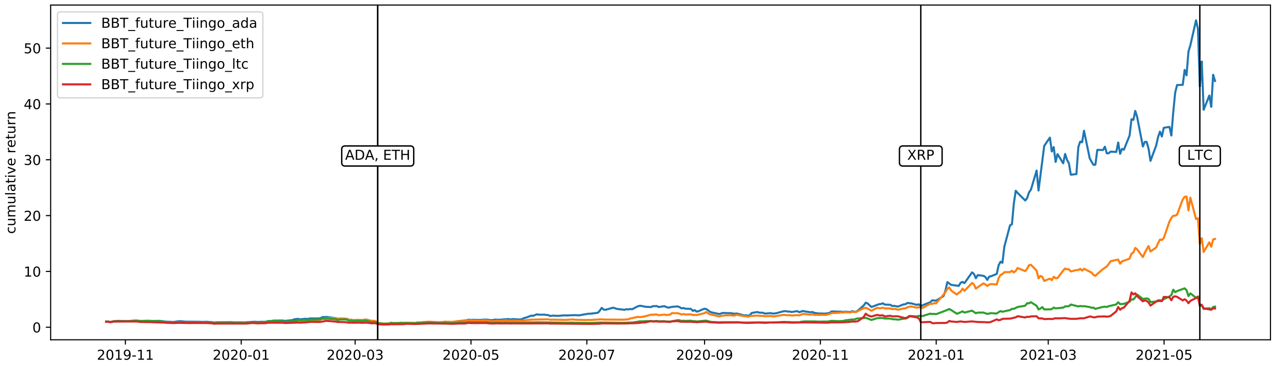

Figure 3: Out-of-sample cumulative return of individual cyptos. The black vertical lines indicate the most negative daily return of cryptos indicated by the labels.

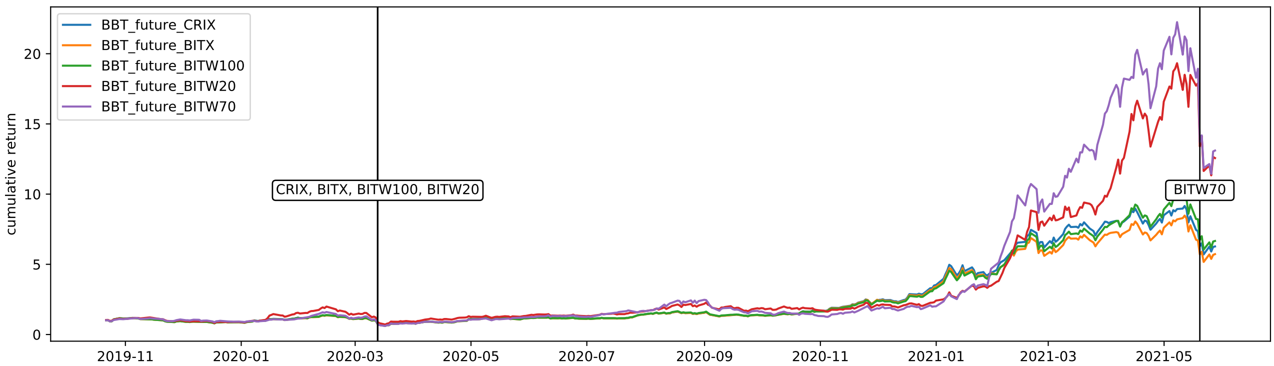

Figure 4: Out-of-sample cumulative return of crypto indices. The black vertical lines indicate the most negative daily return of indices indicated by the labels.

4.3 Copula Selection Results

Next, we inspect the copula selection result in this section. Although the copula selection is only an intermediate step to obtain the OHRs, the result of this step can help us better understand the dependence feature between BTCF and the assets we study in this work and give us valuable information to model the assets in the future. Decisions of the AIC procedure are summarised in table 1.

Overall, t-copula, rotated Gumbel (rotGumbel), and the NIG factor copula are the most frequently chosen copulae by the AIC procedure.

The t-copula is frequently chosen by AIC to model the dependency between the BTC and BTC involved indices, CRIX, BITX, BITW100, and the BTC future. BTC and BTC involved indices exhibit strong tail dependence (both upper and lower tail) with BTCF. We could interpret tail dependence much more of a tendency for one asset to be extreme when another is extreme and vice versa (McNeil et al., 2015). In fact, the t copula has been suggested in various empirical studies to model financial data, such as Zeevi and Mashal (2002) and Breymann et al. (2003). Those studies suggest t-copula is a better model over the Gaussian copula which financial data often seem to exhibit tail dependence.

On the other hand, the radially symmetric feature makes the t-copula not a good choice to model the other hedging pairs. Demarta and McNeil (2005) describe the symmetry feature "strong", because if $(U_1, ..., U_d)$ is a vector distributed in t-copula, then $(U_1, ..., U_d) \stackrel{\mathcal{L}}{=} (1-U_1, ..., 1-U_d)$. This symmetry can be justified in the dependence structure between a futures and its underlying by the theory of

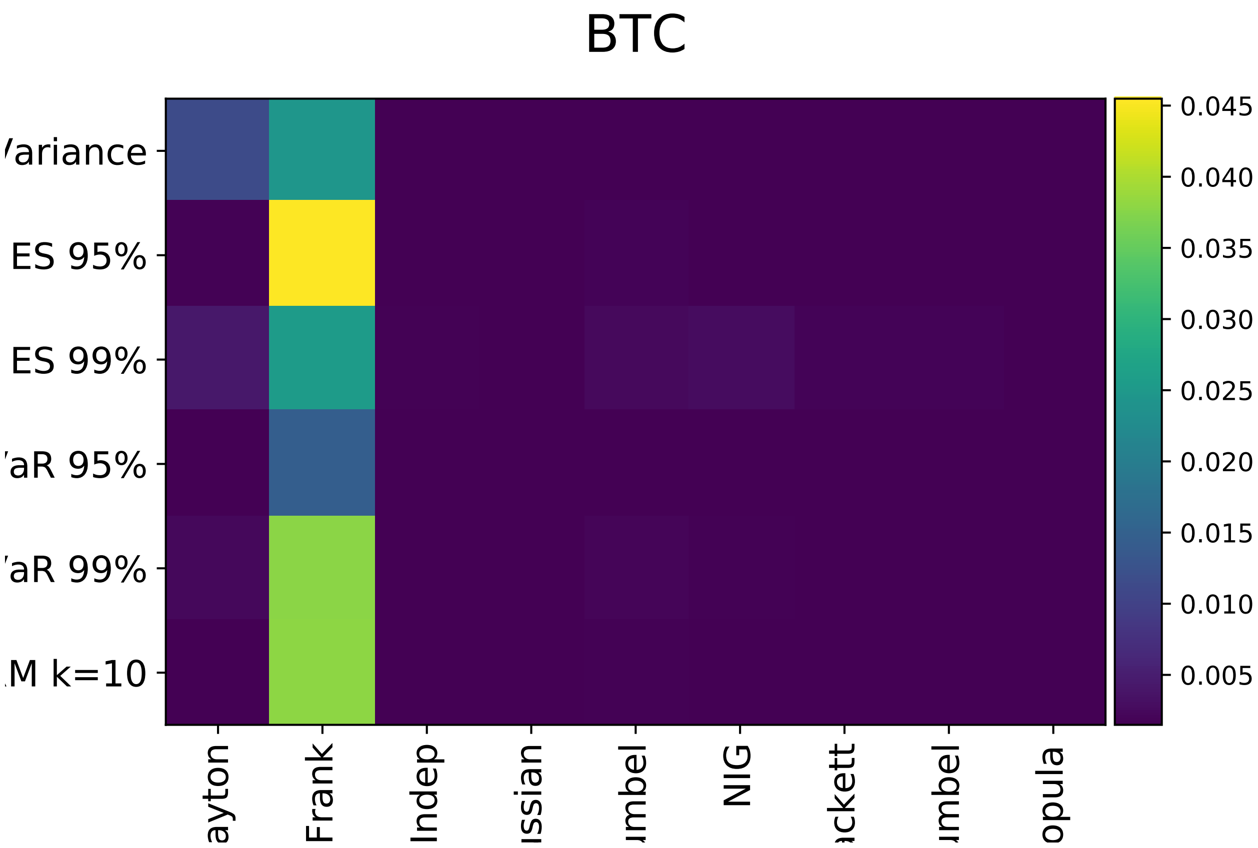

Figure 5: Mean square errors of BTC-BTCF portfolios constructed with different copula and risk minimization objectives. The Frank copula is inferior in the BTC-involved portfolios.

futures pricing, which suggests the price of a futures is a function of the underlying price (Hull, 2003). However, there is no such relationship between a futures and an asset which is not the underlying, and so the radial symmetry becomes a drawback to model other hedging pairs e.g. ETH and BITX70. Another drawback of the t-copula is the lack of flexibility to model off-diagonal region since Rho and nu jointly control the density of the off-diagonal region. This is why sometimes the Gaussian Mix Independence (GMI) better model the dependence.

Among the three popular copulae, rotGumbel copula shows its ability to model the dependency between ETH and BTCF, 94 out of 112 training sets are best fitted with the rotated Gumbel. rotGumbel also performs well when modelling dependency between XRP, BITW20, BITW70, and the BTCF. In particular, the whole time series of the two indices, BITW20 and BITW70, are best fitted solely with the rotated Gumbel copula. The frequently chosen rotated Gumbel indicates the styled fact of financial data: prices tends to drop together.

In fact, Clayton's AIC in many of the training sets is the second lowest, just higher than that of rotated Gumbel. This is because the Clayton copula has the same ability to model the lower quantile dependence. However, Clayton's radial like feature does not match the behaviour of the financial data.

It is worth to mention that although the NIG factor copula is penalised heavily due to its three parameters setup, it is frequently chosen to be the best copula to model the dependency between individual cryptos and the BTC future. An extreme case would be ADA, only NIG factor is chosen in our dataset. Another dependency structure being best described by the NIG factor copula is the pair

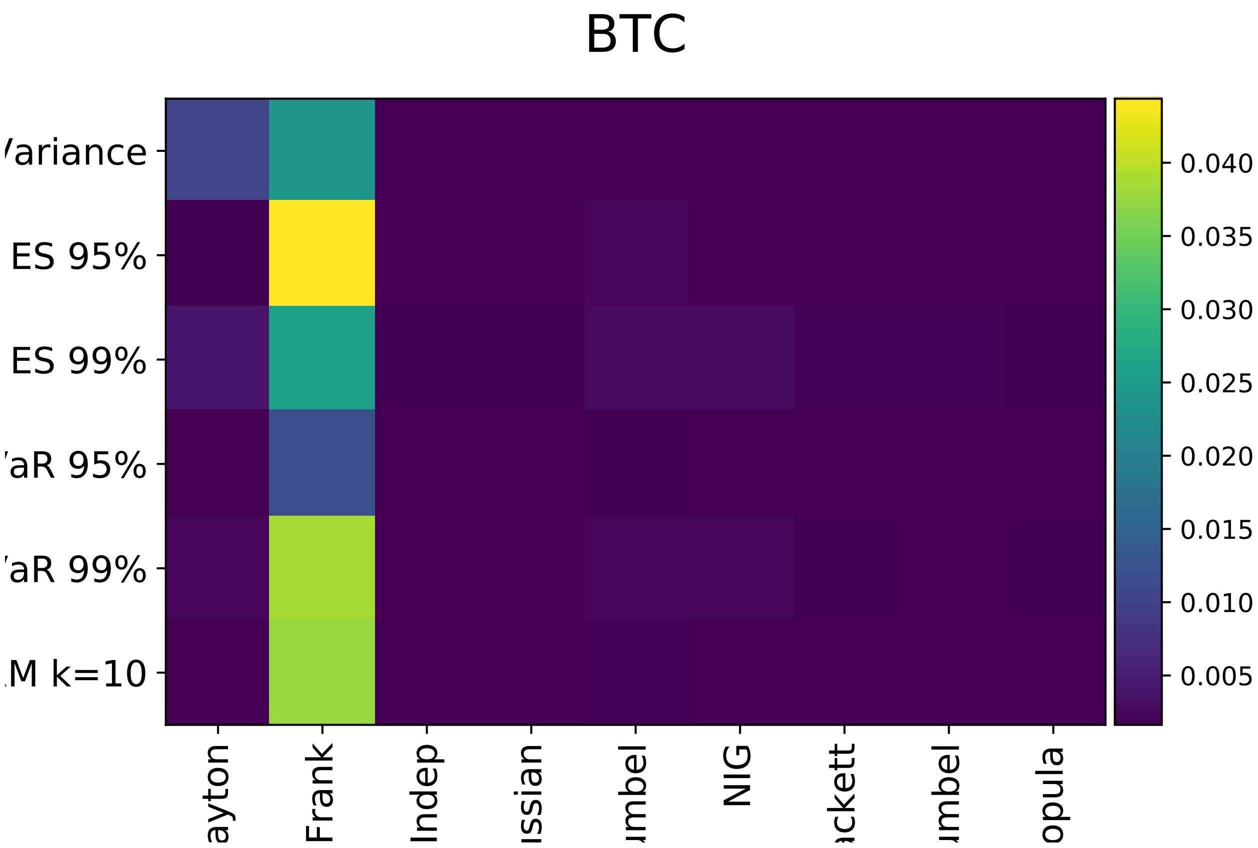

Figure 6: Lower semivariance of BTC-BTCF portfolios constructed with different copula and risk minimization objectives. The Frank copula is obviously inferior.

of LTC-BTCF. 64 out of 112 training sets are best fitted by the NIG factor copula. Indices like BITX and CRIX are sometimes best fitted with the NIG factor copula as well, accounting for modelling 12 and 27 training sets respectively. The popularity of the NIG factor reflects the ability of the copula to model very complex dependency structure. NIG factor copula is able to model the tail, radial asymmetry, and off diagonal behaviour.

Frank copula is generally not a good choice to model financial data just like what Barbi and Romagnoli (2014) has reported. Plackett is known for its dependence parameter can be written as the cross-product ratio (Joe, 1997). However, this feature does not bring the Plackett Copula advantage over other copulae to model the dependence structure between cryptos and BTCF.

4.4 Hedged portfolios with the copula selection step

We presents the TBC

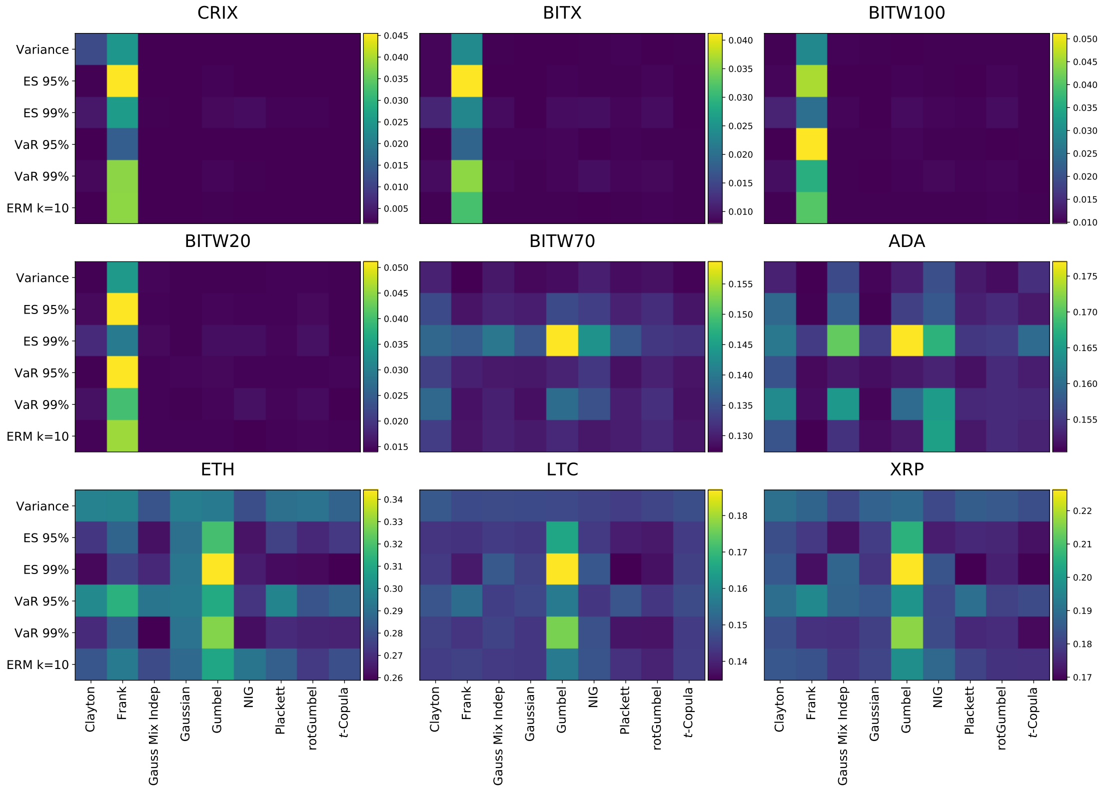

Figure 7: Mean square errors of portfolios constructed with different copula and risk minimization objectives.

Copula/Asset	t	Plackett	GMI	rotGumbel	NIG
Individual Cryptos					
BTC	73	4	2	1	31
ETH	3	6	8	94	1
ADA	0	0	0	0	112
LTC	13	0	3	32	64
XRP	0	31	3	78	0
Crypto Indices with BTC Constituent					
BITX	39	0	14	16	12
CRIX	47	0	11	3	27
BITW100	42	0	8	29	2
Crypto Indices without BTC Constituent					
BITW20	0	0	0	78	3
BITW70	0	0	0	80	1

Table 1: Copula Selection Results.

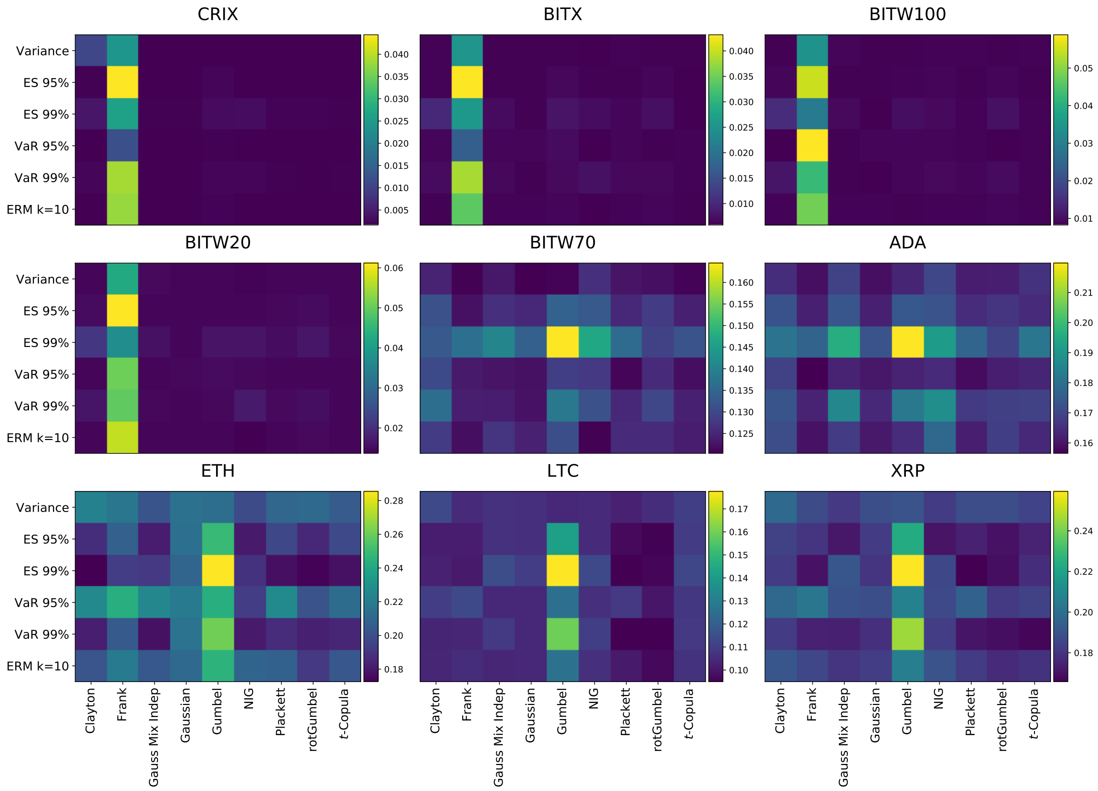

Figure 8: Lower semivariance of portfolios constructed with different copula and risk minimization objectives.

	Mean %	Std %	Skew	Kurt	MD %	MD date	ERM k=10
Individual Cryptos							
BTC	0.0223	0.3221	-1.0008	3.4153	-1.5242	2020-11-30	0.0057
ETH	0.3117	3.8679	1.0345	7.5751	-18.8729	2021-05-19	0.0491
ADA	0.5722	5.3590	1.4203	4.6970	-14.3885	2021-01-08	0.0700
LTC	-0.0512	3.8812	-0.2929	7.7022	-28.0879	2021-05-19	0.0616
XRP	0.0155	7.1579	1.1244	19.8583	-52.5689	2020-12-23	0.0787
Crypto Indices with BTC Constituent							
BITX	0.0590	1.0078	-0.4427	13.0839	-7.8581	2021-05-19	0.0127
CRIX	0.0840	0.9087	0.0488	14.5501	-7.0530	2021-05-19	0.0100
BITW100	0.0853	1.2032	-1.6522	20.5562	-11.1846	2021-05-19	0.0153
Crypto Indices without BTC Constituent							
BITW20	0.2564	3.6009	-0.3446	4.2152	-21.5920	2021-05-19	0.0503
BITW70	0.2818	3.9074	-0.6952	4.8745	-24.5250	2021-05-19	0.0557

Table 2: Summary Statistics of Hedged Portfolios that minimize ERM k=10.

	Mean %	Std %	Skew	Kurt	MD %	MD date	ES 5%
Individual Cryptos							
BTC	0.0204	0.3234	-1.0150	3.4423	-1.5629	2020-11-30	0.0101
ETH	0.3082	3.8890	1.0119	7.4077	-18.7819	2021-05-19	0.0782
ADA	0.5525	5.2673	1.2557	4.2423	-14.9647	2021-05-19	0.0984
LTC	-0.0808	3.9829	-0.4957	7.2302	-28.4608	2021-05-19	0.0962
XRP	0.0176	7.1533	1.1411	19.9176	-52.5698	2020-12-23	0.1354
Crypto Indices with BTC Constituent							
BITX	0.0591	1.0065	-0.3453	12.1335	-7.6211	2021-05-19	0.0215
CRIX	0.0777	0.9207	0.0164	13.5608	-6.9894	2021-05-19	0.0173
BITW100	0.0848	1.2125	-1.6397	19.7472	-11.1357	2021-05-19	0.0274
Crypto Indices without BTC Constituent							
BITW20	0.2608	3.6115	-0.3555	4.2016	-21.5430	2021-05-19	0.0804
BITW70	0.2785	3.9157	-0.6949	4.8047	-24.3474	2021-05-19	0.0908

Table 3: Summary Statistics of Hedged Portfolios that minimize ES 5%.

	Mean %	Std %	Skew	Kurt	MD %	MD date	ES 1%
Individual Cryptos							
BTC	0.014808	0.347559	-0.835432	3.305439	-1.622527	2020-11-30	0.023384
ETH	0.307977	3.895367	0.984009	7.494711	-18.762453	2021-05-19	0.129950
ADA	0.501591	5.404018	1.100767	3.960739	-15.448083	2021-05-19	0.146283
LTC	-0.102863	4.158086	-0.775700	7.437497	-29.172721	2021-05-19	0.164734
XRP	-0.019965	7.288724	1.112146	18.873229	-52.569956	2020-12-23	0.251630
Crypto Indices with BTC Constituent							
BITX	0.059801	1.031183	-0.440995	11.586333	-7.742446	2021-05-19	0.041121
CRIX	0.083524	0.946138	-0.036149	12.404693	-7.020300	2021-05-19	0.034979
BITW100	0.078145	1.263996	-1.964529	21.883557	-11.926340	2021-05-19	0.059302
Crypto Indices without BTC Constituent							
BITW20	0.253793	3.632266	-0.408622	4.446241	-21.986552	2021-05-19	0.128208
BITW70	0.266023	3.932037	-0.759783	5.004985	-24.476390	2021-05-19	0.153457

Table 4: Summary Statistics of Hedged Portfolios that minimize ES 1%.

	Mean %	Std %	Skew	Kurt	MD %	MD date	VaR 5%
Individual Cryptos							
BTC	0.0253	0.3294	-0.9725	3.4373	-1.5347	2020-11-30	0.0063
ETH	0.3084	3.8944	1.0243	7.4297	-19.1750	2021-05-19	0.0514
ADA	0.5726	5.2204	1.2981	4.2544	-14.6974	2021-05-19	0.0769
LTC	-0.0742	3.9145	-0.3836	7.5384	-28.3672	2021-05-19	0.0622
XRP	0.0208	7.1520	1.1269	19.8930	-52.5667	2020-12-23	0.0683
Crypto Indices with BTC Constituent							
BITX	0.0562	0.9930	-0.3117	12.4780	-7.5639	2021-05-19	0.0128
CRIX	0.0863	0.9151	0.0718	13.7915	-6.9744	2021-05-19	0.0092
BITW100	0.0846	1.1980	-1.6592	21.3725	-11.2582	2021-05-19	0.0164
Crypto Indices without BTC Constituent							
BITW20	0.2728	3.5940	-0.3721	4.4896	-22.0733	2021-05-19	0.0546
BITW70	0.2847	3.9133	-0.6580	4.7874	-24.6513	2021-05-19	0.0626

Table 5: Summary Statistics of Hedged Portfolios that minimize VaR 5%.

	Mean %	Std %	Skew	Kurt	MD %	MD date	VaR 5%
Individual Cryptos							
BTC	0.0176	0.3270	-1.0405	3.3742	-1.5689	2020-11-30	0.0134
ETH	0.2977	3.9132	0.9547	7.2414	-18.6061	2021-05-19	0.1026
ADA	0.5562	5.3466	1.1362	3.9334	-15.4795	2021-05-19	0.1106
LTC	-0.0852	4.1503	-0.7234	7.3208	-29.0915	2021-05-19	0.1030
XRP	0.0352	7.1658	1.1582	19.8506	-52.5727	2020-12-23	0.1387
Crypto Indices with BTC Constituent							
BITX	0.0593	1.0178	-0.5331	13.3100	-8.0299	2021-05-19	0.0247
CRIX	0.0738	0.9695	-0.4729	13.6500	-7.0185	2021-05-19	0.0245
BITW100	0.0823	1.2338	-1.9365	23.1938	-11.8752	2021-05-19	0.0347
Crypto Indices without BTC Constituent							
BITW20	0.2499	3.6210	-0.3866	4.3396	-21.6634	2021-05-19	0.0988
BITW70	0.2788	3.9257	-0.7635	5.1288	-24.5294	2021-05-19	0.1147

Table 6: Summary Statistics of Hedged Portfolios that minimize VaR 1%.

	Mean %	Std %	Skew	Kurt	MD %	MD date	VaR 5%
Individual Cryptos							
BTC	0.0176	0.3270	-1.0405	3.3742	-1.5689	2020-11-30	0.0134
ETH	0.2977	3.9132	0.9547	7.2414	-18.6061	2021-05-19	0.1026
ADA	0.5562	5.3466	1.1362	3.9334	-15.4795	2021-05-19	0.1106
LTC	-0.0852	4.1503	-0.7234	7.3208	-29.0915	2021-05-19	0.1030
XRP	0.0352	7.1658	1.1582	19.8506	-52.5727	2020-12-23	0.1387
Crypto Indices with BTC Constituent							
BITX	0.0593	1.0178	-0.5331	13.3100	-8.0299	2021-05-19	0.0247
CRIX	0.0738	0.9695	-0.4729	13.6500	-7.0185	2021-05-19	0.0245
BITW100	0.0823	1.2338	-1.9365	23.1938	-11.8752	2021-05-19	0.0347
Crypto Indices without BTC Constituent							
BITW20	0.2499	3.6210	-0.3866	4.3396	-21.6634	2021-05-19	0.0988
BITW70	0.2788	3.9257	-0.7635	5.1288	-24.5294	2021-05-19	0.1147

Table 7: Summary Statistics of Hedged Portfolios that minimize VaR 1%.

4.5 Hedging Effectiveness Results

In this section, we analyse the out-of-sample hedging effectiveness (HE) of BTCF as hedging. HE is defined as

$$\text{HE} = 1 - \frac{\rho_h}{\rho_s},$$

a measure of the percentage reduction of portfolio risk attribute, in our case the spot ρ_s, to hedged portfolio risk attribute ρ_h. A higher HE indicates a higher hedging effectiveness and larger risk reduction.

The HE above is a generalisation of Ederington measure of hedging performance, where we, in addition to variance, include other risk measures: Expected Shortfall 5% and 1% (ES5 and ES1), Value-at-Risk 5% and 1% (VaR5 and VaR1), and ERM. In particular, ES5 is recommended by the Basel Committee on Banking Supervision (BCBS) to replace VaR as a quantitative risk metrics system. The proposed reform aimed at enhancing the risk metric system's ability to capture tail risk. We obtain a time series of out-of-sample r^h of each hedging pair and each risk reduction objective by concatenating the out-of-sample results. Then, we apply stationary block bootstrapping (SB) to the time series introduced by Politis and Romano (1994) in our analysis in order to preserve the temporal structure of the data while sampling. The SB procedure is as follow. Assume a time series with N observations $\{X_t\}_{t\in[1,N]}$ is a strong stationary, weakly dependence time series of interest, we form blocks of samples $B = \{X_i, ..., X_{i+j-1}\}$. Index i is a random variable uniformly distributed over $[1, 2, ..., N]$ and j is geometric distributed random variable with parameter . The block index i and block length j are independent. For any index k which is greater than N, the sample X_k is defined to be $X_{k(\mod N)}$. For each block, we calculate the hedging effectiveness with different risk measures mentioned above. We choose $p = 0.005$, implying the expected block length is 200. 100 blocks are drawn for each risk minimising objective and spot.

From figure 9, we report, as expected, the BTC involving spots, the BTC, CRIX, BITX and BITW100, are well hedged by the BTCF. Surprisingly, the performances are consistent across different risk reduction objectives and different HE evaluation. The median HE to BTC generated by various risk reduction objectives is ranging from 89.45% to 99.31%, median HE to CRIX is ranging from 81.13% to 95.22%, median HE to BITX is ranging from 79.06% to 94.84%, median HE to BITW100 Is ranging from 71.07% to 92.98%.

The HE of BTCF to other cryptos and indices are substantially lower than to the BTC involving spots, but the consistency the performances across different risk reduction objectives and HE evaluation remains. The median HE to BITW20 generated by various risk reduction objectives is ranging from 24.67% to 47.02%, median HE to BITW70 is ranging from 23.61% to 49.30%, median HE to ADA is ranging from 9.01% to 29.30%, median HE to ETH Is ranging from 30.07% to 36.18%, median HE to LTC Is ranging from 37.74% to 51.30%, median HE to XRP Is ranging from 0.46% to 30.89%.

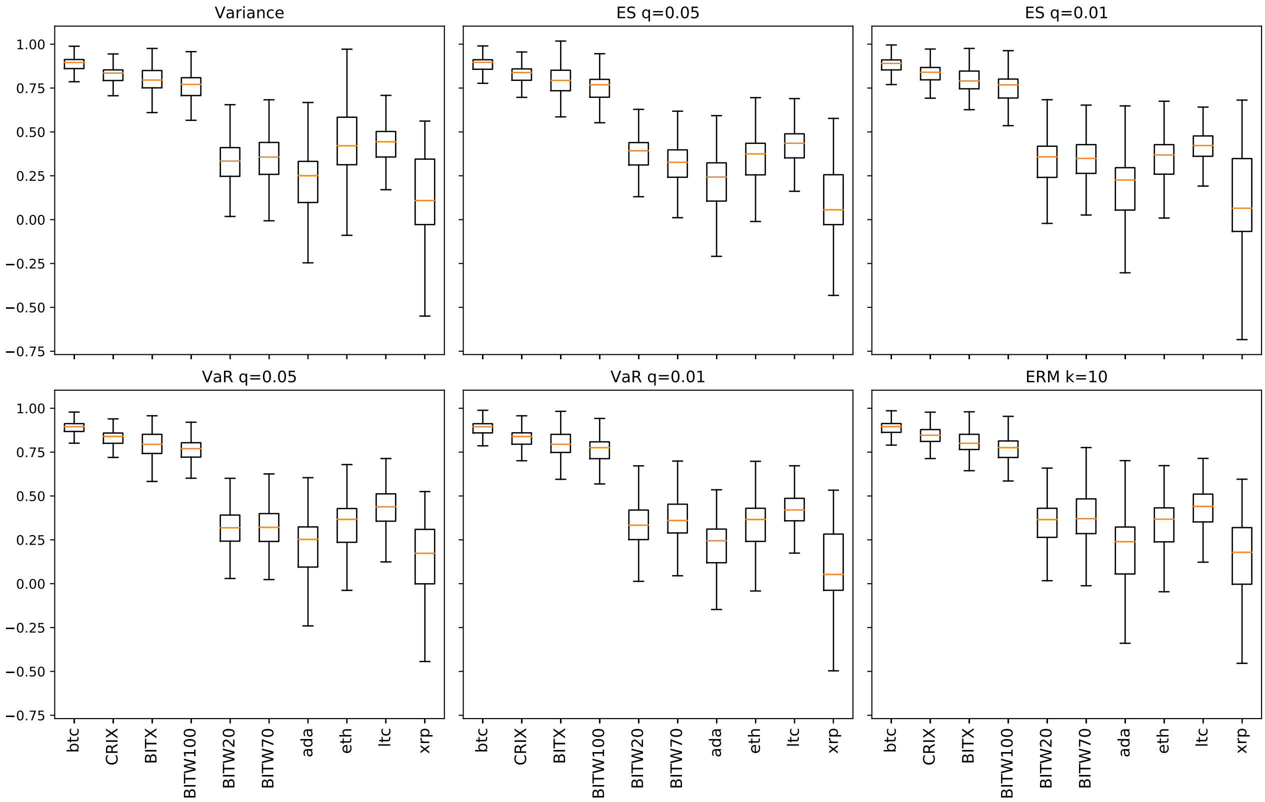

Figure 9: HE evaluated in the corresponding risk minimization objectives. The boxplots indicate the the median, upper quartile, lower quartile, minimium and maximum of the bootstrapped HE. The HE of BTC-involved spots are significantly higher than that of BTC-not-involved spots.

References

Acerbi, C. (2002): "Spectral measures of risk: A coherent representation of subjective risk aversion," *Journal of Banking & Finance*, 26, 1505–1518.

Anderson, D., K. Burnham, and G. White (1998): "Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies," *Journal of Applied Statistics*, 25, 263–282.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): "Coherent measures of risk," *Mathematical Finance*, 9, 203–228.

Barbi, M. and S. Romagnoli (2014): "A Copula-Based Quantile Risk Measure Approach to Estimate the Optimal Hedge Ratio," *Journal of Futures Markets*, 34, 658–675.

Bollerslev, T., V. Todorov, and L. Xu (2015): "Tail risk premia and return predictability," *Journal of Financial Economics*, 118, 113–134.

Breymann, W., A. Dias, and P. Embrechts (2003): "Dependence structures for multivariate high-frequency data in finance," .

Cherubini, U., S. Mulinacci, and S. Romagnoli (2011): "A copula-based model of speculative price dynamics in discrete time," *Journal of Multivariate Analysis*, 102, 1047–1063.

Demarta, S. and A. J. McNeil (2005): "The t copula and related copulas," *International statistical review*, 73, 111–129.

Dowd, K., J. Cotter, and G. Sorwar (2008): "Spectral risk measures: properties and limitations," *Journal of Financial Services Research*, 34, 61–75.

Ederington, L. H. and J. M. Salas (2008): "Minimum variance hedging when spot price changes are partially predictable," *Journal of Banking & Finance*, 32, 654–663.

Embrechts, P., A. McNeil, and D. Straumann (2002): "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, 1, 176–223.

Fama, E. F. (1963): "Mandelbrot and the stable Paretian hypothesis," *The Journal of Business*, 36, 420–429.

Fisher, N. I. and P. K. Sen (2012): *The collected works of Wassily Hoeffding*, Springer Science & Business Media.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995): "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82, 543–552.

Härdle, W. and L. Simar (2019): *Applied Multivariate Statistical Analysis*, Springer, 5th ed.

Härdle, W. K., N. Hautsch, and L. Overbeck (2008): *Applied Quantitative Finance*, Springer Science & Business Media.

Härdle, W. K., M. Müller, S. Sperlich, and A. Werwatz (2004): *Nonparametric and Semiparametric Models*, Springer Science & Business Media.

Härdle, W. K. and O. Okhrin (2010): "De copulis non est disputandum," *AStA Advances in Statistical Analysis*, 94, 1–31.

Hoeffding, W. (1940a): "Masstabinvariante korrelationstheorie," *Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin*, 5, 181–233.

——— (1940b): "Scale-invariant correlation theory (English translation)," 5, 181–233.

——— (1941): "Scale-invariant correlations for discontinuous distributions (English translation)," 7, 49–70.

Hull, J. C. (2003): *Options futures and other derivatives*, Pearson Education India.

Joe, H. (1997): *Multivariate models and multivariate dependence concepts*, CRC Press.

McNeil, A., R. Frey, and P. Embrechts (2005): *Quantitative Risk Management*, Princeton, NJ: Princeton University Press.

——— (2015): *Quantitative Risk Management*, Princeton, NJ: Princeton University Press, 2nd ed.

Nakamoto, S. (2009): "Bitcoin: A Peer-to-Peer Electronic Cash System," .

Nelsen, R. (2002): "Concordance and copulas: A survey," in *Distributions with Given Marginals and Statistical Modelling*, Kluwer Academic Publishers, 169–178.

NELSEN, R. B. (1999): *An Introduction to Copulas*, Springer.

OH, D. H. AND A. J. PATTON (2013): "Simulated method of moments estimation for copula-based multivariate models," *Journal of the American Statistical Association*, 108, 689–700.

POLITIS, D. N. AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 1303–1313.

SCHWEIZER, B., E. F. WOLFF, ET AL. (1981): "On nonparametric measures of dependence for random variables," *Annals of Statistics*, 9, 879–885.

SKLAR, A. (1959): "Fonctions de répartition a n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Universite de Paris*, 8, 229–231.

ZEEVI, A. AND R. MASHAL (2002): "Beyond correlation: Extreme co-movements between financial assets," *Available at SSRN 317122*.

5 Appendix

5.1 Density of linear combination of random variables

Proposition 6 *Let* $\boldsymbol{X} = (X_1, ..., X_d)^\top$ *be real-valued random variables with corresponding copula density* $\boldsymbol{c}_{X_1,...,X_d}$*, and continuous marginals* $F_{X_1}, ..., F_{X_d}$*. Then, pdf of the linear combination of marginals* $Z = n_1 \cdot X_1 + ... + n_d \cdot X_d$ *is*

$$f_Z(z) = \left|n_1^{-1}\right| \int_{[0,1]^{d-1}} \boldsymbol{c}_{X_1,...,X_d}\{F_{X_1} \circ S(z), u_2, ..., u_d\} \cdot f_{X_1} \circ S(z) du_2...du_d \tag{40}$$

$$S(z) = \frac{1}{n_1} \cdot z - \frac{n_2}{n_1} \cdot F_{X_2}^{(-1)}(u_2) - ... - \frac{n_d}{n_1} \cdot F_{X_d}^{(-1)}(u_d) \tag{41}$$

Proof. Rewrite $Z = n_1 \cdot X_1 + ... + n_d \cdot X_d$ in matrix form

$$\begin{bmatrix} Z \\ X_2 \\ \vdots \\ X_d \end{bmatrix} = \begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix}. \tag{42}$$

By transformation variables

$$\boldsymbol{f}_{Z,X_2,...,X_d}(z, x_2, ..., x_d) = \boldsymbol{f}_{X_1,...,X_d}\left(\boldsymbol{A}^{-1} \begin{bmatrix} z \\ x_2 \\ \vdots \\ x_d \end{bmatrix}\right) \cdot |\det \boldsymbol{A}^{-1}| \tag{43}$$

$$= \left|n_1^{-1}\right| \boldsymbol{f}_{X_1,...,X_d}\{S(z), x_2, ..., x_d\} \tag{44}$$

Let $u_i = F_{X_i}(x_i)$ and use the relationship

$$\boldsymbol{c}_{X_1,...,X_d}(u_1, ..., u_d) = \frac{\boldsymbol{f}_{X_1,...,X_d}(x_1, ..., x_d)}{\prod_{i=1}^d f_{X_i}(x_i)}, \tag{45}$$

we have

$$\boldsymbol{f}_{Z,X_2,...,X_d}(z, x_2, ..., x_d) = \tag{46}$$

$$\left|n_1^{-1}\right| \cdot \boldsymbol{c}_{X_1,...,X_d}\{F_{X_1} \circ S(z), u_2, ..., u_d\} \cdot f_{X_1}\{S(z)\} \cdot \prod_{i=2}^d f_{X_i}(x_i) \tag{47}$$

The claim (40) is obtained by integrating out $x_2, ...x_d$ by substituting $dx_i = \frac{1}{f_{X_i}(x_i)} du_i$. ■

	Mean (%)	Std (%)	Skew	Kurt	LQ (%)	MD (%)
Variance	-.19087	6.52591	1.1831	21.605	-2.35520	-52.52360
ES 95%	-.16368	6.46722	1.2248	22.287	-2.28830	-52.56981
ES 99%	-.15243	6.51684	1.2394	22.097	-2.32401	-52.56996
VaR 95%	-.17582	6.52438	1.1665	21.493	-2.34130	-52.56673
VaR 95%	-.14074	6.46029	1.2587	22.457	-2.31024	-52.57274
ERM	-.17083	6.49295	1.1923	21.932	-2.28506	-52.56895

Table 8: Daily Log returns statistics of BTC-BTCF hedged portfolios under different risk minimisation objectives.