Notes on hedging cryptos with spectral risk measures

January 7, 2021

Abstract

We investigate different methods of hedging cryptocurrencies with Bitcoin futures. A useful generalisation of variance-based hedging uses spectral risk measures and copulas.

Contents

1. Optimal hedge ratio

Following (Barbi and Romagnoli, 2014), we consider the problem of optimal hedge ratios by extending the commmonly known minimum variance hedge ratio to more general risk measures and dependence structures.

Hedge portfolio: $R_t^h = R_t^S - hR_t^F$, involving returns of spot and future contract and where h is the hedge ratio
Optimal hedge ratio: $h^* = \text{argmin}_h\, \rho_\phi(s, h)$, for given confidence level $1 - s$ (if applicable, e.g. in the case of VaR, ES), where ρ_ϕ is a spectral risk measure with weighting function ϕ (see below).

The distribution function of R^h can be expressed in terms of the copula and the marginal distributions as Proposition 1 result shows (this is a corrected version of Corollary 2.1 of (Barbi and Romagnoli, 2014)). For practical applications, it is numerically faster and more stable to use additional information about the specific copula and marginal distributions. We therefore derive semi-analytic formulas for a number of special cases, such as the Gaussian-, Student t-, normal inverse Gaussian (NIG) and Archimedean copulas in Section 4.

Proposition 1. *Let R^S and R^F be two real-valued random variables on the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with corresponding absolutely continuous copula $C_{R^S,R^F}(w, \lambda)$ and continuous marginals F_{R^S} and F_{R^F}. Then, the distribution of of R^h is given by*

$$F_{R^h}(x) = 1 - \int_0^1 D_1 C_{R^S,R^F}\left(u, F_{R^F}\left(\frac{F_{R^S}^{-1}(u) - x}{h}\right)\right) \mathrm{d}u. \tag{1}$$

Here $D_1 C(u, v) = \dfrac{\partial}{\partial u} C(u, v)$, which is easily shown to fulfil:[1]

$$D_1 C_{X,Y}(F_X(x), F_Y(y)) = \mathbf{P}(Y \leq y | X = x). \tag{2}$$

[1]Let $F_X(x) = u$, $F_Y(y) = v$. Then, formally,

$$\begin{aligned}
\frac{\partial}{\partial F_X(x)} C(F_X(x), F_Y(y)) &= \frac{\partial}{\partial F_X(x)} \mathbf{P}(U \leq F_X(x), V \leq F_Y(y)) = \mathbf{P}(U \in \mathrm{d}F_X(x), V \leq F_Y(y)) \\
&= \mathbf{P}(V \leq F_Y(y) | U = F_X(x)) \cdot \mathbf{P}(U \in \mathrm{d}F_X(x)) = \mathbf{P}(Y \leq y | X = x) \cdot \mathbf{P}(U \in \mathrm{d}u) \\
&= \mathbf{P}(Y \leq y | X = x).
\end{aligned}$$

Proof. Using the identity (2) gives

$$F_{R^h}(x) = \mathbf{P}(R^s - hR^F \leq x) = \mathbb{E}\left[\mathbf{P}\left(R^F \geq \frac{R^S - x}{h}\Big| R^S\right)\right]$$

$$= 1 - \mathbb{E}\left[\mathbf{P}\left(R^F \leq \frac{R^S - x}{h}\Big| R^S\right)\right] = 1 - \int_0^1 D_1 C_{R^S,R^F}\left(u, F_{R^F}\left(\frac{F_{R^S}^{(-1)}(u) - x}{h}\right)\right) \mathrm{d}u.$$

□

2. Spectral risk measures

Spectral risk measure (Acerbi, 2002; Cotter and Dowd, 2006):

$$\rho_\phi = -\int_0^1 \phi(p)\, q_p \, \mathrm{d}p,$$

where q_p is the p-quantile of the return distribution and $\phi(s)$, $s \in [0, 1]$, is the so-called *risk aversion function*, a weighting function such that[2]

(i) $\phi(p) \geq 0$,

(ii) $\int_0^1 \phi(p)\, \mathrm{d}p = 1$,

(iii) $\phi'(p) \leq 0$.

Examples: VaR, ES
Replacing the last property with $\phi'(p) > 0$ rules out risk-neutral behaviour.
Spectral risk measures are coherent (Acerbi, 2002).

[2] Note that the treatment in (Acerbi, 2002) is measure-based and therefore slightly different.

2.1. Representation of spectral risk measures

To prevent numerical instabilities involving the quantile function, re-write spectral risk measures as follows:

- Integration by substitution: $\int_a^b g(\varphi(x))\,\varphi'(x)\,\mathrm{d}x = \int_{\varphi(a)}^{\varphi(b)} g(u)\,\mathrm{d}u.$
- Spectral risk measures: $-\int_0^1 \phi(p)\,F^{(-1)}(p)\,\mathrm{d}p$
- Set $\varphi(x) = F(x)$, $g(p) = \phi(p)\,F^{(-1)}(p)$.
- Then:

$$-\int_0^1 \phi(p)\,F^{(-1)}(p)\,\mathrm{d}p = -\int_{-\infty}^{\infty} \phi(F(x))\,x\,f(x)\,\mathrm{d}x.$$

2.2. Exponential spectral risk measures

- Choose exponential utiliy function: $U(x) = -\mathbf{e}^{-kx}$, where $k > 0$ is the Arrow-Pratt coefficient of absolute risk aversion (ARA).
- Coefficient of absolute risk aversion: $R_A(x) = -\dfrac{U''(x)}{U'(x)} = k$
- Coefficient of relative risk aversion: $R_R(x) = -\dfrac{xU''(x)}{U'(x)} = xk$
- Weighting function $\phi(p) = \lambda\mathbf{e}^{-k(1-p)}$, where λ is an unknown positive constant.
- Set $\lambda = \dfrac{k}{1-\mathbf{e}^{-k}}$ to satisfy normalisation.

- Exponential spectral risk measure:

$$\rho_\phi = \int_0^1 \phi(p)\, F^{(-1)}(p)\,\mathrm{d}p = \frac{k}{1-\mathbf{e}^{-k}} \int_0^1 \mathbf{e}^{-k(1-p)}\, F^{(-1)}(p)\,\mathrm{d}p.$$

  (If calculation of quantiles is a problem use change of variables above.)

- What exactly is the link between risk measure and utility? I think there is no direct link: the exponential risk measure is *inspired* by ARA utility.

3. D_1 Operator

Introduce copulas first

The D_1 operator is given as

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

In the context of the above notation, we obtain

$$\begin{aligned} D_1 C_{R^s,R^F}\{w, g(w)\} &= \mathbf{P}[R_F \le F_F^{(-1)}\{g(w)\}|R_s = F_S^{(-1)}(w)] = \mathbf{P}\{V \le g(w)|U = w\} \\ &= \frac{\mathbf{P}\{U \in \mathrm{d}w, V \le g(w)\}}{\mathbf{P}(U \in \mathrm{d}w)} = \mathbf{P}\{U \in \mathrm{d}w | V \le g(w)\} = \int_0^{g(w)} c(w,v)\,\mathrm{d}v. \end{aligned}$$

The last line can also be written as Which statement in the last line does the refer to?

$$\frac{\partial}{\partial w} C\{w, g(w')\}\big|_{w'=w}.$$

We give an explicit equation of the D_1 operator for Archimedean copulae.

The D_1 operator is defined as the partial derivatives of the first input to the copula function, so we fix the second argument while taking derivative with respect to the first, and then evaluate the function. we have

Function	Gumbel	Frank	Clayton	Independence
$\phi(t)$	$\{-\log(t)\}^{\theta}$	$-\ln\left\{\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right\}$	$\frac{1}{\theta}(t^{-\theta}-1)$	Same to Gumbel where $\theta=1$
$\phi^{-1}(t)$	$\exp(-t^{1/\theta})$	$\frac{-1}{\theta}\log[1+\exp(-t)\{\exp(-\theta)-1\}]$	$(1+\theta t)^{-\frac{1}{\theta}}$	
$\partial\phi(t)/\partial t$	$\theta\frac{\phi(t)}{t\log(t)}$	$\frac{\theta\exp(-\theta t)}{\exp(-\theta t)-1}$	$-t^{-(\theta+1)}$	
$\partial\phi^{-1}(t)/\partial t$	$\frac{-1}{\theta}t^{\frac{1}{\theta}-1}\phi^{-1}(t)$	$\frac{1}{\theta}\frac{\exp(-t)\{\exp(-\theta)-1\}}{1+\exp(-t)\{\exp(-\theta)-1\}}$	$\theta(1+\theta t)^{-\frac{1}{\theta}-1}$	

Table 1: Archemdean Copulae's Generator, Generator Inverse, and their derivative.

$$\left.\frac{\partial C\{v,g(w)\}}{\partial v}\right|_{v=w} = \left.\frac{\partial\phi^{-1}[\phi(v)+\phi\{g(w)\}]}{\partial[\phi(v)+\phi\{g(w)\}]}\frac{\partial[\phi(v)+\phi\{g(w)\}]}{\partial v}\right|_{v=w} \tag{3}$$

$$= \left.\frac{\partial\phi^{-1}[\phi(v)+\phi\{g(w)\}]}{\partial[\phi(v)+\phi\{g(w)\}]}\frac{\partial\phi(v)}{\partial v}\right|_{v=w} \tag{4}$$

$$= \frac{\partial\phi^{-1}[\phi(w)+\phi\{g(w)\}]}{\partial[\phi(w)+\phi\{g(w)\}]}\frac{\partial\phi(w)}{\partial w} \tag{5}$$

$$, \text{where } g(w) = F_{R^F}\left\{\frac{F^{-1}_{R^S}(w)-r^h}{h}\right\} \tag{6}$$

$$\tag{7}$$

4. Special copulas

Add basics about copulas, Sklar's Theorem

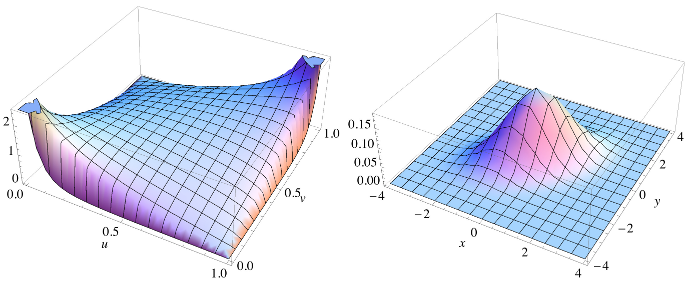

Figure 1: Left: Density of the Gaussian (Normal) copula. Right: Density of the bivariate Normal distribution ($\rho = 0.5$ in both cases).

Currently: Archimedean (Clayton, Gumbel), Gaussian, NVM / elliptical (Student t, NIG), NIG factor
Possibly: double t, Frank, skewed-t

4.1. Gaussian Copula

The **Gaussian copula**, also called **Normal copula**, is the copula generated by jointly normally distributed random variables (given here in bivariate form):

$$C_{\rho,\mathrm{N}}(u,v) := \mathrm{N}_2(\mathrm{N}^{(-1)}(u), \mathrm{N}^{(-1)}(v); \rho),$$

where N_2 and N are the bivariate and univariate normal distribution functions, respectively, and ρ is the correlation parameter.
The Gaussian or Normal copula is

$$C^{Ga}_{\Sigma}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \int_{-\infty}^{\Phi^{-1}(x_1)} \cdots \int_{-\infty}^{\Phi^{-1}(x_d)} \exp\left\{-\frac{1}{2} y^\top \Sigma^{-1} y\right\} dy_1 \ldots dy_d, \quad x \in \mathbb{R}^d. \tag{8}$$

The copula density is

$$c_{\Sigma}^{Ga}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} \Phi^{-1}(x_1) \\ \vdots \\ \Phi^{-1}(x_d) \end{pmatrix}^{\top} \Sigma^{-1} \begin{pmatrix} \Phi^{-1}(x_1) \\ \vdots \\ \Phi^{-1}(x_d) \end{pmatrix} \right\} \tag{9}$$

Simplified notation bivariate Gaussian copula

$$C_{\rho}^{Ga}\{w, g(w)\} = \Phi_{\rho}[\Phi^{-1}(w), \Phi^{-1}\{g(w)\}], \tag{10}$$

where $g(w) : [0,1] \mapsto \mathbb{R}$ is defined above, ρ is the dependency parameter of a bivariate Gaussian copula, Φ_{ρ} is bivariate normal distribution with mean 0 and covariance $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\Phi(\cdot)$ is CDF of standard normal, $\phi(\cdot)$ is PDF of standard normal, $\Phi^{-1}(\cdot)$ is quantile function of standard normal.
[I suggest to move the part below to the section where the formula for R^h is discussed. Also, there is a simpler version that requires only univariate function evluations.]
The bivariate $D_1 C^{Ga}\{w, g(w)\}$ is

$$D_1 C_{\rho}^{Ga}\{w, g(w)\} = \int_{-\infty}^{\Phi^{-1}\{g(w)\}} \phi_{\rho}\{\Phi^{-1}(w), u\} du \cdot \frac{1}{\phi\{\Phi^{-1}(w)\}} \tag{11}$$

Proof.

$$D_1 C_{\rho}\{w, g(w)\} = \left. \frac{\partial C_{\rho}\{w, g(w')\}}{\partial w} \right|_{w'=w} \tag{12}$$

$$= \left. \frac{\partial \Phi_{\rho}[\Phi^{-1}(w), \Phi^{-1}\{g(w)\}]}{\partial \Phi^{-1}(w)} \frac{\partial \Phi^{-1}(w)}{\partial w} \right|_{w'=w} \tag{13}$$

$$= \frac{1}{2\pi\rho} \int_{-\infty}^{\Phi^{-1}\{g(w)\}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \Phi^{-1}(w)^2 - 2\rho\Phi^{-1}(w)u + u^2 \right\} du \cdot \frac{1}{\phi\{\Phi^{-1}(w)\}} \tag{14}$$

□

Hedge distribution

Let $R^S \sim \mathrm{N}(\mu_S, \sigma_S^2)$ and $R^F \sim \mathrm{N}(\mu_F, \sigma_F^2)$ and assume further that they are jointly normally distributed with correlation ρ. Then,

$$R^h = R^S - hR^F \sim \mathrm{N}(\mu_S - h\mu_F, \sigma_s^2 + h^2\sigma_F^2 - 2\rho h\sigma_S\sigma_F).$$

More generally, if $R^k \sim F^k$, $k \in \{S, F\}$, then the distribution of R^h can be expressed with univariate expressions:

$$\begin{aligned}
\mathbf{P}(R^S - hR^F \leq x) &= 1 - \mathbb{E}\left[\mathbf{P}\left(R^F \leq \frac{R^S - x}{h} \big| R^S\right)\right] \\
&= 1 - \mathbb{E}\left[\mathbf{P}\left(\mathrm{N}^{(-1)}(F_F(R^F)) \leq \mathrm{N}^{(-1)}\left(F_F\left(\frac{R^S - x}{h}\right)\right)\big| R^S\right)\right] \\
&= 1 - \mathbb{E}\left[\mathbf{P}\left(\rho\mathrm{N}^{(-1)}(F_S(R^S)) + \sqrt{1-\rho^2}\varepsilon \leq \mathrm{N}^{(-1)}\left(F_F\left(\frac{R^S - x}{h}\right)\right)\big| R^S\right)\right] \\
&= 1 - \mathbb{E}\left[\mathrm{N}\left(\frac{\mathrm{N}^{(-1)}\left(F_F\left(\frac{R^S - x}{h}\right)\right) - \rho\mathrm{N}^{(-1)}(F_S(R^S))}{\sqrt{1-\rho^2}}\right)\right] \\
&= 1 - \int_0^1 \mathrm{N}\left(\frac{\mathrm{N}^{(-1)}\left(F_F\left(\frac{F_S^{(-1)}(u) - x}{h}\right)\right) - \rho\mathrm{N}^{(-1)}(u)}{\sqrt{1-\rho^2}}\right) \mathrm{d}u.
\end{aligned}$$

4.2. Archimedean copulas

- A well-studied one-parameter family of copulas are the **Archimedean copulas**.
- Let $\phi : [0,1] \to [0,\infty]$ be a continuous and strictly decreasing function with $\phi(1) = 0$ and $\phi(0) \leq \infty$.

- We define the **pseudo-inverse** of ϕ as

$$\phi^{(-1)}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) < t \leq \infty. \end{cases}$$

- If, in addition, ϕ is convex, then the following function is a copula:

$$C(u, v) = \phi^{(-1)}(\phi(u) + \phi(v)).$$

- Such copulas are called **Archimedean copulas**, and the function ϕ is called an **Archimedean copula generator**.
- Examples of Archimedean copulas are the **Gumbel** and the **Clayton** copulas:

$$\begin{aligned} C_{\theta,\mathrm{Gu}}(u, v) &= \exp\left\{-((-\ln u)^{\theta} + (-\ln v)^{\theta})^{1/\theta}\right\}, & 1 \leq \theta < \infty, \\ C_{\theta,\mathrm{Cl}}(u, v) &= (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, & 0 < \theta < \infty. \end{aligned}$$

- In the case of the Gumbel copula, the independence copula is attained when $\theta = 1$ and the comonotonicity copula is attained as $\theta \to \infty$.
- Thus, the Gumbel copula interpolates between independence and perfect dependence.
- In the case of the Clayton copula, the independence copula is attained as $\theta \to 0$, whereas the comonotonicity copula is attained as $\theta \to \infty$.

4.3. Elliptical Copulas and normal variance mixtures

See e.g. Theorem 3.22, Definition 3.26 and Theorem 3.28 of (McNeil *et al.*, 2005):

Definition 1. A random vector $\boldsymbol{Z} = (Z_0, \ldots, Z_d)^T$ follows an elliptical distribution if it has a representation

$$\boldsymbol{Z} \stackrel{\mathcal{L}}{=} GA\boldsymbol{U},$$

where $G > 0$ is a scalar random variable, the so-called *mixing variable*, A is a deterministic $(d+1) \times (d+1)$ matrix with $AA^T := \Sigma$, which in turn is a $(d+1) \times (d+1)$ nonnegative definite symmetric matrix of rank $d+1$, and $\boldsymbol{U}$ is a $(d+1)$-dimensional random vector uniformly distributed on the unit sphere $\mathcal{S}_{d+1} := \{\boldsymbol{z} \in \mathbb{R}^{d+1} : \boldsymbol{z}^T\boldsymbol{z} = 1\}$, and $\boldsymbol{U}$ is independent of G.

A subclass of elliptical distributions are the so-called *normal variance mixtures (NVM)*, see Section 3.3 of (McNeil *et al.*, 2005). For the connection between NVM and elliptical distributions, see also Theorem 3.25 of (McNeil *et al.*, 2005).

Definition 2 (Normal variance mixture (NVM))**.** The random vector $\mathbf{X} = (X_1, \ldots, X_k)^T$ follows a multivariate *normal variance mixture (NVM) distribution* if

$$\mathbf{X} \stackrel{\mathcal{L}}{=} \mu + \sqrt{W} A\mathbf{Z},$$

where

(i) $\mathbf{Z} \sim \mathrm{N}_k(\mathbf{0}, I_k)$, i.e., $\mathbf{Z}$ are independent, standard normally distributed,

(ii) $W \geq 0$ is a random variable independent of $\mathbf{Z}$,

(iii) $A \in \mathbb{R}^{d \times k}$ and $\mu \in R^d$ are a matrix and vector of constants, respectively.

It is easily observed that $\mathbf{X}|W = w \sim \mathrm{N}_d(\mu, w\Sigma)$, where $\sigma = AA'$.
In general, we will assume that Σ is positive definite and that $W > 0$ $\mathbf{P}$–a.s.. Then, the density of $\mathbf{X}$ is given by

$$\begin{aligned} f(\mathbf{x}) &= \int f_{\mathbf{X}|W}(\mathbf{x}|w)\,\mathrm{d}H(w) \\ &= \int \frac{w^{-d/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}{2w}\right) \mathrm{d}H(w), \end{aligned}$$

where H is the distribution function of W.
Special cases:

- Normal distribution: W constant
- Student t distribution: $W \sim Ig(1/2\nu, 1/2\nu)$, where Ig is an inverse gamma distribution
- Symmetric generalised hyperbolic distribution: $W \sim N^{-}(\lambda, \chi, \psi)$ where N^{-} refers to the generalised inverse Gaussian (GIG) distribution;
- Normal inverse Gaussian (NIG): W follows a GIG distribution with $\lambda = -0.5$.

Copulas are obtained from elliptical distributions via Sklar's theorem by transforming the margins to uniforms.

Hedge distribution

Let (R^S, R^F) follow a normal variance mixture, i.e., there exists a decomposition such that

$$\begin{aligned} R^S &= \mu_S + \sqrt{W}\sigma_S Z_1 \\ R^F &= \mu_F + \sqrt{W}\sigma_F(\rho Z_1 + \sqrt{1-\rho^2}Z_2), \end{aligned}$$

where W is the mixing variable and Z_1, Z_2 are independent standard normal variables. Then, R^h follows a NVM distribution with

$$R^h = R^S - hR^F = \mu_S - h\mu_F + \sqrt{W}\left((\sigma_S - h\sigma_F\rho)Z_1 - h\sqrt{1-\rho^2}\sigma_F Z_2\right) = \mu_S - h\mu_F + \sqrt{W}Z_3,$$

where $Z_3 \sim \mathrm{N}(0, \sigma_S^2 + h^2\sigma_F^2 - 2\rho h\sigma_S\sigma_F)$.
More generally, let $R^k \sim F^k$, $k \in \{S, F\}$ and write V as the marginal distribution functions of the NVM distribution components. Let $V^{(-1)}(F_S(R^S)) \overset{\mathcal{L}}{=} \sqrt{W}Z_1$ and $V^{(-1)}(F_F(R^F)) \overset{\mathcal{L}}{=} \sqrt{W}\rho Z_1 + \sqrt{W}\sqrt{1-\rho^2}Z_2$, were Z_1, Z_2 are

independent standard normals. Then

$$\begin{aligned}
\mathbf{P}(R^S - hR^F \le x) &= 1 - \mathbb{E}\left[\mathbf{P}\left(R^F \le \frac{R^S - x}{h} | R^S\right)\right] \\
&= 1 - \mathbb{E}\left[\mathbf{P}\left(V^{(-1)}(F_F(R^F)) \le V^{(-1)}\left(F_F\left(\frac{R^S - x}{h}\right)\right) \Big| W, Z_1\right)\right] \\
&= 1 - \mathbb{E}\left[\mathbf{P}\left(\sqrt{W}\rho Z_1 + \sqrt{W}\sqrt{1-\rho^2} Z_2 \le V^{(-1)}\left(F_F\left(\frac{F_S^{(-1)}(V(\sqrt{W}Z_1)) - x}{h}\right)\right) \Big| W, Z_1\right)\right] \\
&= 1 - \mathbb{E}\left[\mathrm{N}\left(\frac{V^{(-1)}\left(F_F\left(\frac{F_S^{(-1)}(V(\sqrt{W}Z_1)) - x}{h}\right)\right) - \rho\sqrt{W}Z_1}{\sqrt{W}\sqrt{1-\rho^2}}\right)\right] \\
&= 1 - \int_0^\infty \int_{-\infty}^\infty \mathrm{N}\left(\frac{V^{(-1)}\left(F_F\left(\frac{F_S^{(-1)}(V(\sqrt{w}z_1)) - x}{h}\right)\right) - \rho\sqrt{w}z_1}{\sqrt{w}\sqrt{1-\rho^2}}\right) \varphi(z_1)\, f_W(w)\, \mathrm{d}z_1\, \mathrm{d}w.
\end{aligned}$$

As in the Gaussian copula case, this expression contains evaluations of only univariate distribution function.

4.4. Normal inverse Gaussian factor copula model

4.4.1. Normal inverse Gaussian distribution

It was established in the previous section that a multivariate normal inverse Gaussian (NIG) distribution is a normal variance mixture. Here the random components have a joint scalar mixing variable. An additional copula model can be derived as a factor model from the NIG distribution: under certain conditions, its distribution type is preserved under linear combinations. This property, together with its infinite divisibility, allow for the construction of Lévy processes from NIG distributions.

Following (Barndorff-Nielsen, 1997), a normal inverse Gaussian (NIG) distribution 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 $0\le|\beta|\le\alpha$, $\mu\in\mathbb{R}$ 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

$$\mathbb{E}X = \mu+\frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} \tag{15}$$

$$\mathrm{Var}(X) = \frac{\alpha^2\delta}{(\alpha^2-\beta^2)^{3/2}}. \tag{16}$$

Let $\mathrm{IG}(\delta,\gamma)$ denote the *inverse Gamma distribution* with density function[3]

$$d(w;\delta,\gamma) = \frac{1}{\sqrt{2\pi}}\exp(\delta\gamma)w^{-3/2}\exp(-\frac{\delta^2/w+\gamma^2 z}{2}). \tag{17}$$

[3]The density of the IG distribution in `Mathematica` is given as

$$f(x)=\sqrt{\frac{\lambda}{x^3}}\frac{1}{\sqrt{2\pi}}\mathbf{e}^{-\frac{\lambda(x-\mu)^2}{2x\mu^2}}, \quad x>0,$$

with parameters $\mu=\delta/\gamma$ and $\lambda=\delta^2$.

The $\mathrm{NIG}(\alpha, \beta, \mu\, \delta)$ distribution is a normal variance-mean mixture: 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 $\mathrm{IG}(\delta, \sqrt{\alpha^2 - \beta^2})$ distribution.
It is easily seen from the moment-generating function that linear combinations of NIG random variables are again NIG-distributed provided they share the parameters α and β. Let $X_i \sim \mathrm{NIG}(\alpha, \beta, \mu_i, \delta_i)$, $i = 1, 2$, be independent NIG variables. Then,

$$\mathbb{E}\left[\mathbf{e}^{u(X_1+X_2)}\right] = \mathbb{E}\left[\mathbf{e}^{uX_1}\right] \mathbb{E}\left[\mathbf{e}^{uX_2}\right] = \exp\left((\delta_1 + \delta_2)\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2}\right) + (\mu_1 + \mu_2)u\right),$$

hence $X_1 + X_2 \sim \mathrm{NIG}(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$. (This is also a direct consequence from the properties of the normal inverse Gaussian Lévy process X_t, which may be represented as Brownian motion with a random time change,

$$X_t = B_{W_t} + \mu t,$$

where $B = (B_t)_{t\geq 0}$ is a Brownian motion and $W = (W_t)_{t\geq 0}$ is a Lévy process with density given by (17). The random variable W_t can be interpreted as a first-passage time of an independent Brownian motion $\overline{B}$, i.e., $W_t = \inf\{s > 0 : \overline{B}_s + \sqrt{\alpha^2 - \beta^2} s = \delta t\}$.)
As a consequence, the NIG distribution gives rise to two copulas:

- a copula determined from a linear combination of independent NIG random variables with identical parameters α, β (essentially a factor model);
- a copula determined from the multivariate normal-mean-variance mixture, which is a linear combination of normal random variables scaled by one scalar inverse Gaussian random variable.

4.4.2. NIG factor copula

We consider a simple factor model consisting of NIG-distributed random variables.

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

$$\begin{aligned} Cov(X_i, X_j) &= Var(Z), \\ Corr(X_i, X_j) &= \frac{\delta}{\sqrt{(\delta + \delta_i)(\delta + \delta_j)}}. \end{aligned} \tag{18}$$

Proof. (i) This follows directly from the moment-generating function.

(ii) 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 (16).

□

The NIG factor copula is obtained by transforming the margins to uniforms (see Sklar's Theorem).

4.4.3. Fitting the NIG factor model

In this section we assume that the marginal distributions are NIG.
Given that the parameters $\alpha, \beta, \mu, \delta$ denote tail heaviness, skewness, location and scale, one can fit the NIG distribution by matching the first four moments. Let $\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ denote the r-th moment of the sample $x_1, \ldots, x_n$. Then moment-matching the NIG distribution corresponds to solving for $\alpha, \beta, \mu, \delta$ the system of equations

$$\begin{pmatrix} \frac{\partial}{\partial u} M(u; \alpha, \beta, \mu, \delta)|_{u=0} \\ \frac{\partial^2}{\partial u^2} M(u; \alpha, \beta, \mu, \delta)|_{u=0} \\ \frac{\partial^3}{\partial u^3} M(u; \alpha, \beta, \mu, \delta)|_{u=0} \\ \frac{\partial^4}{\partial u^4} M(u; \alpha, \beta, \mu, \delta)|_{u=0} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \end{pmatrix}$$

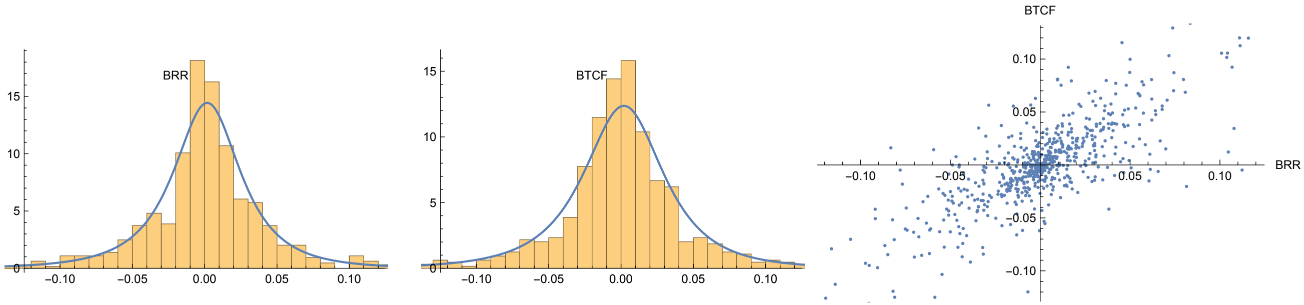

Figure 2: BRR and BTCF return distributions (empirical and fitted to NIG) as well as scatter plot.

Bivariate data, as present in our case, can be fit by moment-matching involving the parameters $(\alpha, \beta, \mu_1, \mu_2, \delta_1, \delta_2)$ and then fitting δ of the joint factor via the empirical correlation. Without loss of generality, we set $\mu = 0$. Then, $\alpha, \beta, \mu_i, \tilde{\delta}_i$, $i = 1, 2$, are determined from

$$\min_{\alpha,\beta,\mu_1,\mu_2,\tilde{\delta}_1,\tilde{\delta}_2} \sqrt{\sum_{k=1}^{2} \sum_{r=1}^{4} \left(\hat{\mu}_{r,k} - \frac{\partial^r}{\partial u^r} M(u; \alpha, \beta, \mu_k, \tilde{\delta}_k)\Big|_{u=0} \right)^2}.$$

Here, $\tilde{\delta}_k$ refers to the scale parameter of X_k. The scale parameters of the independent NIG components Z_k, $k = 1, 2$, are obtained as $\delta_k = \delta - \tilde{\delta}_k$ after solving δ from (18), which fixes the dependence between the margins. Figure 2 shows histograms and the fitted NIG densities of the Bitcoin reference rate (BRR) returns and the Bitcoin Futures contract returns (BTCF).

4.4.4. Fitting the NIG factor copula

If the margins are not NIG-distributed, then the NIG factor copula model is calibrated as follows. Denote the NIG distribution function by $F(\cdot;\alpha,\beta,\mu,\delta)$. Denote the margins by $U_i \sim U(0,1)$, $i=1,2$. The factor model of Proposition 2 is obtained by transforming the uniform margins to standardised NIG distributions by setting $X_i := F^{(-1)}(U_i;\alpha,\beta,\mu,\delta+\delta_i)$ with $\delta+\delta_i = \dfrac{(\alpha^2-\beta^2)^{3/2}}{\alpha^2}$ and $\mu = -\dfrac{(\delta+\delta_i)\beta}{\sqrt{\alpha^2-\beta^2}} = -\dfrac{(\alpha^2-\beta^2)\beta}{\alpha^2} = \dfrac{\beta^3}{\alpha^2}-\beta$. Here, δ refers to the scaling factor of the latent variable Z. In this case, $\mathbb{E}X_i = 0$ and $\mathrm{Var}(X_i) = 1$.
Denote the r-th empirical uncentered moment of i-th NIG-transformed margin by

$$\hat{\mu}_{i,r}(\alpha,\beta,\mu,\delta) := \frac{1}{n}\sum_{j=1}^{n} F^{(-1)}(u_{i,j};\alpha,\beta,\mu,\delta)^r, \tag{19}$$

Given the interpretation of the parameters – α as the tail heaviness, β as the degree of asymmetry, and δ determining the correlation – calibrating the bivariate factor model is the achieved by minimising the RMSE of the third and fourth moments and the correlation between the model and the data:

$$\begin{aligned}\min_{\alpha,\beta,\delta,\delta_1,\delta_2} &\left(\mathrm{Corr}(F^{(-1)}(U_1;\alpha,\beta,\mu,\delta+\delta_1), F^{(-1)}(U_2;\alpha,\beta,\mu,\delta+\delta_2) - \frac{\delta}{\sqrt{(\delta+\delta_1)(\delta+\delta_2)}}\right)^2 \\ &+ \sum_{i=1}^{2}\sum_{r=3}^{4}\left(\hat{\mu}_{i,r}(\alpha,\beta,\delta+\delta_i) - \frac{\partial^r}{\partial u^r} M(u;\alpha,\beta,\mu,\delta+\delta_i)\Big|_{u=0}\right)^2, \text{ subject to } \delta+\delta_i = \frac{(\alpha^2-\beta^2)^{3/2}}{\alpha^2},\end{aligned}$$

with $\mu = \dfrac{\beta^3}{\alpha^2} - \beta$.
This implies that $\delta_1 = \delta_2$. A standardised multivariate model would therefore be homogeneous with one correlation coefficient for all pairwise margins. Different correlations would be achieved by loosening the constraint of unit

variances. Re-writing, with $\tilde{\delta} := \delta + \delta_1 = \dfrac{(\alpha^2 - \beta^2)^{3/2}}{\alpha^2}$ gives

$$\min_{\alpha,\beta,\delta} \left(\mathrm{Corr}(F^{(-1)}(U_1; \alpha, \beta, \mu, \tilde{\delta}), F^{(-1)}(U_2; \alpha, \beta, \mu, \tilde{\delta}) - \frac{\delta}{\tilde{\delta}} \right)^2 + \sum_{i=1}^{2} \sum_{r=3}^{4} \left(\hat{\mu}_{i,r}(\alpha, \beta, \tilde{\delta}) - \frac{\partial^r}{\partial u^r} M(u; \alpha, \beta, \mu, \tilde{\delta}) \Big|_{u=0} \right)^2 .$$

4.4.5. NIG quantiles

Calibrating the model requires an efficient implementation of the NIG quantile function, see Equation (19. One way of implementing this is via the so-called *Cornish-Fisher expansion*.
CF expansion here.
Simulation produces better results but requires more computational power. Hence, use CF for calibration, which performs worse in the tails, and simulation in the hedge calculation / risk measure calculation. Use importance sampling.

4.4.6. Hedge calculation in the NIG factor model

Optimising for the hedge quantity h requires fast calculations of the hedge distribution function (1). The position obtained from hedging Bitcoin with h units of the Bitcoin future returns

$$R^h = X_1 - hX_2 = Z + Z_1 - hZ - hZ_2 = (1-h)Z + Z_1 - hZ_2,$$

with Z, Z_1, Z_2 independent NIG-distributed random variables. Without loss of generally, set $\mu = 0$. Because of the scaling with $1-h$ and h, R^h will not follow an NIG distribution, unless $h = 0$. A direct calculation of the distribution function is achieved by numerically computing the integral

$$\mathbf{P}(R^h \leq x) = \int_0^\infty \int_0^\infty \int_0^\infty \mathrm{N}(y; \mu_1 - h\mu_2 + \beta w_1 - h w_2, (1-h)^2 w + w_1 + h^2 w_2)\, f_W(w)\, f_{W_1}(w_1)\, f_{W_2}(w_2)\, \mathrm{d}w\, \mathrm{d}w_1\, \mathrm{d}w_2,$$

where $\mathrm{N}(x; \mu, \sigma^2)$ denotes the normal distribution function with expectation μ and variance σ^2, and f_W, f_{W_1}, f_{W_2} are the inverse gamma density functions of the scaling variables W, W_1 and W_2 with parameters $\delta, \delta_1, \delta_2$ and $\sqrt{\alpha^2 - \beta^2}$.

However, for determining the optimal hedge parameter h, numerical computation of the three-fold integral may be slow, compared to other methods that make explicit use of the simple form of the moment-generating function. The following two methods achieve a much faster computation for the NIG factor model. The first method uses Fourier inversion to calculate the density from the characteristic function. The second method approximates the distribution of R^h by an NIG distribution. Both methods use the moment-generating function of R^h, which is given by

$$\begin{aligned}\varphi_h(u) := \mathbb{E}\left[\mathbf{e}^{uR^h}\right] &= \mathbb{E}\left[\mathbf{e}^{u[(1-h)Z+Z_1-hZ_2]}\right] = \mathbb{E}\left[\mathbf{e}^{u(1-h)Z}\right]\mathbb{E}\left[\mathbf{e}^{uZ_1}\right]\mathbb{E}\left[\mathbf{e}^{-uhZ_2}\right] \\ &= M\left(u; \frac{\alpha}{1-h}, \frac{\beta}{1-h}, (1-h)\mu, (1-h)\delta\right) M\left(u; \alpha, \beta, \mu_1, \delta_1\right) M\left(-u; \frac{\alpha}{h}, \frac{\beta}{h}, h\mu_2, h\delta_2\right). \end{aligned} \tag{20}$$

Add copula stuff; here the assumption is still that R_S and R_F are NIG.

Calculation via Fourier inversion Because of the simple form of the moment-generating function (20), the density f of R^h can be calculated numerically using the inverse Fourier transform of the characteristic function $\varphi(u)$:

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathbf{e}^{-iux}\varphi_h(u)\,\mathrm{d}u,$$

and the distribution function of R^h can be calculated as

$$\mathbf{P}(R^h \leq x) = \lim_{b\to-\infty}\frac{1}{2\pi}\int_{\mathbb{R}}\frac{\mathbf{e}^{-iux}-\mathbf{e}^{-iub}}{iu}\,\varphi_h(u)\,\mathrm{d}u,$$

see e.g. II.§12, Theorem 3 of (Shiryaev, 1996).

NIG approximation of hedged position If $h = 1$ and $\beta = 0$, then R^h is NIG, hence, if h is close to 1, it may be feasible to approximate R^h by an NIG. The parameters can be determined either by moment-matching or by a first-order Taylor approximation of the moment-generating function's exponent.

For the moment-matching procedure, set $R^h \approx \mathrm{NIG}(\alpha_h, \beta_h, \mu_h, \delta_h)$ and solve for $\alpha_h, \beta_h, \mu_h, \delta_h$,

$$\begin{pmatrix} \frac{\partial}{\partial u}\varphi_h(u)\big|_{u=0} \\ \frac{\partial^2}{\partial u^2}\varphi_h(u)\big|_{u=0} \\ \frac{\partial^3}{\partial u^3}\varphi_h(u)\big|_{u=0} \\ \frac{\partial^4}{\partial u^4}\varphi_h(u)\big|_{u=0} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial u}M(u;\alpha_h,\beta_h,\mu_h,\delta_h)_{u=0} \\ \frac{\partial^1}{\partial u^2}M(u;\alpha_h,\beta_h,\mu_h,\delta_h)_{u=0} \\ \frac{\partial^2}{\partial u^3}M(u;\alpha_h,\beta_h,\mu_h,\delta_h)_{u=0} \\ \frac{\partial^3}{\partial u^4}M(u;\alpha_h,\beta_h,\mu_h,\delta_h)_{u=0} \end{pmatrix}.$$

An alternative to determine the parameters is to assume that $h = 1$ so that $R^1 = Z_1 - Z_2$ and using the following first-order Taylor approximation around zero:

$$\sqrt{\alpha^2 - (\beta+u)^2} - \sqrt{\alpha^2 - (\beta-u)^2} \approx -\frac{2\beta}{\sqrt{\alpha^2-\beta^2}}\,u.$$

This gives

$$\begin{aligned}
\varphi_1(u) &= M(u;\alpha,\beta,\mu_1,\delta_1)M(-u;\alpha,\beta,\mu_2,\delta_2) \\
&= \exp\left(\delta_1(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}) + \mu_1\,u + \delta_2(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta-u)^2}) - \mu_2\,u\right) \\
&= \exp\left(\delta_1(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}) + \mu_1\,u + \delta_2(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}) - \mu_2\,u\right) \\
&\cdot \exp\left(\delta_2(\sqrt{\alpha^2-(\beta+u)^2} - \sqrt{\alpha^2-(\beta-u)^2})\right) \\
&\approx \exp\left((\delta_1+\delta_2)(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}) + \left(\mu_1 - \mu_2 - \frac{2\delta_2\beta}{\sqrt{\alpha^2-\beta^2}}\right)u\right) \\
&= M\left(u;\alpha,\beta,\mu_1-\mu_2-\frac{2\delta_2\beta}{\sqrt{\alpha^2-\beta^2}},\delta_1+\delta_2\right).
\end{aligned}$$

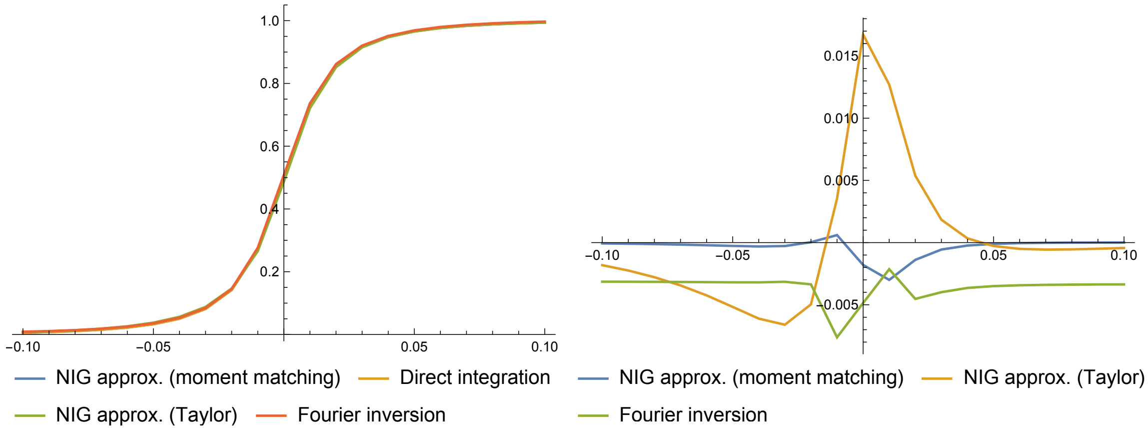

Figure 3: Left: CDF of $R^{0.95}$ using Bitcoin data. Right: Difference of different methods to direct integration. CPU times: NIG approximation (moment matching): 0.2 seconds plus 18.6 seconds for parameter calibration; Direct integration: 82 seconds; NIG approximation (Taylor): 0.18 seconds; Fourier inversion: 0.034 seconds plus 1.55 seconds to generate smooth density function.

Figure 3 shows an example of the distribution of $R^{0.95}$ for the different approximations as well as their error relative to direct integration. It turns out that the NIG moment-matching approximation performs best in terms of the error, while the Taylor approximation performs worst. Taking into account CPU times, the Fourier inversion technique performs best in terms of balancing error and CPU time.

[The definition below is one way of introducing the elliptical copula, but not the most practical one for our purposes.]

Definition 3. Elliptical Distribution. The d-dimensional random vector $\boldsymbol{y}$ has an elliptical distribution if and only if the characteristic function $\boldsymbol{t} \mapsto \mathbb{E}\{\exp(i\boldsymbol{t}^\top \boldsymbol{y})\}$ with $\boldsymbol{t} \in \mathbb{R}^d$ has the representation

$$\phi_g(\boldsymbol{t}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \exp(i\boldsymbol{t}^\top \boldsymbol{\mu}) g(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}; \boldsymbol{\nu}) \tag{21}$$

where $g(\cdot;\nu) : [0, \infty[\mapsto \mathbb{R}$, $\nu \in \mathbb{R}^d$, and Σ is a symmetric positive semidefinite $d \times d$-matrix.

If r has a density, then the density of $\boldsymbol{y}$ is of the form

$$|\Sigma|^{\frac{1}{2}} g\{(\boldsymbol{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})\}. \tag{22}$$

The function $g(\cdot;\nu)$ is known as characteristic generator, whereas $\boldsymbol{\nu}$ is parameter that determines the shape, in particular the tai index of the distribution.

Corollary 3. *(?, equation 2.12) If $\boldsymbol{y}$ follows an elliptical distribution, then $\boldsymbol{y}$ has a stochastic representation*

$$\boldsymbol{y} = \boldsymbol{\mu} + r\boldsymbol{A}^\top \boldsymbol{u}, \tag{23}$$

where $r \in \mathbb{R}_+$ is independent of $\boldsymbol{u}$, and $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{\Sigma}$.

Distribution	$r \sim$	$g(\boldsymbol{t})$
Gaussian	χ_n	

Table 2: Generators of Elliptical Distributions summarised from (?, Chapter 2)

4.5. t-copulae

The t copula is to represent the dependency structure by t distribution (??). ? extend this idea to skewed t copula and grouped t copula to allow more flexibility in the modelling of dependency structure.

4.5.1. Vanilla t-copula

The t-copula is

$$C^t_{\nu,\Sigma}(x) = \int_{-\infty}^{t_\nu^{-1}(x_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(x_n)} \frac{\Gamma\left\{\frac{\nu+i}{2}\right\}}{\Gamma\left\{\frac{\nu}{2}\right\}(\pi\nu)^{i/2}|\Sigma|^{1/2}} \left(1 + \frac{y^\top \Sigma^{-1} y}{\nu}\right)^{-\frac{\nu+i}{2}} dy_1 \dots dy_n, \tag{24}$$

where t_ν^{-1} is the quantile function of a univariate student-t distribution with degree of freedom ν.

4.5.2. Skewed t copula

Mean variance mixture

4.5.3. Double-t copula

[It is OK to introduce the copula without any reference to CDO's.] **?** present an alternative way to the Gaussian copula for valuing CDO tranches. The double-t copula model is a weighted sum of a common (or market) variable M and a idiosyncratic variable Z_i. *[They are t-distributed, right?]* The double-t copula is

$$X_i = w_i M + \sqrt{1 - w_i^2} Z_i \tag{25}$$

where M and Z_i are independent random variables with zero mean and unit variance, and X_i is an indicator variable for i^{th} asset. The authors map the time to default of the i^{th} obligor, t_i, to X_i,

$$F_{X_i}(x) = F_{t_i}(t). \tag{26}$$

In our case, we map X_i to log-returns of portfolio constituents,

$$F_{X_1}(x) = F_{r^S}(s) \text{ and } F_{X_2}(x) = F_{r^F}(t). \tag{27}$$

This is also known as percentile-to-percentile mapping(**?**).

[The percentile-to-percentile mapping is just the property that applying the cdf to a random variable yields a $U(0,1)$ variable:

$$F(x) = \mathbf{P}(X \leq x) = \mathbf{P}(F(X) \leq F(x)) = \mathbf{P}(U \leq F(x)),$$

since by definition, $U \sim U(0,1)$ fulfills $\mathbf{P}(U \leq u) = u$, $0 \leq u \leq 1$. This could be introduced when copulas and Sklar's Theorem are introduced.] The reason for this mapping is to turn incomprehensible dependency structures into known structure.

4.5.4. Normal Inverse Gaussian Copula

Normal Inverse Gaussian (NIG) distribution is a flexible 4-parameter distribution that can produce fat tails and skewness, unlike student-t distribution, NIG's convolution is stable under certain conditions and the CDF, PDF and quantile function can still be computed sufficiently fast (**?**, chapter 5). NIG distribution is a mixture of normal and inverse Gaussian distribution.

Definition 4. Inverse Gaussian Distribution. A non-negative random variable Y has an Inverse Gaussian (IG) distribution with parameters $\alpha > 0$ and $\beta > 0$ if its density funcion is of the form

$$f_{\text{IG}}(y; \alpha, \beta) = \frac{\alpha}{\sqrt{2\pi\beta}} y^{-1.5} \exp\left\{ -\frac{(\alpha - \beta z)^2}{2\beta z} \right\} \tag{28}$$

The corresponding distribution function is:

$$F_{\text{IG}}(y; \alpha, \beta) = \frac{\alpha}{\sqrt{2\pi\beta}} \int_0^y z^{-1.5} \exp\left\{ -\frac{(\alpha - \beta z)^2}{2\beta z} \right\} dz. \tag{29}$$

We write $Y \sim \text{IG}(\alpha, \beta)$.

Definition 5. Normal Inverse Gaussian Distribution. A random variable X has an Normal Inverse Gaussian (NIG) distribution with parameters α, β, μ and δ if its density funcion is of the form

$$X|Y = y \sim \Phi(\mu + \beta y, y) \tag{30}$$

$$Y \sim \text{IG}(\delta\gamma, \gamma^2) \text{ with } \gamma \stackrel{\text{def}}{=} \sqrt{\alpha^2 - \beta^2} \tag{31}$$

The corresponding distribution function is:

$$F_{\mathrm{NIG}}(y; \alpha, \beta) = \frac{\alpha}{\sqrt{2\pi\beta}} \int_0^y z^{-1.5} \exp\left\{-\frac{(\alpha - \beta z)^2}{2\beta z}\right\} dz. \tag{32}$$

5. Estimation

5.1. Two-Stage Estimation

? study the efficiency of a two-stage estimation procedure of copula estimation. The authors also call this method inference function for margins IFM.

Pros

1. Almost as efficient as MLE methods but easier to be implemented
2. Yields an asymptotically Gaussian, unbiased estimate

Cons

1. Subject to specification of marginals **?**

Our data

$$\boldsymbol{y} = \begin{bmatrix} y_{11} & \cdots & y_{1i} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{ni} \end{bmatrix} \tag{33}$$

Let F and f be the joint cdf and joint density of $\boldsymbol{y}$ with parameters $\boldsymbol{\delta}$, and let F_i and f_i be the marginal cdf and marginal density for the i^{th} random variable with parameters $\boldsymbol{\theta}_i$, we have

$$f(\boldsymbol{y};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots\boldsymbol{\theta}_i,\boldsymbol{\delta}) = c\{F_1(\boldsymbol{y}_1;\boldsymbol{\theta}_1),F_2(\boldsymbol{y}_2;\boldsymbol{\theta}_2),\ldots,F_i(\boldsymbol{y}_1;\boldsymbol{\theta}_i);\boldsymbol{\delta}\}\prod_{j=1}^{i} f_i(\boldsymbol{y}_j;\boldsymbol{\theta}_j) \tag{34}$$

For a sample of size n, the log-likelihood of functions of the i^{th} univariate margin is

$$L_i(\theta_i) = \sum_{m=1}^{n} \log f_i(y_{mi};\theta_i), \tag{35}$$

and the log-likelihood function for the joint distribution is

$$L(\delta, \theta_1, \theta_2, \ldots, \theta_i) = \sum_{m=1}^{n} \sum_{j=1}^{i} \log f(y_{mj}; \delta, \theta_1, \theta_2, ..., \theta_i) \tag{36}$$

In most cases, one does not have closed form estimators and numerical techniques are needed. Numerical ML estimation difficulty increase when the total number of parameters increases. The two-stage estimation is designed to overcome this problem.
The two-stage procedure is

1. estimate the univariate parameters from separate univariate likelihoods to get $\tilde{\boldsymbol{\theta}}_1, ..., \tilde{\boldsymbol{\theta}}_i$
2. maximize $L(\boldsymbol{\delta}, \tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_i)$ over $\boldsymbol{\delta}$ to get $\tilde{\boldsymbol{\delta}}$

Under regularity conditions [4] , $(\tilde{\boldsymbol{\theta}}_1, \ldots \tilde{\boldsymbol{\theta}}_i, \tilde{\boldsymbol{\delta}})$ is the solution of

$$(\partial L_1 / \partial \boldsymbol{\theta}_1^{\intercal}, \ldots, \partial L_i / \partial \boldsymbol{\theta}_i^{\intercal}, \partial L / \partial \boldsymbol{\delta}_1^{\intercal}) = \mathbf{0} \tag{37}$$

For comparison, if we optimize L directly without the two-stage procedure (i.e. MLE), we solve for

$$(\partial L / \partial \boldsymbol{\theta}_1^{\intercal}, \ldots, \partial L / \partial \boldsymbol{\theta}_i^{\intercal}, \partial L / \partial \boldsymbol{\delta}_1^{\intercal}) = \mathbf{0} \tag{38}$$

We denote the two solutions as $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\theta}}_1, \ldots \tilde{\boldsymbol{\theta}}_i, \tilde{\boldsymbol{\delta}})$ for two-stage procedure; $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\theta}}_1, \ldots \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\delta}})$ for MLE procedure. and compare the asymptotic relative efficiency of $\tilde{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\eta}}$.
Asymptotics: yet to be done.
? show the estimation of $\boldsymbol{\theta}$ may be seriously affected. They compare the two-stage approach and Canonical Maximum Likelihood Method by simulation and conclude that Canonical Maximum Likelihood is prefered from a computational statistics and data analysis point of view.

[4] Regularity conditions include 1. $\exists \frac{\partial \log f(x;\theta)}{\partial \theta}, \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}, \frac{\partial^3 \log f(x;\theta)}{\partial \theta^3}$ for all x; 2. $\exists g(x), h(x) and H(x)$ such that for θ in a neighborhood $N(\theta_0)$ the relations $\left|\frac{\partial f(x;\theta)}{\partial theta}\right| \leq g(x)$, $\left|\frac{\partial^2 f(x;\theta)}{\partial \theta^2}\right| \leq h(x)$, $\left|\frac{\partial^3 f(x;\theta)}{\partial \theta^3}\right| \leq H(x)$ hold for all x, and $\int g(x)dx < \infty$, $\int h(x)dx < \infty$, $\mathbb{E}_\theta\{H(X)\} < \infty$ for $\theta \in N(\theta_0)$; 3. For each $\theta \in \Theta$, $0 < \mathbb{E}_\theta \left\{ \left(\frac{\partial \log f(X;\theta)}{\partial \theta} \right)^2 \right\}$. For detail see section 4.2.2 of **?**

5.2. Canonical Maximum Likelihood Method

This approach was studied by **?** and **?**. One estimates the margins using empirical CDF

$$F_X(x) = \frac{1}{n+1} \sum_{i=1}^{n} 1(X_i \leq x) \tag{39}$$

,
we maximize the log-likelihood

$$L(\delta) = \sum_{i=1}^{n} \log[c_\delta\{F_X(X_i), F_Y(Y_i)\}] \tag{40}$$

This procedure does not require specification of marginals.

References

Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.

Barbi, M. and S. Romagnoli. A copula-based quantile risk measure approach to estimate the optimal hedge ratio. *Journal of Futures Markets*, 34(7):658–675, 2014.

Barndorff-Nielsen, O. E. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, 24(1):1–13, 1997.

Cotter, J. and K. Dowd. Extreme spectral risk measures: an application to futures clearinghouse margin requirements. *Journal of Banking & Finance*, 30(12):3469–3485, 2006.

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

Shiryaev, A. N. *Probability*. Springer, Berlin, 2nd edition, 1996.