Hedging Cryptos with Bitcoin Futures

Francis Liu[*]

Meng-Jou Lu [†]

Natalie Packham[‡]

Wolfgang Karl Härdle[§][¶]

This version: November 22, 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`.

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. *[Provide references.]* 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 future by 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. 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

3

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., $\text{VaR}_{1-\alpha} = -F_{R^h}^{(-1)}(1-\alpha) = -\inf\{x \in \mathbb{R} : 1 - \alpha \le F_{R^h}(x)\}$, 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 a sense that one needs to calibrate a new joint pdf $f_{r^S, -hr^F}$ when updating h. This is too time consuming and unnecessary. Another problem of using the joint pdf is that one lacks the flexibility to model the margins. A joint pdf completely determines the form of its marginals, for example, margins of a bivariate t-distribution are univariate t-distributions.

To overcome such a problem, we use copulae. The benefit of using copulae is two folded. First, copulae allow us to model the margins and dependence structure separately, see Sklar's Theorem. Second, copulae are invariant under strictly monotone increasing function (Schweizer et al., 1981), see lemma below.

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 (1) 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.

Lemma 1

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

Leveraging the two features of copulae, Barbi and Romagnoli (2014) introduces the distribution of linear combination of random variables using copulae. We slightly edit the Corollary 2.1 of their work and yield the following correct 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 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 $D_1 C(u, v) = \dfrac{\partial}{\partial u} C(u, v)$ see e.g. Equation (5.15) of (McNeil et al., 2005):

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

Proof.   Using the identity (4) gives

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

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

∎

Corollary 1 *Given the formulation of the above portfolio, 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

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

The two expressions are equivalent. Notice 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

We discuss the empirical procedure to obtain the OHR.

First, we split the time series of spots and future into sets of training and testing data. The training data is the first 300 observations and its corresponding testing data is the consecutive 5 observations. We roll 5 observations forward to obtain the next training and test data and repeat this until the end of the time series. Notice that the testing data are non-overlapping.

Next, we obtain the OHR as follows:

1. **Construct Univariate Kernel Density Function (KDE)**. With the training data, we obtain the spot and future's univariate kernel density function 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 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 lowest AIC is used for the next step. Discussion of this step is in 3.5.

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

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

3.1 Copulae

As we saw from the last section, risk measures we considered are all functionals of the joint distribution of r^S and r^F. We test different copulae: Gaussian-, t-, Frank-, Gumbel-, Clayton-, Plackett-, mixture, and factor copula. This hedging exercise concerns only portfolios with two assets, we only present the bivariate version of copulae and some important features of a copula, they include Kendall's τ, Spearman's ρ, 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 as $\boldsymbol{W}$, product copula as $\boldsymbol{\Pi}$, and the Fréchet-Hoeffding upper bound as $\boldsymbol{M}$, they represent cases of perfect negative dependence, independence, and perfect positive dependence respectively. For further detail, we refer readers to Joe (1997) and Nelsen (1999). See also Härdle and Okhrin (2010).

3.1.1 Elliptical Copulae

Elliptical copulae define dependence structures associated with elliptical distributions. The bivariate Gaussian copula is:

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

where $\Phi_{2,\rho}$ is the cdf of bivariate Normal distribution with zero mean, unit variance, and correlation ρ. , and Φ^{-1} is quantile function univariate standard normal distribution. 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 density is

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

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

The Kendall's τ_K and Spearman's ρ_S of a bivariate Gaussian Copula are

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

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

The t-Copula has a form

$$\boldsymbol{C}(u,v) = \boldsymbol{T}_{2,\rho,\nu}\{T_\nu^{-1}(u), T_\nu^{-1}(v)\}$$

$$= \int_{-\infty}^{T_\nu^{-1}(u)} \int_{-\infty}^{T_\nu^{-1}(v)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi\nu\sqrt{1-\rho^2}} \tag{11}$$

$$\left(1 + \frac{s^2 - 2st\rho + t^2}{\nu}\right)^{-\frac{\nu+2}{2}} ds dt, \tag{12}$$

where $\boldsymbol{T}_{2,\rho,\nu}(\cdot,\cdot)$ denotes the cdf of bivariate t distribution with scale parameter ρ and degree of freedom ν, $T_\nu^{-1}(\cdot)$ is the quantile function of a standard t distribution with degree of freedom ρ.

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 $\boldsymbol{T}_{2,\rho,\nu}(\cdot,\cdot)$, and t_ν the density of standard t distribution.

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

3.1.2 Archimedean Copulae

The family of Archimedean copulae forms a large class of copulae with many convenient features. In general, they take a form

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

where $\psi : [0,1] \rightarrow [0,\infty)$ 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.

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

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

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 $D_n = \frac{n}{x^n}\int_0^x \frac{t^n}{e^t - 1}dt$.

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.

$$\boldsymbol{C}_\theta(u, v) = \exp -\{(-\log(u))^\theta + (-\log(v))^\theta\}^{\frac{1}{\theta}}, \tag{19}$$

where $\theta \in [1, \infty)$ is the dependence parameter. While Gumbel copula cannot model perfect counter dependence (Nelsen, 2002), $\boldsymbol{C}_1 = \boldsymbol{\Pi}$ models the independence, and $\lim_{\theta\to\infty} \boldsymbol{C}_\theta = \boldsymbol{W}$ models the perfect dependence.

$$\tau_K(\theta) = \frac{\theta - 1}{\theta} \tag{20}$$

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

The Clayton copula takes the form

$$\boldsymbol{C}_\theta(u, v) = \left\{ \max(u^{-\theta} + v^{-\theta} - 1, 0) \right\}^{-\frac{1}{\theta}}, \tag{21}$$

where $\theta \in (-\infty, \infty)$ is the dependency parameter. $\lim_{\theta\to-\infty} \boldsymbol{C}_\theta = \boldsymbol{M}$, $\boldsymbol{C}_0 = \boldsymbol{\Pi}$, and $\lim_{\theta\to\infty} \boldsymbol{C}_\theta = \boldsymbol{W}$.

Kendall's τ to this copula dependency is

$$\tau_K(\theta) = \frac{\theta}{\theta + 2}. \tag{22}$$

3.1.3 Mixture Copula

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 K copulae

$$\mathbf{P}(X_1 \leq x_1, X_2 \leq x_2) = \sum_{k=1}^{K} p^k \cdot \boldsymbol{C}^{(k)}\{F_{X_1}^{(k)}(x_1; \boldsymbol{\gamma}_1^{(k)}), F_{X_2}^{(k)}(x_2; \boldsymbol{\gamma}_2^{(k)}); \boldsymbol{\theta}^{(k)}\} \tag{23}$$

where $p^{(k)} \in [0, 1]$ is the weight of each component, $\boldsymbol{\gamma}^{(k)}$ is the parameter of the marginal distribution in the kth component, and $\boldsymbol{\theta}^{(k)}$ is the dependence parameter with the copula of the kth 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 mixture. By Sklar's theorem one may write

$$\boldsymbol{C}(u, v) = \sum_{k=1}^{K} p^{(k)} \cdot \boldsymbol{C}^{(k)}\{F_{X_1}^{-1}(u), F_{X_2}^{-1}(v); \boldsymbol{\theta}^{(k)}\}. \tag{24}$$

The copula density is again a linear combination of copula density

$$\boldsymbol{c}(u, v) = \sum_{k=1}^{K} p^{(k)} \cdot \boldsymbol{c}^{(k)}\{F_{X_1}^{-1}(u), F_{X_2}^{-1}(v); \boldsymbol{\theta}^{(k)}\}. \tag{25}$$

While Kendall's τ of mixture copula is not known in close form, the Spearman's ρ is

Proposition 3 *Let $\rho_S^{(k)}$ be the Spearman's ρ of the kth component and $\sum_{k=1}^{K} p^{(k)} = 1$ holds, the Spearman's ρ of a mixture copula is*

$$\rho_S = \sum_{k=1}^{K} p^{(k)} \cdot \rho_S^{(k)} \tag{26}$$

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

$$\rho_S = 12 \int_{\mathbb{I}^2} \boldsymbol{C}(s, t) ds dt - 3. \tag{27}$$

Rewrite the mixture copula into sumation of components

$$\rho_S = 12 \int_{\mathbb{I}^2} \sum_{k=1}^{K} p^{(k)} \cdot \boldsymbol{C}^{(k)}(s, t) ds dt - 3. \tag{28}$$

$\blacksquare$

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)*

$$\boldsymbol{C}_{\alpha,\beta}(u, v) = \alpha \boldsymbol{M}(u, v) + (1 - \alpha - \beta)\boldsymbol{\Pi}(u, v) + \beta \boldsymbol{W}(u, v), \tag{29}$$

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

$$\tau_K(\alpha, \beta) = \frac{(\alpha - \beta)(\alpha + \beta + 2)}{3} \tag{30}$$

, and

$$\rho_S(\alpha, \beta) = \alpha - \beta \tag{31}$$

We use a mixture of Gaussian and independent copula in our analysis. We write the copula

$$\boldsymbol{C}(u, v) = p \cdot \boldsymbol{C}^{\text{Gaussian}}(u, v) + (1 - p)(uv). \tag{32}$$

The corresponding copula density is

$$\boldsymbol{c}(u, v) = p \cdot \boldsymbol{c}^{\text{Gaussian}}(u, v) + (1 - p). \tag{33}$$

This mixture allows us to model how much "random noise" appear in the dependency structure. In this hedging exercise, the structure of the "random noise" is not of our concern nor we can hedge the noise by a two-asset portfolio. However, the proportion of the "random noise" does affect the distribution of r^h, so as the optimal hedging ratio h^*. One can consider the mixture copula as a handful 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.

3.2 Other Copula

The Plackett copula has an expression

$$\boldsymbol{C}_\theta(u, v) = \frac{1 + (\theta - 1)(u + v)}{2(\theta - 1)} - \frac{\sqrt{\{1 + (\theta - 1)(u + v)\}^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)} \tag{34}$$

$$\rho_S(\theta) = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \log \theta}{(\theta - 1)^2} \tag{35}$$

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

$$\frac{\mathbf{P}(U \leq u, V \leq v) \cdot \mathbf{P}(U > u, V > v)}{\mathbf{P}(U \leq u, V > v) \cdot \mathbf{P}(U > u, V \leq v)}$$
$$= \frac{\boldsymbol{C}_\theta(u, v)\{1 - u - v + \boldsymbol{C}_\theta(u, v)\}}{\{u - \boldsymbol{C}_\theta(u, v)\}\{v - \boldsymbol{C}_\theta(u, v)\}}$$
$$= \theta. \tag{36}$$

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

3.3 Estimation

3.3.1 Simulated Method of Moments

This method is suggested by (Oh and Patton, 2013). 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 a pair (X, Y) with copula C are defined as

$$\rho_S = 12 \int \int_{I^2} C_{\boldsymbol{\theta}}(u, v)\, \mathrm{d}u\, \mathrm{d}v - 3 \tag{37}$$

$$\tau_K = 4\, \mathsf{E}[C_{\boldsymbol{\theta}}\{F_X(x), F_Y(y)\}] - 1, \tag{38}$$

$$\lambda_q = \begin{cases} \mathbf{P}(F_X(X) \leq q | F_Y(Y) \leq q) = \dfrac{C_{\boldsymbol{\theta}}(q, q)}{q}, & \text{if } q \in (0, 0.5], \\ \mathbf{P}(F_X(X) > q | F_Y(Y) > q) = \dfrac{1 - 2q + C_{\boldsymbol{\theta}}(q, q)}{1 - q}, & \text{if } q \in (0.5, 1). \end{cases} \tag{39}$$

The empirical counterparts are

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

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

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

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

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

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

The SMM estimator is

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

where $\hat{W}$ is some positive definite weight matrix.

In this work, we use $\tilde{\boldsymbol{m}}(\boldsymbol{\theta}) = (\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95})^\top$ for calibration of Bitcoin price and CME Bitcoin future.

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

$$\boldsymbol{f_X}(x_1, ..., x_d) = \boldsymbol{c}\{F_{X_1}(x_1), ..., F_{X_d}(x_d)\} \prod_{j=1}^{d} f_{X_i}(x_i). \tag{40}$$

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

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{c}\{F_{X_1}(x_{i,1}; \boldsymbol{\delta}_1), ..., F_{X_d}(x_{i,d}; \boldsymbol{\delta}_d); \boldsymbol{\gamma}\} + \sum_{i=1}^{n}\sum_{j=1}^{d} f_{X_j}(x_{i,j}; \boldsymbol{\delta}_j) \tag{41}$$

$$= L_C(\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_d, \boldsymbol{\gamma}) + \sum_{j=1}^{d} L_j(\boldsymbol{\delta}_j) \tag{42}$$

where $\boldsymbol{\delta}_j$ is the parameter 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 $\boldsymbol{\theta}$ is 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 $L_C(\hat{\boldsymbol{\delta}}_1, ..., \hat{\boldsymbol{\delta}}_d, \boldsymbol{\gamma})$.

That is, under regularity conditions, $(\hat{\boldsymbol{\delta}}_1, ..., \hat{\boldsymbol{\delta}}_d, \hat{\boldsymbol{\gamma}})$ is the solution of

$$\left(\frac{\partial L_1}{\partial \boldsymbol{\delta}_1}, ..., \frac{\partial L_d}{\partial \boldsymbol{\delta}_d}, \frac{\partial L_C}{\partial \boldsymbol{\gamma}} \right) = \boldsymbol{0}. \tag{43}$$

However, the IFM requires making assumption on the distribution of the margins. Genest et al. (1995) suggests to replace the estimation of marginals parameters estimation by non-parameteric estimation. Given non-parametric estimator $\hat{F}_i$ of the margins F_i, the estimator of the dependence parameters $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\mathrm{argmax}} \sum_{i=1}^{n} \boldsymbol{c}\{\hat{F}_{X_1}(x_{i,1}), ..., \hat{F}_{X_d}(x_{i,d}); \boldsymbol{\gamma}\}. \tag{44}$$

3.4.1 Comparison

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 from 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, and those points generate PnL to the hedged portfolio. MM is preferred visually as it produces a better fit to the dependence structure in the two extremes. Therefore, we deploy the method of moments throughout the analysis. We choose the 5th-, 10th-, 90th-, 95th-quantile, and Spearman's ρ as the moments.
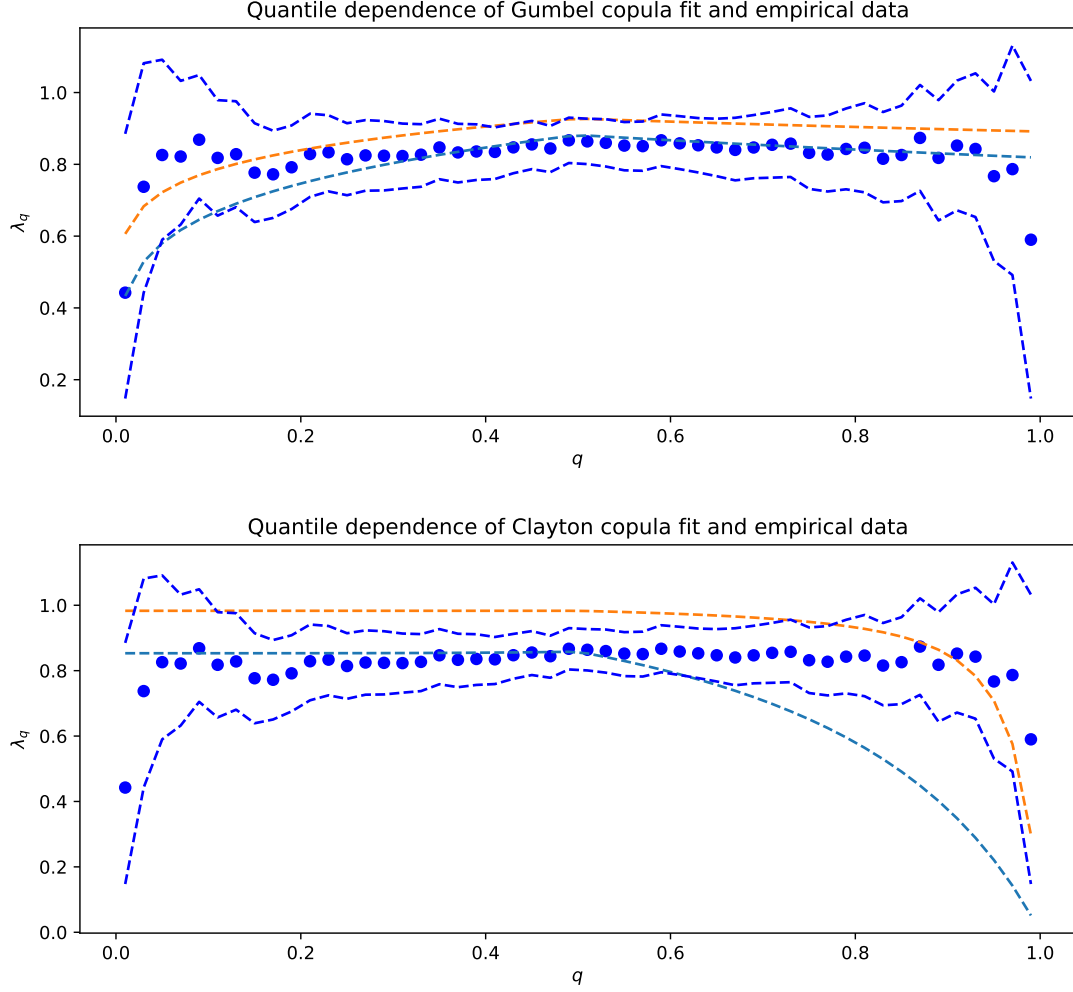
Figure 1: Quantile dependences of Gumbel, and Clayton Copula. The blue circle dots are the quantile dependence estimates of Bitcoin and CME future, blue dotted lines are the estimates' 90% confidence interval. Orange dotted line is the copula implied quantile dependence by MM estimation. Light blue dotted line is the copula implied quantile dependence by MLE estimation.

3.5   Copula Selection

The dependency structure of price data changes across time, in which both the dependency parameters and the type of dependency, dependencies between cryptos and the BTCF are no exception. 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 that provider 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.

Notice that there are other model selection procedure and criteria, e.g. TIC , likelihood ratio test. 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, 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 in kernel density. The selected copula will then be used to search for OHR.

3.6 Risk Measures

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).

Let Z be a random variable of distribution F_Z.

1. Variance is $\text{Var}(F_Z)$

2. VaR of a given confidence level α is $\text{VaR}(F_Z) = -F_Z^{(-1)}(1-\alpha)$

3. ES with parameter α is $\text{ES}(F_Z) = -\frac{1}{1-\alpha}\int_0^{1-\alpha} F_Z^{(-1)}(p)dp$

4. ERM with Arrow-Pratt coefficient of absolute risk aversion k is $\text{ERM}_k(F_Z) = \int_0^{1-\alpha} \phi(p)F_Z^{(-1)}(p)dp$ where ϕ is a weight function described in (46) below.

VaR, ES, and ERM fall into the class of Spectral Risk Measure (SRM). SRM has the from (Acerbi, 2002)

$$\rho_\phi(r^h) = -\int_0^1 \phi(p)F_Z^{(-1)}(p)dp, \tag{45}$$

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

  i ϕ is positive

  ii ϕ is decreasing

 iii   integrates to one.

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, hence violates the ii property of admissible risk spectra. The ES' $\phi(p)$ gives all tail quantiles the same weight of $\frac{1}{1-\alpha}$ and non-tail quantiles zero weight. ERM assumes investor's risk preference is in a form of exponential utility function $U(x) = 1 - e^{kx}$, its corresponding risk spectrum is defined as

$$\phi(p) = \frac{ke^{-k(1-p)}}{1-e^{-k}}, \tag{46}$$

where k is the Arrow-Pratt coefficient of absolute risk aversion.

The parameter k has an economic interpretation of being the ratio between the second derivative and first derivative of investor's utility function on an risky asset

$$k = -\frac{U''(x)}{U'(x)}, \tag{47}$$

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

4 Data

In the empirical analysis, we consider the risk reduction capability of the BTC-future (BTCF) on five cryptos , BTC, ETH, ADA, LTC, and XRP, and five crypto indexes, BITX, BITW100, CRIX, 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, and the BTCF is held in short position of OHR unit in order to reduce the risk of the spot. All the hedging portfolios are cross asset hedging except the BTCF portfolio. 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 track 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.

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 BTCFs based on the trading activities on CME Globex between 14:59 and 15:00 CT. 15:00 CT is also the reporting time of the daily closing price by the Bloomberg Terminal (BBT). Cryptos data are collected from a data provider called Tiingo. Tiingo aggregates crypto OHLC (open, high, low, and close) prices fed by APIs from various exhcanges. Tiingo covers major exchanges, e.g. Binance, Gemini, Poloniex etc., so Tiingo's aggregated OHLC price is a good representation a market tradable price. For each crypto, we match the opening price at 15:00 CT from Tiingo with the daily closing price of BTCF from BBT. Since CRIX is not available at 15:00 CT, we recalculate a 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 public via S&P.

5 Results

5.1 Overview

We visualize the out-of-sample hedged portfolio returns and the corresponding OHR in figure 2. We also indicate the dates when the five smallest returns occur in the spot return in black vertical lines.

Overall, the BTC-involved spots, i.e. BTC, CRIX, BITX, and BITW100, are well hedged by the BTCF. The overall magnitude of the hedged portfolio returns are greatly reduced, as well as the extreme losses in spot are no longer the extreme losses in the hedged portfolio returns. Most of the time different risk reduction objectives generate very similar hedged portfolio returns as one can from the plot that the hedged portfolio returns are overlapping to each other.

The OHRs of these portfolios are close to one regardless of which risk minimization objective is taken. It is natural for the OHRs of the BTC-BTCF close to one as the futures price F is governed

by the underlying price S and interest rate r

$$F = Se^{rT},$$

where T is the futures' time to maturity. It s not difficult to show that the arithmetic difference between the daily log returns is almost zero, i.e. $\log \frac{F_{t+1}}{F_t} \approx \log \frac{S_{t+1}}{S_t}$. This relationship explains the every OHRs in BTC-BTCF portfolio are close to one.

Most of the time the OHRs in the BTC-BTCF portfolio are steady and well-performing, but we can find fluctuations in MES99%HR and MVaR99%HR. In particular, MES99%HR fluctuates greatly from Jun 2019 to Aug 2019. The MES99%HR is lower than its counterparts and so the resulting hedge fail to capture the magnitude of the negative return of BTC. The MVaR99%HR performs similarly in that period time with less fluctuations. We suspect the fluctuating and malfunctioning MES99%HR and MVaR99%HR come from fact that ES 99% and VaR 99% are sensitive to small changes in the left tail as they consider a small fraction of the data.

On the other hand, BTC-not-involved spots'portfolios, i.e. BITW20-, BITW70-, ETH-, ADA-, LTC-, and XRP-BTCF, are less promising. The hedged portfolio returns are almost as volatile as the spots. Nonetheless, the the extreme downward jumps of the spots are not taken care by the hedge. We will further discuss the effectiveness of hedge in section 5.4.

In addition to the time series plots, 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 $= \frac{1}{n} \sum_{i=1}^{n} \{E(r^h)r_i^h -\}^2$. All results here are out-of-sample results obtained without the copula selection step in order to compare the performances across copulae.

Figure 3 and 4 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 5 and 6, 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 5 and 6 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 Xs also provide the information about the risk reduction objectives. In general, portfolios that minimise variance, ES 95%, VaR 95%, and ERM are very similar in means square error and lower semi variance. Contrarily, portfolios that minimise ES 99% and VaR 99% are have high mean square error and lower semi variance. This may due to the fact that ES 99% and VaR 99% are easily affected extreme events in the training data.

5.2 Copula Selection Results

The copula selection result provides insight that help us understand the crypto market better, so we illustrate the copula selection result in this section. 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 $(U_1, ..., U_d) \overset{\mathcal{L}}{=} (1 - U_1, ..., 1 - U_d)$. This symmetry can be justified in the dependence structure between a future and its underlying by the theory of future pricing, which suggests the price of a future is a function of the underlying price (Hull, 2003). However, there is no such relationship between a future 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 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.

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.

	Mean (%)	Std (%)	Skew	Kurt	LQ (%)	MD (%)
BTC	.01948	.39138	-0.8949	11.867	-.12785	3.18969
CRIX	.08048	.88261	0.0677	14.939	-.36619	7.05299
BITX	.58960	1.00776	-0.4427	13.084	-.35729	7.85815
BITW100	.85260	1.20317	-1.6522	20.556	-.42111	11.18459
BITW20	.25642	3.60091	-0.3446	4.215	-1.35096	21.59201
BITW70	.28181	3.90745	-0.6952	4.874	-1.60984	24.52498
ADA	.13983	5.40242	1.0344	4.201	-2.65629	19.40610
ETH	.10541	3.79113	0.8743	6.759	-1.70680	18.87288
LTC	-.20647	4.27759	-0.0729	5.628	-2.22628	28.08793
XRP	-.17083	6.49295	1.1923	21.932	-2.28506	52.56895

Table 2: Statistics of daily log returns of hedged portfolios aimed at minimizing ERM $k = 10$.

5.3 OHRs and Returns of Hedged Portfolios

We present an overview of the OHR and hedged portfolios log returns results in this section. Overall, the performance of the hedge largely depend on the spot assets, while the choice of risk minimisation objective does not affect much of the log return of hedge ratio. We present the summary statistics of the hedged portfolios returns in table 2 that aimed at minimising ERM $k = 10$.

We can see from table 2 that the summary statistics of hedged portfolio returns differ. As expected, the BTC-involved portfolios, i.e. portfolios with BTC, CRIX, BITX, and BITW100 as spot, have a relatively small standard deviations (std) and maximum drawdowns (MD). However, the kurtosis (kurt) of the portfolio returns are higher than the other portfolios except XRP. This result suggests that the variance of the BTC-involved spots are well hedged but not the jump risk (extreme events).

In fact, extreme events are hard to model and predict, even the BTC-BTCF portfolio cannot escape from this. Nonetheless, the skewness (skew) are close to zero.

Portfolios with spot of crypto indices that are not BTC-involved, i.e. BITW20 and BITW70, have relatively lower std and MD than portfolio with individual cryptos as spot. This indicates the overall market is indeed follow the BTC or the BTCF's development, but individual cryptos can have their own price movement just as discuss in the last section.

Next, we illustrate the OHR time series of the BTC-BTCF and other hedge portfolios in figure 7 and ?? respectively. The two most fluctuating OHR time series are generated by minimising ES 99% and VaR 99%. ES 99% and VaR 99% are calculated based on a small fraction of data from the left tail of portfolio return, so the resulting OHR time series are relatively sensitive to extremes in the training data. The other OHR time series do not deviate from each other a lot.

In figure ??, we can observe a large range of OHR in other hedge portfolios. For BTC-involved indices, the OHRs are ranging from 0.75 to 1.05, slightly wider than range of BTC-BTCF's OHR.

Again, we present the OHR time series generated by minimising ES 99% and VaR 99% are more fluctuating in different hedge portfolios. Furthermore, we can see more clearly in figure ?? that OHR minimising ES 99% are more extreme.

5.4 Hedging Effectiveness Results

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

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

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 8, 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%.

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*.

6 Appendix

6.1 Density of linear combination of random variables

Proposition 5 *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*

$$f_Z(z) = |n_1^{-1}| \int_{[0,1]^{d-1}} \boldsymbol{c}_{X_1,...,X_d}\{F_{X_1} \circ S(z), u_2, ..., u_d\} \cdot f_{X_1} \circ S(z) du_2...du_d \tag{48}$$

$$S(z) = \frac{1}{n_1} \cdot z - \frac{n_2}{n_1} \cdot F_{X_2}^{(-1)}(u_2) - ... - \frac{n_d}{n_1} \cdot F_{X_d}^{(-1)}(u_d) \tag{49}$$

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

$$\begin{bmatrix} Z \\ X_2 \\ \vdots \\ X_d \end{bmatrix} = \begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix}. \tag{50}$$

By transformation variables

$$\boldsymbol{f}_{Z,X_2,...,X_d}(z, x_2, ..., x_d) = \boldsymbol{f}_{X_1,...,X_d}\left(\boldsymbol{A}^{-1}\begin{bmatrix} z \\ x_2 \\ \vdots \\ x_d \end{bmatrix}\right) \cdot |\det \boldsymbol{A}^{-1}| \tag{51}$$

$$= |n_1^{-1}| \, \boldsymbol{f}_{X_1,...,X_d}\{S(z), x_2, ..., x_d\} \tag{52}$$

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

$$\boldsymbol{c}_{X_1,...,X_d}(u_1, ..., u_d) = \frac{\boldsymbol{f}_{X_1,...,X_d}(x_1, ..., x_d)}{\prod_{i=1}^d f_{X_i}(x_i)}, \tag{53}$$

we have

$$\boldsymbol{f}_{Z,X_2,...,X_d}(z, x_2, ..., x_d) = \tag{54}$$

$$|n_1^{-1}| \cdot \boldsymbol{c}_{X_1,...,X_d}\{F_{X_1} \circ S(z), u_2, ..., u_d\} \cdot f_{X_1}\{S(z)\} \cdot \prod_{i=2}^d f_{X_i}(x_i) \tag{55}$$

The claim (48) is obtained by integrating out $x_2, ...x_d$ by substituting $dx_i = \frac{1}{f_{X_i}(x_i)} du_i$. ∎
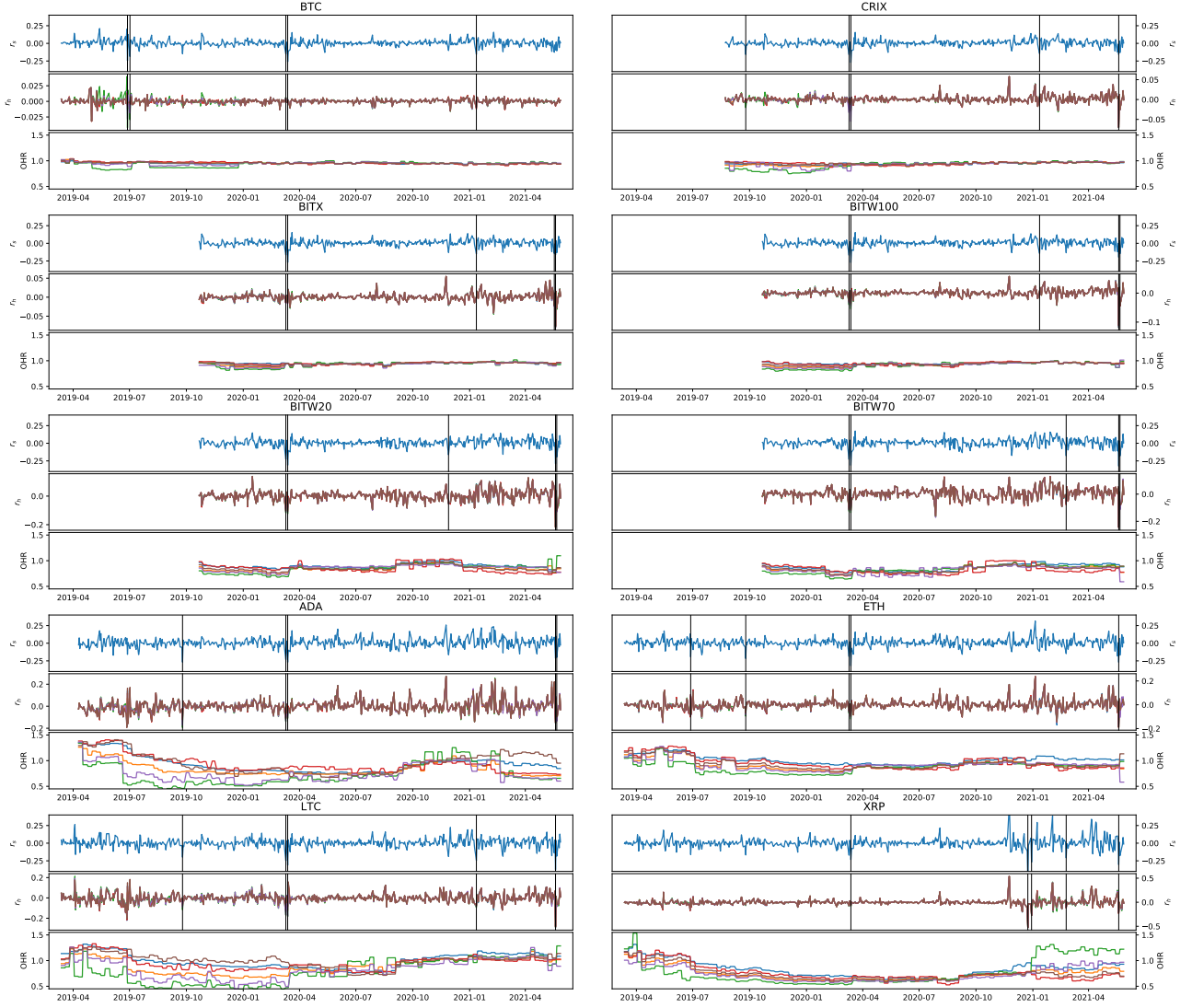
Figure 2: An overview of out-of-sample hedged portfolios. For each panel, the line plot on the top is the daily log returns of the spot; line plot in the middle is the daily log returns of hedged portfolio with different risk reduction objectives; line plot at the bottom is the OHR of different risk reduction objectives. The coloring of risk reduction objectives is as follows: blue line for variance, orange line for ES 95%, green line for ES 99%, red line for VaR 95%, purple line for VaR 99%, and brown line for ERM $k = 10$. The vertical black lines in the top and middle line plots indicate the dates when the five smallest returns occur in the spot return. Most of the time the portfolio returns of different risk minimisation objectives overlap wth each other. We can see that those BTC-involved spots, i.e. BTC, CRIX, BITX, and BITW100, are well hedged by the BTCF as the magnitude of the hedged portfolio returns are greatly reduced, as well as the five extreme losses of the spots are well managed. One exception is the BTC-BTCF portfolio that minimize ES 99% (first panel green line in middle plot), which fluctuate greatly from Jun 2019 to Aug 2019. In that period of time, the portfolio OHR is lower than its counterparts. Hedges of spots that are not BTC related are less promising. The effectiveness of a hedge is discussed in section 5.4.
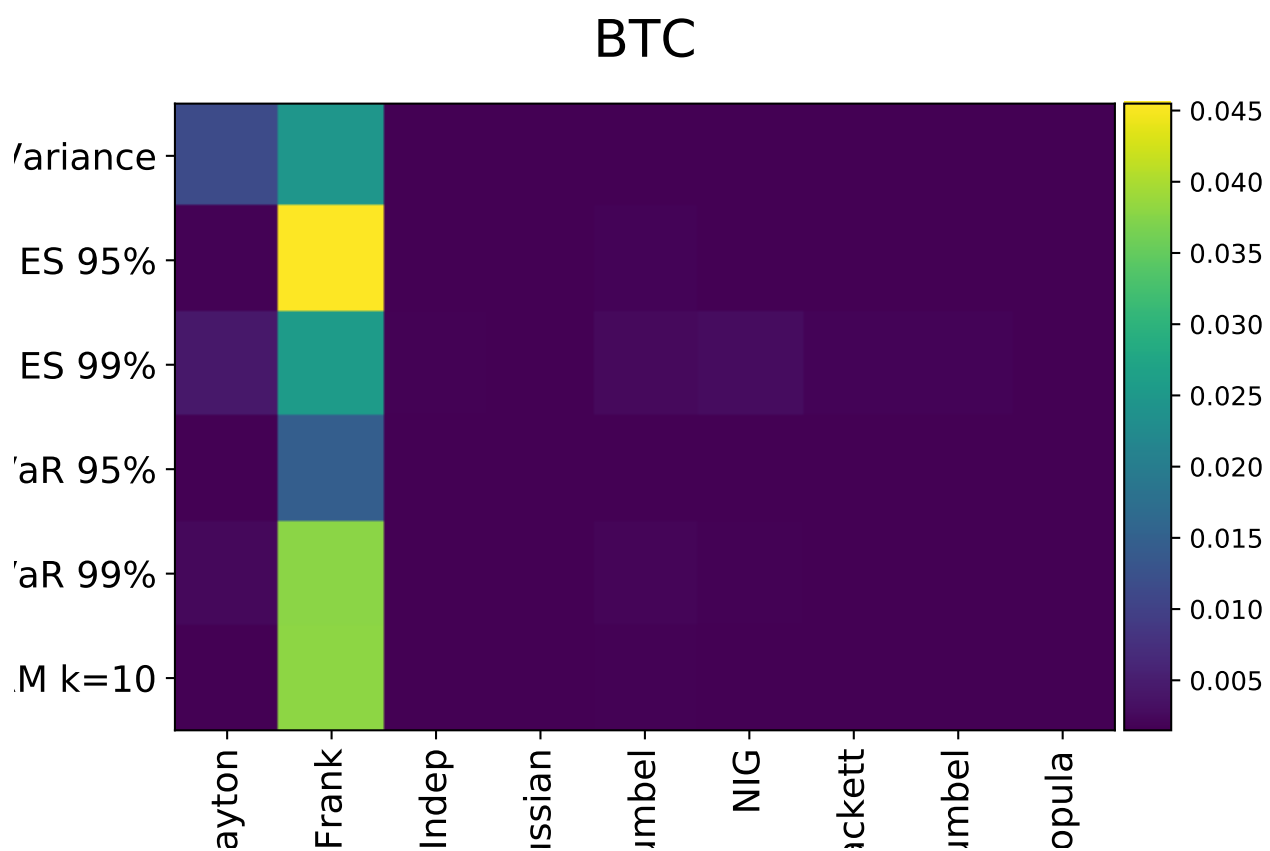
24

Figure 3: 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.
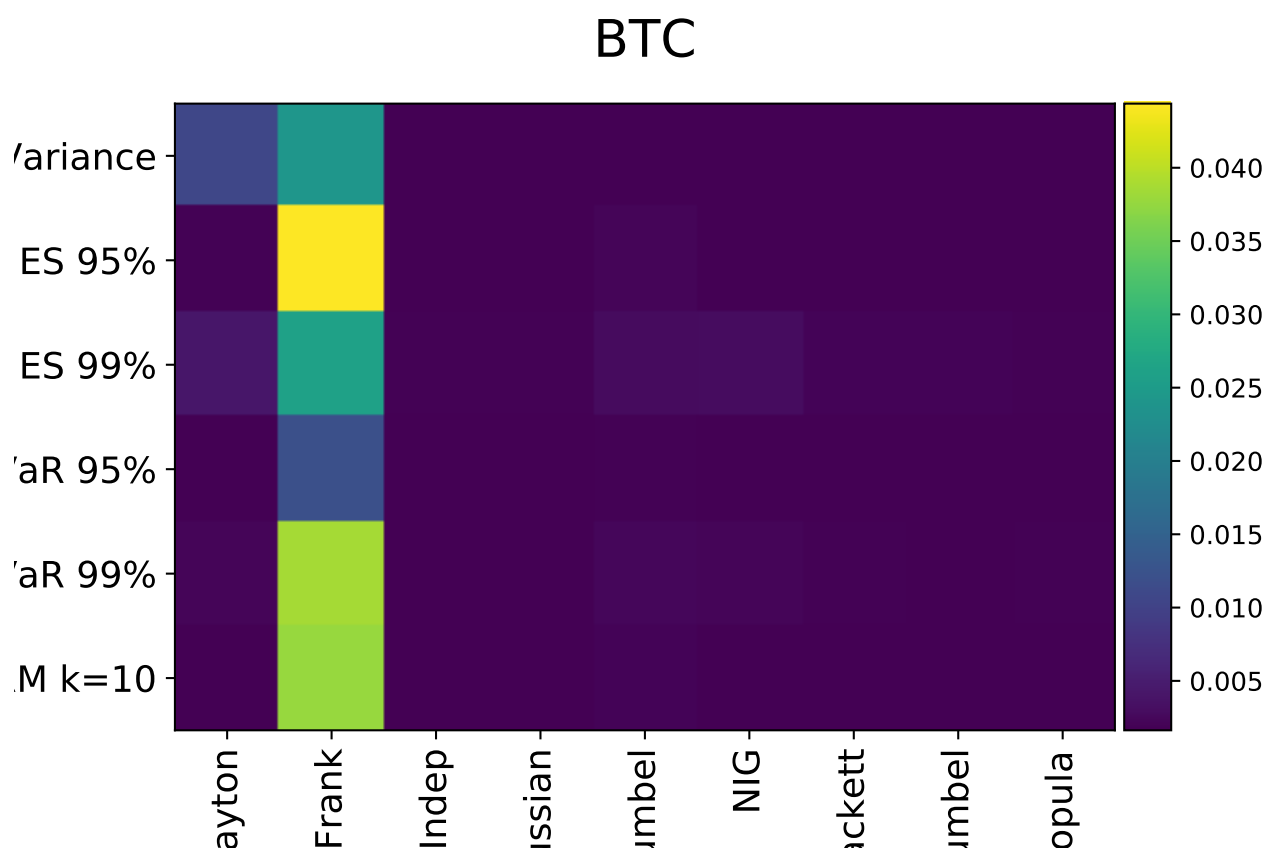
Figure 4: Lower semivariance of BTC-BTCF portfolios constructed with different copula and risk minimization objectives. The Frank copula is obviously inferior.
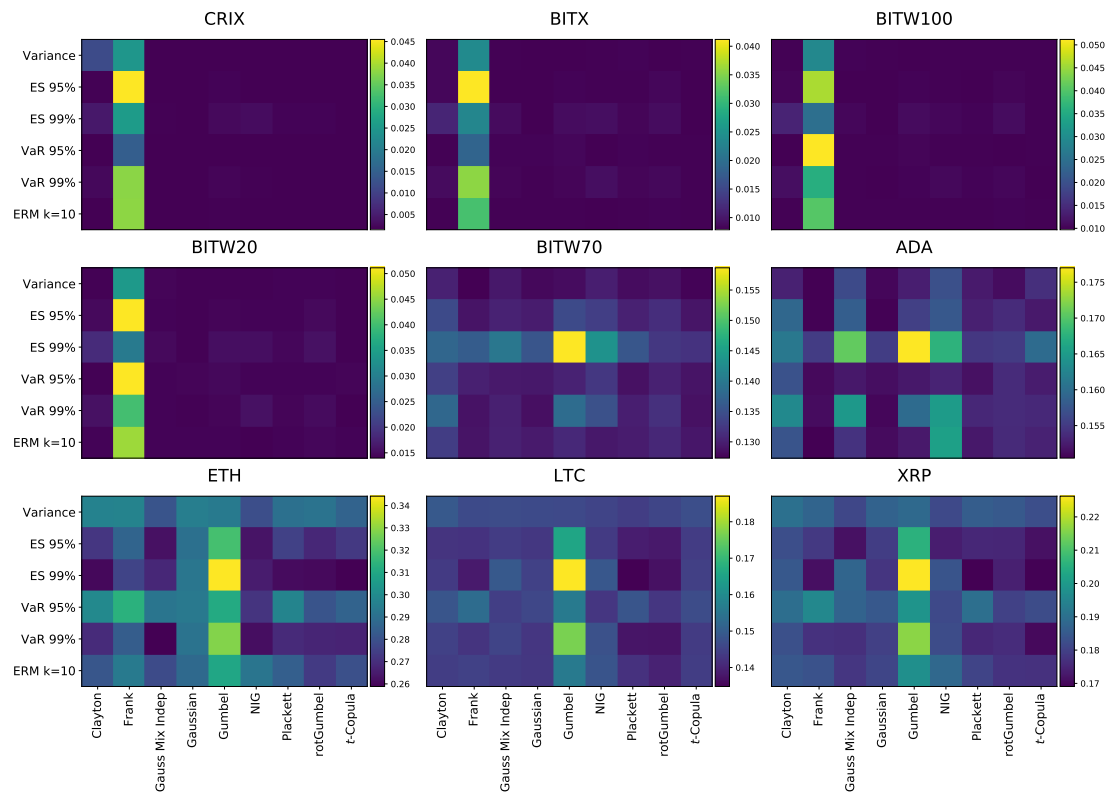
Figure 5: Mean square errors of portfolios constructed with different copula and risk minimization objectives.
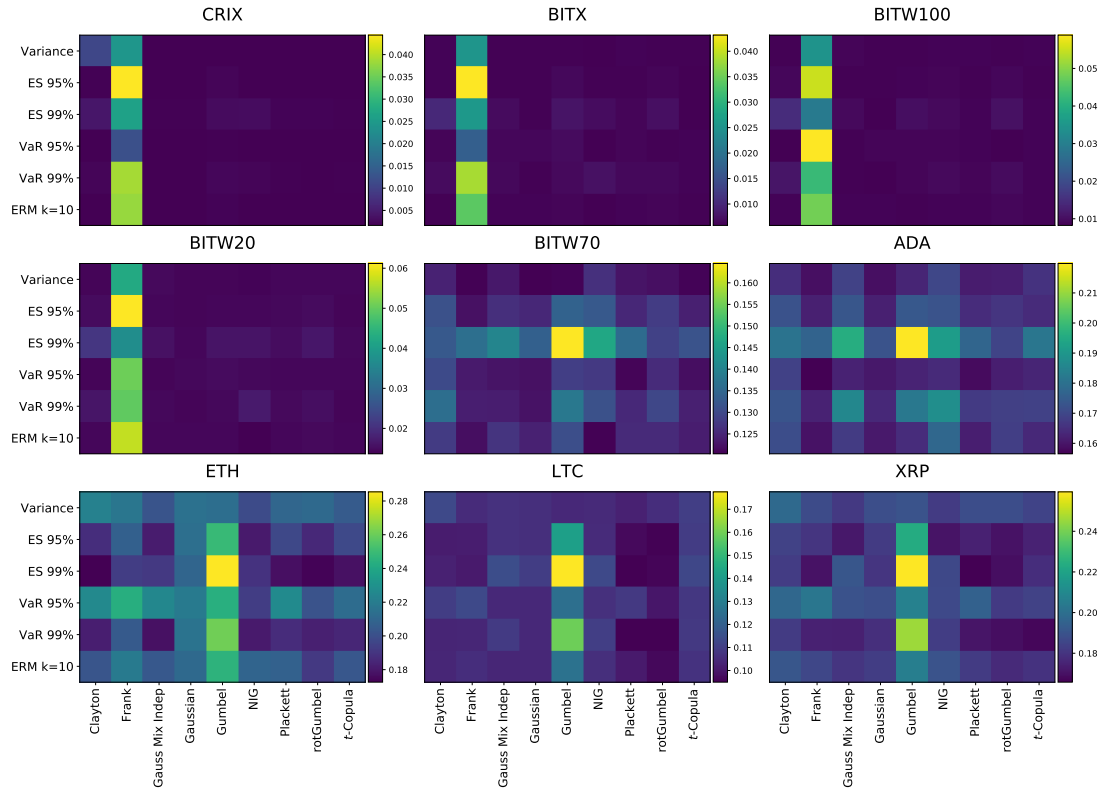
Figure 6: Lower semivariance of portfolios constructed with different copula and risk minimization objectives.
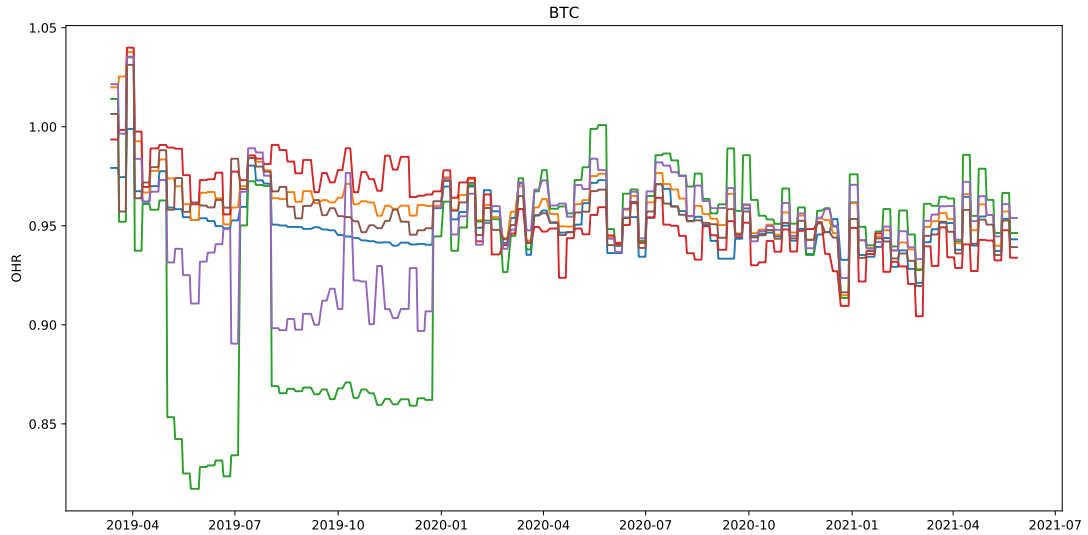

Figure 7: OHR of BTC-BTCF portfolio with different risk minimizing objectives. The blue line is OHR minimizing variance, orange line is that of ES 95%, green line is that of ES 99%, red line is that of VaR 95%, purple line is that of VaR 99%, and brown line is that of ERM $k = 10$.
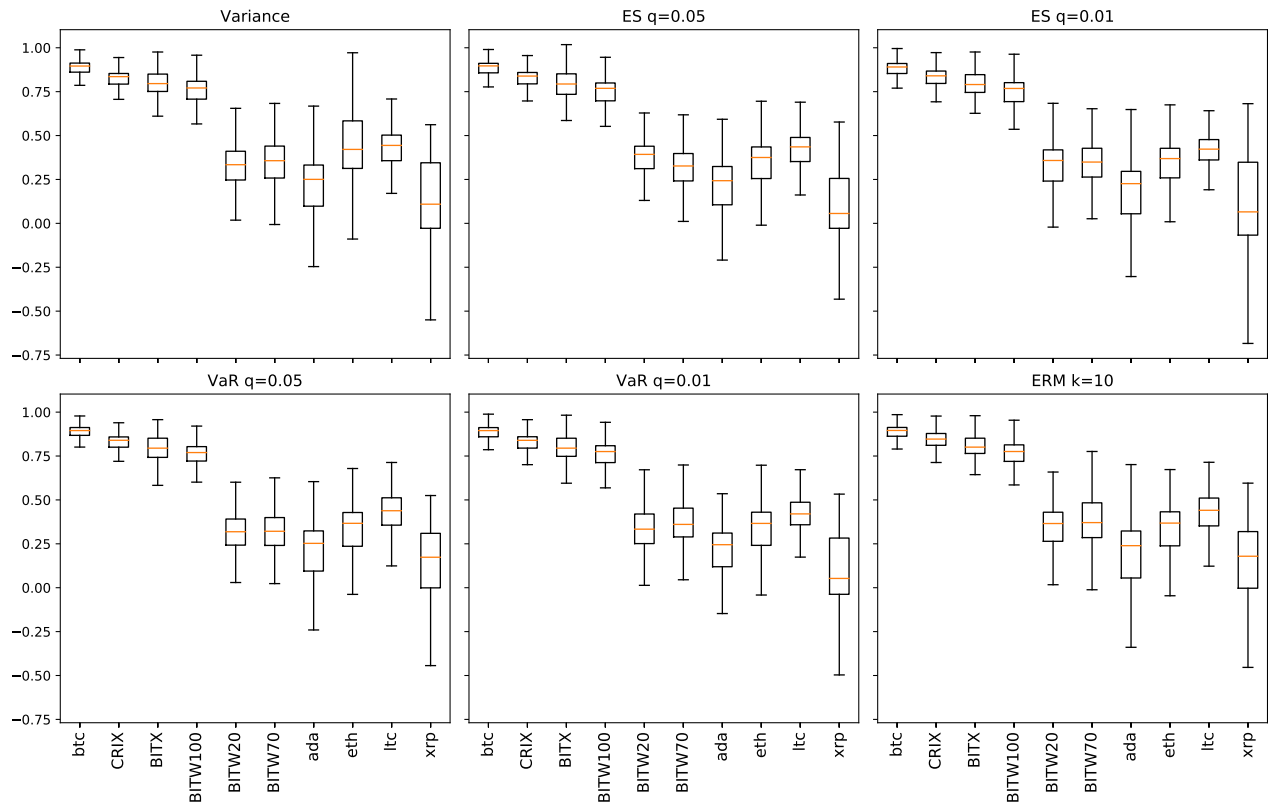
Figure 8: HE evaluated in the corresponding risk minimization objectives. The boxplots indicate the the median, upper quartile, lower quartile, minimium and maximum of the bootstrapped HE. The HE of BTC-involved spots are significantly higher than that of BTC-not-involved spots.

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