Swiss Finance Institute

Research Paper Series

N°21-73

Heterogeneous Tail Generalized Common
Factor Modeling

Simon Hediger

University of Zurich

Jeffrey Näf

ETH Zurich

Marc S. Paoletta

University of Zurich and Swiss Finance Institute

Pawel Polak

Stony Brook University

Heterogeneous Tail Generalized Common Factor Modeling

Simon Hediger^{a*} Jeffrey Näf^b Marc S. Paoletta^{a,c} Paweł Polak^d

^a*Department of Banking and Finance, University of Zurich, Switzerland*

^b*Seminar for Statistics, ETH Zürich, Switzerland*

^c*Swiss Finance Institute*

^d*Department of Applied Mathematics and Statistics, Stony Brook University, United States*

October 27, 2021

Abstract

A multivariate normal mean-variance heterogeneous tails mixture distribution is proposed for the joint distribution of financial factors and asset returns (referred to as Factor-HGH). The proposed latent variable model incorporates a Cholesky decomposition of the dispersion matrix to ensure a rich dependency structure for capturing the stylized facts of the data. It generalizes several existing model structures, with or without financial factors. It is further applicable in large dimensions due to a fast ECME estimation algorithm of all the model parameters. The advantages of modelling financial factors and asset returns jointly under non-Gaussian errors are illustrated in an empirical comparison study between the proposed Factor-HGH model and classical financial factor models. While the results for the Fama-French 49 industry portfolios are in line with Gaussian-based models, in the case of highly tail heterogeneous cryptocurrencies, the portfolio based on the Factor-HGH model doubles the average return while keeping the volatility, the maximum drawdown, the turnover, and the expected-shortfall at a low level.

Keywords: Asset Pricing Model, Cryptocurrencies, Expectation Maximization Algorithm, Heterogeneous Tails, Mixture Distribution, Portfolio Optimization

1 Introduction

In explaining the behavior of stock returns, common risk factor models are well established in the field of asset pricing. The impetus came from the famous CAPM-model of [Treynor \(1961\)](#), [Sharpe \(1964\)](#), [Lintner \(1965\)](#), and [Mossin \(1966\)](#), which was later extended to a three-factor, a four-factor, and a five-factor model by [Fama and French \(1993\)](#), [Carhart \(1997\)](#) and [Fama and French \(2015\)](#), respectively. The momentum factor, which extended the three-factor model to the Carhart four-factor model, was first introduced and

*Corresponding author at: Department of Banking and Finance, University of Zurich, Switzerland.

E-mail address: simon.hediger@uzh.ch (S. Hediger).

analysed by [Jegadeesh and Titman \(1993\)](#). Nowadays, a zoo of factors is available (see, e.g., [Feng et al., 2020](#), and the references therein) with a large number of model variants. While the five factors of [Fama and French \(2015\)](#), and the momentum factor, remain among the most important common risk factors, a large literature now exists on how to decide whether or not to include a particular factor among the dozens, if not hundreds, available: See, e.g., [Bai and Ng \(2002\)](#), [Stock and Watson \(2002\)](#), [Tsai and Tsay \(2010\)](#), [Bai and Ng \(2013\)](#), [Bai and Liao \(2016\)](#), and the references therein.

One of the most common assumptions shared across papers in the financial risk factor literature regards the statistical distributional assumption of the financial asset returns. In particular, the assumption of Gaussianity is nearly ubiquitous. This is tenable for returns measured at the monthly level, but for daily returns (and higher frequency), the evidence for non-Gaussianity of asset returns is strong and well established; see for example [Pagan \(1996\)](#), [Chicheportiche and Bouchaud \(2012\)](#), [McNeil et al. \(2015\)](#), and the references therein. Moreover, and of great relevance for our task at hand, some models proposed in the literature that account for the non-elliptic nature of the returns data outperform their Gaussian-based counterparts in terms of out-of-sample portfolio performance (in a variety of measures, notably lower risk and higher return). Examples include [Paolella and Polak \(2015b\)](#), in which the joint distribution of asset returns is allowed to follow a multivariate non-elliptic generalized hyperbolic distribution; and [Paolella et al. \(2019\)](#), in which a two-component Markov switching structure, each governed by symmetric generalized hyperbolic distributions (and such that the unconditional distribution is non-elliptic) is used. A recent example of a non-Gaussian distributional assumption in a factor asset pricing model is [Kan and Zhou \(2017\)](#), in which they propose use of a multivariate Student's t . This distributional assumption is nested within our more general framework presented herein. While arguably a better empirical choice of distribution than the Gaussian, the use of the Student's t has the disadvantage in some applications of not possessing a moment generating function; see, e.g., [Paolella and Polak \(2015b\)](#), and the references therein. Further, it necessarily imposes the same tail behavior of each constituent component, this being highly unrealistic, and also one of the reasons for popularity of copula-based models; see [Marinelli et al. \(2012\)](#), [Paolella and Polak \(2015a, 2018\)](#), [Näf et al. \(2019\)](#), and the references therein. Other recent examples of using non-Gaussian distributions within the context of factor models in asset pricing include [Chung et al. \(2006\)](#), [Zhou and Li \(2016\)](#), and [Bao et al. \(2018\)](#).

The use of factor models has also emerged in the realm of cryptocurrencies. A Crypto-CAPM model and a three-factor model in the spirit of [Fama and French \(1993\)](#) was proposed in [Shen et al. \(2020\)](#). More recently, [Liu and Tsyvinski \(2021\)](#) look at a broader range of cryptocurrency-specific factors and their predictive performance. The common stylized facts of financial asset returns (e.g., heavy tails, volatility clustering, and non-ellipticity) are equally present in cryptocurrencies; see [Zhang et al. \(2018\)](#). Additionally, [Hu et al. \(2019\)](#) find that cryptocurrency portfolios constructed via optimizations that minimize variance and expected shortfall outperform a major stock market index (the S&P 500).

In this paper, we focus on a statistically interpretable factor model that is able to incorporate stylized facts of asset returns. Factors and asset returns are modeled jointly based on a model that incorporates heavy-tails and heterogeneous tail behavior, namely the so-called HGH-model (Heterogeneous Generalized Hyperbolic) of [Näf et al. \(2019\)](#). The modeling structure in [Näf et al. \(2019\)](#) avoids the use of copula, and in so doing, is able to derive exact distribution theory, obviating the need for simulation. As those authors

show, the model parameters associated with the resulting multivariate non-elliptical distribution can be estimated (quickly and efficiently no less, and thus suitable for large dimensions) by an ECME algorithm. This in turn permits computation of portfolio weights (using expected shortfall as the risk measure), suitable for medium-sized dimension, e.g., on the order of, say, 100.

The use of factor models in areas such as covariance matrix estimation, risk management applications, portfolio selection, and trading strategies has become essentially standard now; and given the ever-increasing set of available factors (e.g., via sentiment data), one cannot imagine a model without their use. In this paper, we focus on portfolio optimization. More precisely, we apply the proposed Factor-HGH-model in portfolio optimization on 49 value-weighted industry portfolio returns and six common factors, including momentum, from the [Kenneth R. French homepage](#). Further, we conduct portfolio optimization in a universe of 42 cryptocurrencies, in which the market factor is taken to be the only predictor. This approach is inspired by [Shen et al. \(2020\)](#) and [Brauneis and Mestel \(2019\)](#). While the results for the classical 49 industry portfolios are in line with Gaussian-based models, in the case of the highly tail heterogeneous cryptocurrencies, the portfolio based on the Factor-HGH model doubles the average return while keeping the volatility, the maximum drawdown, the turnover, and the expected-shortfall at a low level. These results illustrate the importance of non-Gaussian tail-heterogeneous generalized common factor modelling in this emerging asset class.

Perhaps stating the obvious, in recent years, machine learning is playing an ever-increasing role in finance. In particular, deep learning offers a promising alternative to standard financial models; see the results in [Heaton et al. \(2017\)](#). However, such optimism is tempered by the results in [Gu et al. \(2020\)](#), where one sees that the performance difference between linear factor models and use of non-linear machine learning tools is not substantially large. It is important to note that [Gu et al. \(2020\)](#) assume a non-stock-specific model structure, and use stock-specific characteristics instead of common factors as predictor variables. Further, their sample universe contains a wide variety of different companies, from potentially short-living micro-stocks, up to well-established large stocks. The difference in performance between classical models and sophisticated machine learning methods is even smaller when using stock-specific models, and focusing only on long-living companies; see [De Nard et al. \(2020\)](#).

Compared to standard machine learning methods as in [Gu et al. \(2020\)](#), which build the portfolios based on the raw excess return predictions, we focus on the prediction of the *joint* multivariate distribution of the factors and the returns of the assets in the portfolio. Moreover, all of this probabilistic information is used in building the portfolio through the minimization of the expected shortfall (ES). Modelling factors and asset returns jointly is not entirely new: See [Pourahmadi \(1999\)](#) for what appears to be the first contribution in this regard; and [Darolles et al. \(2018\)](#) for an extended version.

The paper proceeds as follows. Section 2 presents the new model in full generality, explains the parameter estimation procedure, and connects the new model to the common ones. Section 3 focuses on how the model, despite its seeming complexity, can straightforwardly be used to generate portfolio weights. Section 4 details an empirical investigation, showcasing the benefit of jointly modeling factors and asset returns, and doing so under a non-Gaussian distribution assumption. Section 5 introduces a model extension. Section 6 concludes. Appendix A contains details on model estimation. Appendix B provides proofs of two theoretical results deployed in the paper. Finally, Appendix C presents further performance tables.

2 Model

Let $\mathbf{f}_t$ and $\mathbf{r}_t$ denote the value of K_f factors, and excess simple returns of K_r assets, at time $t = 1, \dots, T$, respectively. We consider a joint multivariate model for $\mathbf{Z}_t = [\mathbf{f}_t^\top, \mathbf{r}_t^\top]^\top$ given by

$$\mathbf{Z}_t = \boldsymbol{\mu} + \mathbf{L}\mathbf{v}_t, \quad (1)$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(f)} \\ \boldsymbol{\mu}^{(r)} \end{bmatrix}, \mathbf{L} = \left[\begin{array}{c|c} \mathbf{L}^{(f)} & \mathbf{0} \\ \hline \mathbf{B} & \mathbf{L}^{(r)} \end{array} \right], \mathbf{v}_t = \begin{bmatrix} \mathbf{v}_t^{(f)} \\ \mathbf{v}_t^{(r)} \end{bmatrix} = \mathbf{C}^{1/2}\boldsymbol{\epsilon}_t, \text{ and } \mathbf{C} = \left[\begin{array}{c|c} \mathbf{C}^{(f)} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C}^{(r)} \end{array} \right],$$

each component of which is now defined. Vectors $\boldsymbol{\mu}^{(f)}$ and $\boldsymbol{\mu}^{(r)}$ correspond to the expected factor values and expected returns, respectively. Matrices $\mathbf{L}^{(f)}$, $\mathbf{L}^{(r)}$, and $\mathbf{C}^{(f)}$, $\mathbf{C}^{(r)}$, are $K_f \times K_f$, and $K_r \times K_r$, respectively, defined as:

$$\mathbf{L}^{(\cdot)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ q_{21}^{(\cdot)} & 1 & 0 & \dots & 0 \\ q_{31}^{(\cdot)} & q_{32}^{(\cdot)} & 1 & \dots & 0 \\ & \ddots & \ddots & \ddots & 0 \\ q_{K(\cdot)1}^{(\cdot)} & q_{K(\cdot)2}^{(\cdot)} & \dots & q_{K(\cdot)K(\cdot)-1}^{(\cdot)} & 1 \end{pmatrix}, \quad \mathbf{C}^{(\cdot)} = \begin{pmatrix} c_1^{(\cdot)} & & \\ & \ddots & \\ & & c_{K(\cdot)}^{(\cdot)} \end{pmatrix}, \quad (2)$$

with each $\mathbf{C}^{(\cdot)}$ being diagonal with elements $c_i^{(\cdot)} > 0$, for $i = 1, \dots, K(\cdot)$. Matrix $\mathbf{B}$ is $K_r \times K_f$, given by:

$$\mathbf{B} = \begin{pmatrix} b_{1,1} & \dots & b_{1,K_f} \\ \vdots & \ddots & \vdots \\ b_{K_r,1} & \dots & b_{K_r,K_f} \end{pmatrix}. \quad (3)$$

Finally,

$$\boldsymbol{\epsilon}_t = [\boldsymbol{\epsilon}_t^{(f)\top}, \boldsymbol{\epsilon}_t^{(r)\top}]^\top = [\epsilon_{1,t}^{(f)}, \dots, \epsilon_{K_f,t}^{(f)}, \epsilon_{1,t}^{(r)}, \dots, \epsilon_{K_r,t}^{(r)}]^\top$$

denotes the error term, these being mean zero and independent and identically distributed (iid) over time. In this paper, we choose to restrict $\mathbf{C}$ to be time-invariant, in order to concentrate on the efficacy of the new model and its features, without introducing a yet further generalization. However, all models discussed here may incorporate GARCH dynamics into $\mathbf{C}$ to allow for time-varying conditional variance of the error term (see also the Appendix in [Näf et al., 2019](#)).

First consider model (1) with iid Gaussian errors, i.e., $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I})$ (referred to as Gaussian-Cholesky). This is in fact a special case of the model in [Darolles et al. \(2018\)](#). Maximum likelihood estimation of the parameters in this Gaussian special case of our model can be easily achieved in two steps, as detailed in Appendix A.1. It is worth noting that we can allow for sparse $\mathbf{L}$, by incorporating ℓ_1 regularization into the estimation. This generalization corresponds to a maximum a posteriori estimation with Laplace prior ([Murphy, 2012](#)).

If we further impose $\mathbf{L}^{(r)} = \mathbf{I}$, then the model in (1) simplifies, and corresponds to a “standard” financial factor model with time series regression of the returns on the factors, and normally distributed error term

$$r_{i,t} = \beta_{i,0} + \boldsymbol{\beta}_i^\top \mathbf{f}_t + \epsilon_{i,t}, \quad (4)$$

where $\boldsymbol{\beta}_i = [\beta_{i,1}, \dots, \beta_{i,K_f}]^\top$, and $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ iid over t , for $i = 1, \dots, K_r$, as derived by [Ross \(1976, 1977\)](#) using the Arbitrage Pricing Theory (APT); and by [Chamberlain and Rothschild \(1983\)](#) in a large economy setting. Indeed, for $\mathbf{L}^{(r)} = \mathbf{I}$, the model in (1) reduces to

$$\begin{aligned} \mathbf{r}_t &= \boldsymbol{\mu}^{(r)} + \mathbf{B} \mathbf{v}_t^{(f)} + \mathbf{v}_t^{(r)} \\ &= \left(\boldsymbol{\mu}^{(r)} - \mathbf{B}(\mathbf{L}^{(f)})^{-1} \boldsymbol{\mu}^{(f)} \right) + \mathbf{B}(\mathbf{L}^{(f)})^{-1} \mathbf{f}_t + \mathbf{v}_t^{(r)}. \end{aligned} \quad (5)$$

Since by assumption $\mathbf{v}_t^{(r)} = (\mathbf{C}^{(r)})^{1/2} \boldsymbol{\epsilon}_t^{(r)} \sim N(\mathbf{0}, \mathbf{C}^{(r)})$, iid over t , with $\mathbf{C}^{(r)}$ being a diagonal matrix, it follows that the regressions obtained in (4) and (5) are equivalent, with

$$\beta_{i,0} = \left[\boldsymbol{\mu}^{(r)} - \mathbf{B}(\mathbf{L}^{(f)})^{-1} \boldsymbol{\mu}^{(f)} \right]_i, \quad \boldsymbol{\beta}_i^\top = \left[\mathbf{B}(\mathbf{L}^{(f)})^{-1} \right]_{i,\bullet}, \quad (6)$$

where $[\mathbf{A}]_{i,\bullet}$ denotes the i th row of a matrix $\mathbf{A}$.

One may also compare the resulting mean and covariance of $\mathbf{r}_t$ of the two approaches: combining (1) with (6), we find that

$$\mathbb{E}[\mathbf{r}_t] = \boldsymbol{\mu}^{(r)} = \begin{pmatrix} \beta_{1,0} \\ \vdots \\ \beta_{K_r,0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1^\top \\ \vdots \\ \boldsymbol{\beta}_{K_r}^\top \end{pmatrix} \mathbb{E}[\mathbf{f}_t] \quad (7)$$

and

$$\text{Cov}(\mathbf{r}_t) = \begin{pmatrix} \boldsymbol{\beta}_1^\top \\ \vdots \\ \boldsymbol{\beta}_{K_r}^\top \end{pmatrix} \text{Cov}(\mathbf{f}_t) \begin{pmatrix} \boldsymbol{\beta}_1 & \dots & \boldsymbol{\beta}_{K_r} \end{pmatrix} + \mathbf{C}^{(r)}, \quad (8)$$

which are two moments obtained from the regression in (4). Formula (8) is the basis for many shrinkage approaches to estimate the covariance matrix of $\mathbf{r}_t$ ([Ledoit and Wolf, 2020](#), Section 5), and, for $K_f < K_r$, the factors provide a dimension reduction as in [Fan et al. \(2008\)](#).

In the general (but still Gaussian) case with $\mathbf{L}^{(r)} \neq \mathbf{I}$, (5) becomes

$$\mathbf{r}_t = \left(\boldsymbol{\mu}^{(r)} - \mathbf{B}(\mathbf{L}^{(f)})^{-1} \boldsymbol{\mu}^{(f)} \right) + \mathbf{B}(\mathbf{L}^{(f)})^{-1} \mathbf{f}_t + \mathbf{L}^{(r)} \mathbf{v}_t^{(r)}, \quad (9)$$

which is a linear regression with an error vector $\mathbf{L}^{(r)} \mathbf{v}_t^{(r)}$ that is Gaussian with mean vector $\mathbf{0}$ and non-diagonal covariance matrix $\mathbf{L}^{(r)} \mathbf{C}^{(r)} (\mathbf{L}^{(r)})^\top$. Thus, instead of estimating the regression equation over the returns independently, they are interconnected through a correlation structure in the corresponding error terms as in seemingly unrelated regression equations (SURE) of [Zellner \(1962\)](#).

We now turn to the non-Gaussian case. Financial returns, and financial factors, measured at the daily or higher frequency, exhibit leptokurtosis and mild asymmetry. (This is illustrated below in Table 1.) In order to accommodate these stylized facts, we can impose in (1) a (semi-)heavy-tailed distribution such as the generalized hyperbolic (GHyp) distribution for the error terms. Its use was arguably popularized in McNeil et al. (2015), and continues to be used; see, e.g., Bianchi et al. (2020). That is, we take $\varepsilon_{i,t} \sim \text{GHyp}(\lambda, \alpha, \delta, \mu)$, with the probability density function

$$f_{\text{GHyp}}(y; \lambda, \alpha, \delta, \mu) = \frac{k_{\lambda-\frac{1}{2}}((y-\mu)^2 + \delta^2, \alpha^2)}{\sqrt{2\pi} k_{\lambda}(\delta^2, \alpha^2)}, \quad (10)$$

with $\lambda \in \mathbb{R}$, $\alpha > 0$ and $\delta > 0$, where

$$k_{\lambda}(\chi, \psi) = 2(\chi/\psi)^{\lambda/2} K_{\lambda}(\sqrt{\chi\psi}), \quad (11)$$

and $K_{\nu}(x)$ is the modified Bessel function of the third kind with index ν , given for all $x > 0$ by

$$K_{\nu}(x) = \frac{1}{2} \int_0^{\infty} t^{\nu-1} e^{-x(t+t^{-1})/2} dt. \quad (12)$$

It is noteworthy that each marginal distribution is endowed with its own set of shape parameters, allowing for tail heterogeneity, thus differentiating it from the use of the multivariate generalized hyperbolic distribution, as used in Paolella and Polak (2015b). Model (1) with GHyp innovations adapts the HGH model of Näf et al. (2019) by jointly modeling the multivariate distribution of the returns and financial factors. This distributional construction was proposed independently by Schmidt et al. (2006) and Näf et al. (2019), the latter authors not having known about the former. The set of parameters to estimate consists of

$$\mu = [\mu^{(f)\top}, \mu^{(r)\top}]^{\top}, \mathbf{L} = \left[\begin{array}{c|c} \mathbf{L}^{(f)} & \mathbf{0} \\ \hline \mathbf{B} & \mathbf{L}^{(r)} \end{array} \right], \Phi = (\alpha^{\top} \quad \delta^{\top} \quad \lambda^{\top} \quad \text{diag}(\mathbf{C})^{\top})^{\top}, \quad (13)$$

where Φ gathers the parameters corresponding to the distribution of $\mathbf{v}_t$. Näf et al. (2019) derive an ECME algorithm to estimate the parameters iteratively, by exploiting the mixed normal representation of the GHyp distribution, with the mixing distribution being the generalized inverse Gaussian distribution (GIG). Its use yields a closed form expression for the conditional moments computed in the E-step of the algorithm, as well as the lower-triangular form of matrix $\mathbf{L}$. This allows for sequential updating of the parameters in $\mathbf{L}$. In Appendix A.2, we adapt the algorithm to estimate the parameters in (13) that characterize the *joint* multivariate distribution of the financial factors and the returns of the assets in the portfolio. The returns and factors exhibit relation (9), but now with errors in the regression that are both correlated and heavy-tailed.

3 Portfolio Optimization

For a given set of K_r assets, we would like to produce a portfolio vector $\mathbf{w}$, using the additional information of K_f factors. In addition, we impose a no-short-selling constraint, this being, firstly, quite common in

practice; see, e.g., [Almazan et al. \(2004\)](#), who report that 70% of mutual funds explicitly state that short selling is not permitted; and, secondly, its imposition can be interpreted as a useful form of shrinkage that leads to better performance; see, e.g., [Jagannathan and Ma \(2003\)](#), and [DeMiguel et al. \(2009\)](#). We find optimal portfolio weights by minimizing, for a given level $\alpha \in (0, 1)$, the expected shortfall of the portfolio returns

$$\min_{\mathbf{w} \in \mathfrak{B}_\theta} \text{ES}_\alpha(\mathbf{w}^\top \mathbf{r}_{t+1}), \quad (14)$$

conditional on the information up to time t . $\mathfrak{B}_\theta$ characterizes the set of feasible portfolios for a long-only strategy

$$\mathfrak{B}_\theta = \left\{ \mathbf{w} \in \mathbb{R}^{K_r} : \mathbf{w}' \boldsymbol{\mu}^{(r)} \geq \theta, \sum_{k=1}^K w_k = 1, w_k \geq 0, \text{ for } k = 1, \dots, K \right\}, \quad (15)$$

which includes lower bound for the expected portfolio return, fully invested portfolio, and the short-selling constraint. In the case of Gaussian errors from Section 2, as shown in [Embrechts et al. \(2002\)](#), (14) is equivalent to a min-variance portfolio

$$\min_{\mathbf{w} \in \mathfrak{B}_\theta} \text{Var}(\mathbf{w}^\top \mathbf{r}_{t+1}) = \min_{\mathbf{w} \in \mathfrak{B}_\theta} \mathbf{w}^\top \text{Cov}(\mathbf{r}_{t+1}) \mathbf{w}, \quad (16)$$

with, according to our model,

$$\text{Cov}(\mathbf{r}_{t+1}) = \mathbf{B} \mathbf{C}^{(f)} \mathbf{B}^\top + \mathbf{L}^{(r)} \mathbf{C}^{(r)} (\mathbf{L}^{(r)})^\top.$$

In the case of a non-elliptical HGH model, as shown in [Näf et al. \(2019\)](#), direct calculation of the expected shortfall in (14) requires numerical integration over a density that has itself an infinite series representation. Therefore, we use the [Rockafellar and Uryasev \(2000\)](#) result, together with the saddle point approximation from [Broda and Paoletta \(2009\)](#). In particular, let $S_t = \mathbf{w}^\top \mathbf{r}_t$ denote the portfolio return at time t , then we use

$$F_\alpha(y; \mathbf{w}) = -\mathbf{w}^\top \boldsymbol{\mu} + y \frac{2\alpha - 1}{2\alpha} + \frac{1}{\pi\alpha} \int_0^{+\infty} \text{Im} \left(\frac{\mathbb{M}_{S_t}(iz)}{e^{-iz\mathbf{w}^\top \boldsymbol{\mu}}} (\mathbb{K}'_{S_t}(iz) - \mathbf{w}^\top \boldsymbol{\mu} + y) e^{-izy} \right) \frac{dz}{z}, \quad (17)$$

where $\mathbb{M}_S$ and $\mathbb{K}_S$ denote, respectively, the moment generating function (mgf) and the cumulant generating function (cgf) of random variable S . Solving the minimization problem

$$(y^*, \mathbf{w}^*) = \arg \min_{(y, \mathbf{w}) \in \mathbb{R} \times \mathfrak{B}_\theta} F_\alpha(y; \mathbf{w}), \quad (18)$$

gives back the optimal portfolio $\mathbf{w}^*$. Moreover, the ES of the optimal portfolio is given as $\text{ES}_\alpha(\mathbf{w}^{*\top} \mathbf{Y}) = \min_{(y, \mathbf{w})} F_\alpha(y; \mathbf{w})$. Evaluation of (17) requires (only) computable expressions for $\mathbb{M}_{S_t}$ and $\mathbb{K}_{S_t}$ of the portfolio distribution. The distribution of $\mathbf{r}_t$ is not HGH anymore, but we can still derive the mgf of $\mathbf{r}_t$, $\mathbb{M}_{\mathbf{r}_t}$, which is necessary for portfolio optimization. Define $\mathbf{0}_{k_1 \times k_2}$ as a matrix of zeros of dimension $k_1 \times k_2$,

$\mathbf{I}_{k_1 \times k_1}$ to be the identity matrix of dimension $k_1 \times k_1$ and

$$\mathbf{A}_1 := \begin{pmatrix} \mathbf{I}_{K_f \times K_f} & \mathbf{0}_{K_f \times K_r} \end{pmatrix}, \quad \mathbf{A}_2 := \begin{pmatrix} \mathbf{0}_{K_r \times K_f} & \mathbf{I}_{K_r \times K_r} \end{pmatrix}, \quad (19)$$

such that $\mathbf{f}_t = \mathbf{A}_1 \mathbf{Z}_t$, $\mathbf{r}_t = \mathbf{A}_2 \mathbf{Z}_t$, $\mathbf{L}^{(f)} = \mathbf{A}_1 \mathbf{L} \mathbf{A}_1^\top$, $\mathbf{L}^{(r)} = \mathbf{A}_2 \mathbf{L} \mathbf{A}_2^\top$ and $\mathbf{B} = \mathbf{A}_1 \mathbf{L} \mathbf{A}_2^\top$.

Lemma 1 *Let $\mathbf{X}, \mathbf{Y}$ be random vectors in $\mathbb{R}^{K_f}$ and $\mathbb{R}^{K_r}$ respectively, such that $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L})$, with $K = K_f + K_r$. Then the mgf of $\mathbf{Y}$ is given as*

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) = \exp(\mathbf{u}^\top \boldsymbol{\mu}^{(\mathbf{Y})}) \prod_{i=1}^K \left(\frac{\alpha_i^2}{\alpha_i^2 - \tilde{u}_i^2} \right)^{\lambda_i/2} \frac{K_{\lambda_i}(\delta_i \sqrt{\alpha_i^2 - \tilde{u}_i^2})}{K_{\lambda_i}(\delta_i \alpha_i)}, \quad (20)$$

with, for $i = 1, \dots, K$,

$$\tilde{u}_i = \begin{cases} c_{ii,t}^{1/2} \sum_{\ell=k+1}^K u_\ell q_{\ell i}, & \text{if } i < k+1, \\ c_{ii,t}^{1/2} \left(u_i + \sum_{\ell=i+1}^K u_\ell q_{\ell i} \right), & \text{if } i \geq k+1. \end{cases}$$

The changes induced by considering factors very much resembles the Gaussian case. For instance, and similar to the comparable result stated in Section 2, it holds that

$$\text{Cov}(\mathbf{r}_t) = \mathbf{B} \text{Cov}(\mathbf{v}_t^{(f)}) \mathbf{B}^\top + \mathbf{L}^{(r)} \mathbf{C}^{(r)} (\mathbf{L}^{(r)})^\top. \quad (21)$$

In general, the mgf in (20) for $\mathbf{Y} = \mathbf{r}_t$ depends on $\mathbf{f}_t$ through $q_{\ell i}$, for $\ell \geq k+1$ and $i < k+1$, which corresponds to the effects of the factors on the returns. The proof of Lemma 1 can be found in Appendix B.

4 Empirical Results

In a comprehensive empirical analysis, we compare our proposed modeling approach with classical methods from the literature. We consider two datasets, the 49 daily industry portfolios (FF49) from the [Kenneth R. French homepage](#) and 42 hourly cryptocurrencies from the [Bitfinex homepage](#). We excluded the cryptocurrencies that exhibit almost no variation over time, like DAI. In the case of the FF49 dataset, we consider the Fama-French three factors (FF-3), the Fama-French five factors (FF-5) and an additional Momentum factor (FF-3 + MOM and FF-5 + MOM). So the common risk factors include the excess return on the market factor, which includes all NYSE, AMEX, and NASDAQ firms (Mkt-RF), the high minus low factor (HML), the conservative minus aggressive factor (CMA), the robust minus weak factor (RMW), the small minus big factor (SMB) and the momentum factor (MOM). Again, the data is downloaded from the [Kenneth R. French homepage](#).

In case of the cryptocurrencies, we only consider the market factor (CMkt) as a predictor. Similar to [Shen et al. \(2020\)](#), the market factor is constructed from 65% Bitcoin Cash (BTC), 25% Ethereum (ETH) and giving equal weight to the remaining 10% in the universe. Table 1 shows the first four moments of the seven

factors, with added bootstrap standard errors in brackets. It appears that the factors do not exhibit strong asymmetry as measured by the sample skewness, the use of which assumes existence of third moments. The sample kurtosis values serve as a heuristic to indicate the presence of strong leptokurtosis or the possible presence of heavy tails, though there appears to be no large discrepancy between the tail behaviors in terms of kurtosis between all factors of the FF49. We also would like to emphasize that pooling the data as we do here tends to drive up kurtosis values, whereas they appear closer to Gaussian for many of the much shorter moving windows for the FF49 data. On the other hand, the cryptocurrencies appear to exhibit heavy tails also in rolling windows. However these remain heuristic arguments, as assessing the maximally existing moment of the underlying distribution is nearly futile (see, e.g., [McNeil et al., 2015](#); [Paolella, 2016](#), and the numerous references therein). Interestingly the market factor of the FF49 (Mkt-RF) and the market factor of the cryptocurrencies (CMkt) exhibit very similar behavior in these summary statistics. Overall, the summary statistics provide a heuristically based argument that modeling the factors themselves with a heavy-tailed distribution *might* be beneficial. Formal testing procedures will not help: They have very low power, and the mapping from a Neyman-Pearson test, or use of Fisherian p -values, to efficacy in forecasting is anyway not established. As such, it is not clear whether the ability to model vastly different tail behaviors of the HGH is needed for the six factors in the FF49 application. It might make more sense to model a single tail behavior for all factors; see Section 5. Finally, the risk-free rate is subtracted from the returns of the FF49 data.

We conduct mean-expected shortfall portfolio optimization for the models described in Section 2, as described in Section 3, setting $\theta = 0$ and $\alpha = 0.15$. In case of the HGH model structure, we focus on a subclass of our distribution by taking each G_k to be standard gamma distributed, i.e., $G_k \sim \text{Gam}(\lambda_k, 1)$ independent for all $k \in \{1, \dots, K\}$. The gamma distribution arises as a special case of the $\text{GIG}(\lambda, \alpha, \delta)$ in the limit as $\delta \rightarrow 0$.

The results are summarized in Table 2 and Figure 1. As performance measures we report the average annualized return (*Return*), the annualized standard deviation (*Volatility*), the final return (*Total Return*), the maximum drawdown (*Max. Drawdown*), the annualized percentage turnover (*Turnover*), the annualized Sharpe ratio (*Sharpe*), the annualized Sortino ratio (*Sortino*), the annualized percentage STARR-ratio at the 98.5% level ($\text{STARR}_{98.5\%}$) and the annualized empirical expected shortfall at the 98.5% level ($\text{ES}_{98.5\%}$). Note, the STARR-ratio is defined as the average return divided by the empirical expected shortfall at a given level. The Factor-HGH model is estimated with $L^{(r)}$ equal to the identity matrix, while the Gaussian-Cholesky model has an unconstrained $L^{(r)}$. We propose this method as the optimal combination between good performance and estimation complexity. Further, in the case of the FF49 dataset, the five Fama-French factors are used as predictors. For an extensive performance overview of all combinations of model structures and factors, see Appendix C.

An issue that arises for the practical use of the HGH model (and thus also in the Factor-HGH) is the choice of the ordering of marginals. For this rolling window exercise, the reordering of the components in $\mathbf{Y}_t$ is done beforehand with 1'000 datapoints. On those datasets, the same approach as in [Näf et al. \(2019\)](#) is used, but separately for factors and returns. That is, while we change the ordering within factors and

Table 1: Various sample statistics of the 5386 respectively 3987 percentage returns for the $K_f = 6$ factors of the FF49 and $K_f = 1$ factor of the cryptocurrencies under study. Values in parentheses indicate standard errors computed using a simple nonparametric bootstrap.

	Mkt-RF	HML	CMA	RMW	SMB	MOM	CMkt
Mean	0.023 (0.018)	0.007 (0.010)	0.011 (0.006)	0.018 (0.007)	0.012 (0.009)	0.004 (0.014)	0.009 (0.017)
Variance	1.598 (0.072)	0.564 (0.025)	0.183 (0.010)	0.278 (0.010)	0.401 (0.014)	1.168 (0.065)	1.252 (0.068)
Skewness	-0.440 (0.287)	0.288 (0.224)	-0.728 (0.523)	0.178 (0.167)	-0.090 (0.183)	-1.322 (0.453)	-0.329 (0.418)
Kurtosis	12.893 (1.745)	11.262 (0.858)	16.457 (6.295)	8.418 (0.773)	7.777 (1.106)	17.530 (6.028)	13.652 (3.14)

Table 2: Comparison of the gross performance for mean-expected shortfall portfolios for the FF49 dataset and the crypto dataset. * in the top row denotes annualized quantities. The FF49 dataset contains daily returns of 49 industry portfolios from 03-01-2000 until 28-05-2021. The crypto dataset contains hourly returns of 42 cryptocurrencies from 01-01-2021 until 16-06-2021. Rebalancing is done every 5 days for the FF49 data and every 5 hours for the crypto data, respectively. The Factor-HGH model is estimated with $L^{(r)}$ equal to the identity matrix. Within a dataset, the best performance statistics are marked in **bold** (not including the market portfolio).

	Return*	Volatility*	Total Return	Max. Drawdown	Turnover*	Sharpe*	Sortino*	STARR _{98.5%} %*	ES _{98.5%} *
FF49 Data									
Mkt-RF	8.49	19.86	281.86	-55.81	0.00	0.43	0.60	1.77	478.85
Intercept-Only	17.34	14.96	2710.22	-45.73	943.05	1.16	1.65	5.27	328.72
Fama-French	17.87	13.89	3135.25	-44.84	790.88	1.29	1.83	5.66	315.93
Gaussian-Cholesky	17.34	14.96	2710.76	-45.76	943.88	1.16	1.65	5.27	328.79
HGH	16.79	18.04	2152.58	-60.30	2013.54	0.93	1.31	4.66	360.42
Factor-HGH	16.73	14.52	2410.52	-52.11	895.51	1.15	1.62	5.10	328.24
Crypto Data									
CMkt	57.36	102.88	1.91	-50.10	0.00	0.56	0.78	0.39	420.81
Intercept-Only	12.32	8.55	5.30	-5.62	104.35	1.44	1.93	1.35	26.28
CAPM	15.46	10.71	6.64	-8.98	128.84	1.44	1.87	1.33	33.35
Gaussian-Cholesky	18.18	8.60	8.00	-5.62	110.98	2.11	2.85	2.00	26.12
HGH	21.16	9.72	9.35	-3.11	99.52	2.18	3.28	2.40	25.32
Factor-HGH	25.50	10.04	11.40	-2.90	80.63	2.54	3.94	2.79	26.28

returns by a permutation matrix $\mathbf{P}$, we still have that for $t = 1, \dots, T$,

$$\mathbf{PZ}_t = \begin{bmatrix} \mathbf{P}_1 \mathbf{f}_t \\ \mathbf{P}_2 \mathbf{r}_t \end{bmatrix}$$

for two permutation matrices $\mathbf{P}_1, \mathbf{P}_2$.

In the case of the FF49 dataset, using the FF5 factors, we barely see a performance difference between the model approaches. Even more surprisingly, the intercept-only model (using no factor information at all) is performing similarly compared to the more sophisticated model structures. Although better than the base HGH model, which uses no factor information, our Factor-HGH model falls behind all the classical Gaussian models. It seems that heavy tail adjusted residuals bring no benefit in this rather Gaussian uni-

Figure 1: Gross percentage long-only cumulative portfolio returns for the 49 industry portfolios and the 42 cryptocurrencies. The industry portfolios contain daily observations from 03-01-2000 until 28-05-2021. The cryptocurrencies contain hourly observations from 01-01-2021 until 16-06-2021. The Factor-HGH model is estimated with $L^{(r)}$ equal to the identity matrix.

verse. Though, with a Sharpe-ratio difference of only 0.14 to the Fama-French model, the damage appears limited. Further, the classical Fama-French model is slightly better than the Gaussian-Cholesky model structure. Finally, as can be studied in Appendix C, the model ranking is consistent among the different factor structures FF-3, FF-3 + MOM, FF-5 and FF-5 + MOM.

If we switch to cryptocurrencies, which by all means fulfill stylized facts like heavy tails and varying tail behavior even on a rolling window level, our proposed model structures start to excel. Compared to the benchmark models, the portfolio based on the Factor-HGH model doubles the average return while keeping the volatility, the maximum drawdown, the turnover, and the expected-shortfall at a low level. Notably as of the beginning of January 2021, we observe that the two HGH models are more robust against downside risk. Also, the cumulative return climbs way above the Gaussian models afterward. However, it must be emphasized that during the period of mid-January until the beginning of February the volatility of the HGH models is considerably higher than for the Gaussian models. Further, after May, the Factor-HGH and the

basic (no factor) HGH remain on a stable path, while the Gaussian models perform more volatile and their cumulative returns begin a gentle decline. The stable performance of the HGH-type models is due to the investment into the more stable coins, such as PAX. Compared to the base HGH, the Factor-HGH achieves this stable performance with a much lower turnover. This low turnover is remarkable also in comparison with the other models. It shows that, in terms of net returns, the difference between the Factor-HGH and the competing models is even more pronounced. As shown in Table 5 and Figure 2 in Appendix C, this strong performance in net gains remains, even after having removed the (very stable) PAX coin from the investment universe.

If we would simply invest in the market factor, which is mainly driven by BTC, one would end up with much worse performance. Unlike in our crypto universe, for the 49 industry portfolios, investing in the market factor only does not lead to a catastrophic performance. However, the summary statistics of Mkt-RF and CMkt in Table 1 would not necessarily predict such a different outcome. This further highlights the difficulty of investing in a crypto universe.

In summary, when considering the 49 industry portfolios from the [Kenneth R. French homepage](#), the Factor-HGH can keep up with classical Gaussian models. On the other hand, when we draw our attention to a more non-Gaussian scenario, like cryptocurrencies, we observe a sound performance benefit of the Factor-HGH model compared to the benchmark models.

5 Extension: Hybrid Models

So far, we only considered having K independent GHyp realizations in model (1). However we demonstrate now that a wide variety of models can be generated by different distributional choices on ϵ_t . For simplicity we will in this section differentiate between $\epsilon_{i,t} \sim \text{GHyp}(\lambda, \alpha, \delta, \mu)$, $(\epsilon_{i_1,t}, \dots, \epsilon_{i_n,t}) \sim \text{MGHyp}(\lambda, \alpha, \delta, \mu, \Sigma)$ and $\epsilon_{i,t} \sim N(\mu, 1)$, where MGHyp denotes the multivariate generalized hyperbolic distribution, see e.g., [Paolella and Polak \(2015b\)](#) and [McNeil et al. \(2015\)](#). The relevant difference between those distributions in our context, lies in the choice of mixture variables G . Let for the following

$$\mathbf{D}_t = \text{diag}(G_{1,t}, \dots, G_{K,t}), \quad (22)$$

such that we may rewrite model (1) as

$$\mathbf{Z}_t = \boldsymbol{\mu} + \mathbf{L}\mathbf{C}^{1/2}\mathbf{D}_t^{1/2}\boldsymbol{\epsilon}_t, \quad (23)$$

with $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I})$.

We then impose the following rules on the joint distribution of $G_{1,t}, \dots, G_{K,t}$: For any k , either let $G_{k,t} \sim \text{GIG}(\lambda_k, \alpha_k, \delta_k)$, or set $G_{k,t} = 1$. Moreover for each group of size $s < K$, $(G_{k_1,t}, \dots, G_{k_s,t})$ either is independent or perfectly dependent, in the sense that $G_{k_1,t} = \dots = G_{k_s,t}$. Note that for $G_{k,t} = 1$ for all $k = 1, \dots, K$ and $t = 1, \dots, T$, this gives the Gaussian model of Section 2 with its variations depending on the choice of $\mathbf{L}$. On the other hand, choosing $G_{k,t} \sim \text{GIG}(\lambda_k, \alpha_k, \delta_k)$ independent for all k results in the HGH model. Alternatively, we might choose to only allow for $G_{1,t} \sim \text{GIG}(\lambda_1, \alpha_1, \delta_1)$, while for $k > 1$, $G_{k,t} = 1$

for all t . This would mean that there is a single shock $G_{1,t}$ propagating through the whole distribution, but only ever working through $\varepsilon_{1,t}$. Finally imagine $G_{1,t} = \dots = G_{K,t} = \mathbf{G}_t \sim \text{GIG}(\lambda, \alpha, \delta)$. Now again there is just a single shock, but this time it affects all independent errors $\varepsilon_{1,t}, \dots, \varepsilon_{K,t}$ simultaneously. In this case we obtain the MGHyp distribution.

To formalize the above, assume there exists $j = 1, \dots, d$ independent random variables $\mathbf{G}_{1,t}, \dots, \mathbf{G}_{d,t}$ such that either $\mathbf{G}_{j,t} \sim \text{GIG}(\lambda_j, \alpha_j, \delta_j)$ for all t or $\mathbf{G}_{j,t} = 1$ for all t . Let for the following,

$$\begin{aligned}\mathcal{K}_0 &:= \{k : G_{k,t} = 1\} \\ \mathcal{K}_1 &:= \{k : G_{k,t} = \mathbf{G}_{1,t} \sim \text{GIG}(\lambda_1, \alpha_1, \delta_1)\} \\ \mathcal{K}_2 &:= \{k : G_{k,t} = \mathbf{G}_{2,t} \sim \text{GIG}(\lambda_2, \alpha_2, \delta_2)\} \\ &\vdots \\ \mathcal{K}_d &:= \{k : G_{k,t} = \mathbf{G}_{d,t} \sim \text{GIG}(\lambda_d, \alpha_d, \delta_d)\},\end{aligned}\tag{24}$$

where $\mathcal{K}_0$ is allowed to be empty. Then, for a given choice of these sets, the model can be estimated by the ECME algorithm presented in Appendix A.3. We will denote the resulting distribution with $\mathbf{Z} \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L}, \mathcal{K})$ with $\mathcal{K} = (\mathcal{K}_j)_{j=0}^d$.

Lemma 2 Define for $\mathbf{v} = (v_1, \dots, v_K) \in \mathbb{R}^K$,

$$\tilde{v}_i = c_{ii}^{1/2} \left(v_i + \sum_{k=i+1}^K v_k q_{ki} \right), \quad i = 1, \dots, K,$$

and

$$\omega_j = \left(\sum_{i \in \mathcal{K}_j} \tilde{v}_i^2 \right)^{1/2}.\tag{25}$$

The moment generating function of $\mathbf{Z} \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L}, \mathcal{K})$ is given by

$$\mathbb{M}_{\mathbf{Z}}(\mathbf{v}) = \exp(\mathbf{v}^\top \boldsymbol{\mu}) \exp(\omega_0^2/2) \cdot \prod_{j=1}^d \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right)^{\lambda_j/2} \frac{K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2})}{K_{\lambda_j}(\delta_j \alpha_j)},\tag{26}$$

for $\mathbf{v}$ such that $\omega_j \in (-\alpha_j, \alpha_j)$ for all j . This leads to the cumulant generating function

$$\mathbb{K}_{\mathbf{Z}}(\mathbf{v}) = \mathbf{v}^\top \boldsymbol{\mu} + \frac{\omega_0^2}{2} + \sum_{j=1}^d \left\{ \frac{\lambda_j}{2} \ln \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right) + \ln K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2}) - \ln K_{\lambda_j}(\delta_j \alpha_j) \right\}.$$

The proof of Lemma 2 can be found in Appendix B. In our context of factors, one can again derive the marginal mgf for this model class:

Lemma 3 Let $\mathbf{X}, \mathbf{Y}$ be random vectors in $\mathbb{R}^{K_f}$ and $\mathbb{R}^{K_r}$ respectively, such that $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L}, \mathcal{K})$,

with $K = K_f + K_r$. Then the mgf of $\mathbf{Y}$ is given as

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) = \exp(\mathbf{u}^\top \boldsymbol{\mu}) \exp(\omega_0^2/2) \cdot \prod_{j=1}^d \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right)^{\lambda_j/2} \frac{K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2})}{K_{\lambda_j}(\delta_j \alpha_j)},$$

where for $j = 0, \dots, d$, ω_j is defined as in (25) with, for $i = 1, \dots, K$,

$$\tilde{u}_i = \begin{cases} c_{ii,t}^{1/2} \sum_{\ell=k+1}^K u_\ell q_{\ell i}, & \text{if } i < k+1, \\ c_{ii,t}^{1/2} \left(u_i + \sum_{\ell=i+1}^K u_\ell q_{\ell i} \right), & \text{if } i \geq k+1. \end{cases} \quad (27)$$

From the above it follows that the linear combination $S = \mathbf{w}^\top \mathbf{Y}$ has cgf

$$\begin{aligned} \mathbb{K}_S(t) &= \log(\mathbb{E}[\exp(t\mathbf{w}^\top \mathbf{Y})]) \\ &= t\mathbf{w}^\top \boldsymbol{\mu} + \frac{t^2 \omega_0^2}{2} + \sum_{j=1}^d \left\{ \frac{\lambda_j}{2} \ln \left(\frac{\alpha_j^2}{\alpha_j^2 - t^2 \omega_j^2} \right) + \ln K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - t^2 \omega_j^2}) - \ln K_{\lambda_j}(\delta_j \alpha_j) \right\} \end{aligned} \quad (28)$$

with $\omega_j = \left(\sum_{i \in \mathcal{K}_j} \tilde{u}_i^2 \right)^{1/2}$, $\tilde{u}_i$, $i = 1, \dots, K$, defined analogously to (27). This can be used for ES-based portfolio optimization as outlined in Section 3.

As an example, consider the case when there is one G only and $\mathbf{C} = \mathbf{I}$ for simplicity. In the case when this G is influencing all returns simultaneously, but is not present in the factors, it holds that

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu}^{(\mathbf{Y})} + \mathbf{B}\boldsymbol{\epsilon}^{(\mathbf{X})} + \sqrt{G}\mathbf{L}^{(\mathbf{Y})}\boldsymbol{\epsilon}^{(\mathbf{Y})}. \quad (29)$$

If we instead have a joint G in the factors and returns, then

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu}^{(\mathbf{Y})} + \sqrt{G} \left(\mathbf{B}\boldsymbol{\epsilon}^{(\mathbf{X})} + \mathbf{L}^{(\mathbf{Y})}\boldsymbol{\epsilon}^{(\mathbf{Y})} \right), \quad (30)$$

so that

$$\mathbf{Y} \sim \text{MGHyp}(\lambda, \alpha, \delta, \boldsymbol{\mu}^{(\mathbf{Y})}, \mathbf{B}\mathbf{B}^\top + \mathbf{L}^{(\mathbf{Y})}\mathbf{L}^{(\mathbf{Y})\top}),$$

which may formally be checked by using the mgf from Lemma 2.

6 Conclusion

This paper details a method for incorporating exogenous common factor information into a tractable model that allows for non-elliptic behavior in the factors and asset returns, notably semi-heavy tails, margin asymmetry, and, notably, heterogeneous tail behavior. Parameter estimation is straightforward and fast, via an ECME algorithm; as is conducting mean-ES portfolio optimization. The proposed model structure is shown to generalize several existing factor models, including the classical Gaussian factor models of Fama and French (1993), Carhart (1997), and Fama and French (2015).

The empirical analysis indicates that classical Gaussian models and the new Factor-HGH models per-

form similarly for the standard FF49 data set. Irrespective of the model used, incorporation of the common factor information does not lead to markedly improved performance in the FF49 universe: The intercept-only model is almost as good as the more sophisticated factor models. However, in the case of cryptocurrencies, where stylized facts such as heavy tails and heterogeneous tail behavior are rather pronounced, the proposed HGH structure clearly outperform the classical models. Intriguingly, the dimension reduction by modelling dependence only through the factors leads to a performance boost compared to the full HGH distribution. Further, in case of the cryptocurrency data, incorporation of the common market factor information into the model structure results in a clear performance increase. To conclude, in the world of cryptocurrencies, the proposed Factor-HGH opens an interesting playground for modeling factors and returns jointly under non-Gaussian errors.

References

- Almazan, A., Brown, K. C., Carlson, M., and Chapman, D. A. (2004). Why Constrain Your Mutual Fund Manager? *Journal of Financial Economics*, 73(2):289–321.
- Bai, J. and Liao, Y. (2016). Efficient Estimation of Approximate Factor Models via Penalized Maximum Likelihood. *Journal of Econometrics*, 191(1):1–18.
- Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.
- Bai, J. and Ng, S. (2013). Principal Components Estimation and Identification of Static Factors. *Journal of Econometrics*, 176(1):18–29.
- Bao, T., Diks, C., and Li, H. (2018). A Generalized CAPM Model with Asymmetric Power Distributed Errors with an Application to Portfolio Construction. *Economic Modelling*, 68:611–621.
- Bianchi, M. L., Hitaj, A., and Tassinari, G. L. (2020). Multivariate Non-Gaussian Models for Financial Applications. *arXiv preprint arXiv:2005.06390*.
- Brauneis, A. and Mestel, R. (2019). Cryptocurrency-Portfolios in a Mean-Variance Framework. *Finance Research Letters*, 28:259–264.
- Broda, S. A. and Paoletta, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 7(4):412–436.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, 51(5):1281–1304.
- Chicheportiche, R. and Bouchaud, J.-P. (2012). The Joint Distribution of Stock Returns is Not Elliptical. *International Journal of Theoretical and Applied Finance*, 15(3).

- Chung, Y. P., Johnson, H., and Schill, M. J. (2006). Asset Pricing When Returns Are Nonnormal: Fama-French Factors versus Higher-Order Systematic Comoments. *The Journal of Business*, 79(2):923–940.
- Darolles, S., Francq, C., and Laurent, S. (2018). Asymptotics of Cholesky GARCH Models and Time-Varying Conditional Betas. *Journal of Econometrics*, 204(2):223–247.
- De Nard, G., Hediger, S., and Leippold, M. (2020). Subsampled Factor Models for Asset Pricing: The Rise of Vasa. *Available at SSRN 3557957*.
- DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management science*, 55(5):798–812.
- Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and Dependency in Risk Management: Properties and Pitfalls. In Dempster, M. A. H., editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, Cambridge.
- Fama, E. F. and French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116(1):1–22.
- Fan, J., Fan, Y., and Lv, J. (2008). High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, 147(1):186–197.
- Feng, G., Giglio, S., and Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, 75(3):1327–1370.
- Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273.
- Heaton, J., Polson, N., and Witte, J. H. (2017). Deep Learning for Finance: Deep Portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.
- Hu, Y., Rachev, S. T., and Fabozzi, F. J. (2019). Modelling Crypto Asset Price Dynamics, Optimal Crypto Portfolio, and Crypto Option Valuation. *arXiv preprint arXiv:1908.05419*.
- Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683.
- Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91.
- Kan, R. and Zhou, G. (2017). Modeling Non-Normality Using Multivariate t: Implications for Asset Pricing. *China Finance Review International*, 7(1):2–32.

- Ledoit, O. and Wolf, M. (2020). The Power of (Non-) Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*.
- Lintner, J. (1965). Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance*, 20(4):587–615.
- Liu, Y. and Tsyvinski, A. (2021). Risks and Returns of Cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727.
- Marinelli, C., d’Addona, S., and Rachev, S. T. (2012). Multivariate Heavy-Tailed Models for Value-at-Risk Estimation. *International Journal of Theoretical and Applied Finance*, 15(04):1250029.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton, revised edition.
- Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica: Journal of the Econometric Society*, pages 768–783.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.
- Näf, J., Paoletta, M. S., and Polak, P. (2019). Heterogeneous Tail Generalized COMFORT Modeling via Cholesky Decomposition. *Journal of Multivariate Analysis*, 172(C):84–106.
- Pagan, A. (1996). The Econometrics of Financial Markets. *Journal of Empirical Finance*, 3(1):15–102.
- Paoletta, M. S. (2007). *Intermediate Probability: A Computational Approach*. John Wiley & Sons, Chichester.
- Paoletta, M. S. (2016). Stable-GARCH Models for Financial Returns: Fast Estimation and Tests for Stability. *Econometrics*, 4(2). Article 25.
- Paoletta, M. S. and Polak, P. (2015a). ALRIGHT: Asymmetric LaRge-Scale (I)GARCH with Hetero-Tails. *International Review of Economics and Finance*, 40:282–297.
- Paoletta, M. S. and Polak, P. (2015b). COMFORT: A Common Market Factor Non-Gaussian Returns Model. *Journal of Econometrics*, 187(2):593–605.
- Paoletta, M. S. and Polak, P. (2018). COBra: Copula-Based Portfolio Optimization. In Vladik Kreinovich, Songsak Sriboonchitta, N. C., editor, *Studies in Computational Intelligence: Predictive Econometrics and Big Data*. Springer.
- Paoletta, M. S., Polak, P., and Walker, P. S. (2019). Regime Switching Dynamic Correlations for Asymmetric and Fat-Tailed Conditional Returns. *Journal of Econometrics*, 213(2):493–515.
- Pourahmadi, M. (1999). Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterisation. *Biometrika*, 86(3):677–690.

- Rockafellar, R. T. and Uryasev, S. P. (2000). Optimization of Conditional Value at Risk. *Journal of Risk*, 2:21–41.
- Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13(3):341–360.
- Ross, S. A. (1977). The Capital Asset Pricing Model (CAPM), Short-Sale Restrictions and Related Issues. *The Journal of Finance*, 32(1):177–183.
- Schmidt, R., Hrycej, T., and Stützel, E. (2006). Multivariate Distribution Models with Generalized Hyperbolic Margins. *Computational Statistics & Data Analysis*, 50(8):2065–2096.
- Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, 19(3):425–442.
- Shen, D., Urquhart, A., and Wang, P. (2020). A Three-Factor Pricing Model for Cryptocurrencies. *Finance Research Letters*, 34:101248.
- Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Treynor, J. L. (1961). Market Value, Time, and Risk. *Time, and Risk* (August 8, 1961).
- Tsai, H. and Tsay, R. S. (2010). Constrained Factor Models. *Journal of the American Statistical Association*, 105(492):1593–1605.
- Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298):348–368.
- Zhang, W., Wang, P., Li, X., and Shen, D. (2018). Some Stylized Facts of the Cryptocurrency Market. *Applied Economics*, 50(55):5950–5965.
- Zhou, W. and Li, L. (2016). A New Fama-French 5-Factor Model Based on SSAEPD Error and GARCH-Type Volatility. *Journal of Mathematical Finance*, 6(05):711.

A Estimation

This section presents ECME algorithms to estimate the parameters of the models in the main text. All algorithms shown here are essentially slight modifications of the ECME algorithms in [Paoletta and Polak \(2015b\)](#) and [Näf et al. \(2019\)](#). As a consequence, it can be shown that all of them monotonically increases the likelihood in each iteration. We start with the important moment expression of a GIG random variable,

given for instance in (Paolella, 2007, Section 9.4): If $G \sim \text{GIG}(\lambda, \alpha, \delta)$, the r th moment, $r \in \mathbb{R}$, can be expressed as

$$\mathbb{E}[G^r] = k_{\lambda+r}(\delta^2, \alpha^2) / k_{\lambda}(\delta^2, \alpha^2), \quad (31)$$

where $k_{\lambda}(\chi, \psi)$ is given in (11). To abbreviate, let

$$\boldsymbol{\theta}_P = (\boldsymbol{\mu}, \mathbf{L}, \mathbf{C}) \quad \text{and} \quad \boldsymbol{\theta}_D = (\lambda_1, \dots, \lambda_K, \alpha_1, \dots, \alpha_K, \delta_1, \dots, \delta_K). \quad (32)$$

A.1 Gaussian Errors

For the case of Gaussian errors, the ECME algorithm becomes trivially:

Step 1: Estimate $\mathbf{L}$ and $\boldsymbol{\mu}$ by

$$\arg \max_{\mathbf{L}, \boldsymbol{\mu}} \left[\sum_{k=1}^K -\frac{1}{2} c_k^{-1} \sum_{t=1}^T \left\{ y_{k,t} - (\mu_k + q_{k1} v_{1,t} + q_{k2} v_{2,t} + \dots + q_{k(k-1)} v_{k-1,t}) \right\}^2 \right], \quad (33)$$

where, for each k , the term in the first sum in (33) corresponds to a multiple linear regression problem of the dependent variable $y_{k,t}$ regressed onto the exogenous variables and $v_{j,t}$, $j = 1, \dots, k-1$. In particular, each of these regressions is independent from each other and from c_k^{-1} .

Step 2: Obtain the residuals from Step 1 using $\mathbf{v}_t = \mathbf{L}^{-1}(\mathbf{Y}_t - \boldsymbol{\mu})$ and solve

$$\arg \max_{(c_k)_{k=1}^K} \left[\sum_{k=1}^K -\frac{1}{2} \sum_{t=1}^T \left\{ \ln(c_k) + c_k^{-1} v_{k,t}^2 \right\} \right], \quad (34)$$

which has a closed-form solution given by

$$c_k = \frac{1}{T} \sum_{t=1}^T v_{k,t}^2. \quad (35)$$

In order to allow for sparse $\mathbf{L}$, this algorithm can be augmented by incorporating ℓ_1 regularization into (33). This generalization corresponds to a maximum a posteriori estimation with Laplace prior $f_{q_k}(q) = \eta_k \exp(-\eta_k |q|)$, where $\eta_k > 0$. Define, for $k = 1, \dots, K$,

- $\mathbf{L}_{k,\bullet} = (q_{k,1}, \dots, q_{k,k-1})^\top$ the first $k-1$ elements of k th row of $\mathbf{L}$ written as a column vector,
- $\hat{\mathbf{x}}_{k,t} = (1, v_{1,t}, \dots, v_{k-1,t})^\top \in \mathbb{R}^k$,
- $\mathbf{X}_k = (\hat{\mathbf{x}}_{k,1}, \dots, \hat{\mathbf{x}}_{k,T})^\top \in \mathbb{R}^{T \times k}$,
- $\mathbf{Y}_{\bullet,k} = (Y_{k,1}, \dots, Y_{k,T})^\top \in \mathbb{R}^T$.

We can set the penalized regression problem

$$\arg \min_{\mu_k, \mathbf{L}_{k,\bullet}} \left\| \mathbf{Y}_{\bullet,k} - \mathbf{X}_k \begin{pmatrix} \mu_k, \mathbf{L}_{k,\bullet}^\top \end{pmatrix}^\top \right\|_2^2 + \eta_k \left\| \begin{pmatrix} \mu_k, \mathbf{L}_{k,\bullet}^\top \end{pmatrix}^\top \right\|_1,$$

with $\|\mathbf{x}\|_1 = \sum_{j=1}^k |x_j|$ the ℓ_1 -norm on $\mathbb{R}^k$, and $\eta_k > 0$ a penalty strength parameter. This is simply a LASSO regression (Tibshirani, 1996) and allows for automatically selecting the relevant factors and estimating the parameters.

A.2 GHyp Errors

We utilize the algorithm developed in Näf et al. (2019) with some modifications. First, Näf et al. (2019) consider potentially different regularization parameters for each regression. Instead, we propose to use different regularization parameters within each of the K regressions depending on whether one regresses on the factors or the returns.

Thus, the CM1-Step is of particular interest: Let for $\hat{G}_{1,t}^{-1}, \dots, \hat{G}_{K,t}^{-1}$, $t = 1, \dots, T$ given by the E-Step below, $\kappa_{k,t}^2 := c_{kk}^{-1} \hat{G}_{k,t}^{-1}$. Moreover, define $\mathbf{1}_k$ to be a vector of ones of length k and similarly with $\mathbf{0}_k = \mathbf{0}_{k \times 1}$. The CM1-step is, for each $k = 1, \dots, K$, simply a regularized regression problem of the dependent variable $\kappa_{k,t} y_{k,t}$ onto the exogenous variables $\kappa_{k,t}$ and $\kappa_{k,t} v_{j,t}$, $j = 1, \dots, k-1$, with regularizer vector $\boldsymbol{\eta}_k = (\eta_{1,k}, \dots, \eta_{k-1,k})$. In particular, we define

- $\mathbf{L}_{k,\bullet} = (q_{k,1}, \dots, q_{k,k-1})^\top$ the first $k-1$ elements of k th row of $\mathbf{L}$ written as a column vector,
- $\hat{\mathbf{v}}_j = (v_{1,j}, \dots, v_{T,j})^\top$, the vector of residuals, with $j \leq k-1$,
- $\mathbf{B}_k = (\mathbf{1}_T, \hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_{k-1}) \in \mathbb{R}^{T \times k}$,
- $\hat{\boldsymbol{\kappa}}_k = c_{kk}^{-1/2} \left((\hat{G}_{k,1}^{-1})^{1/2}, \dots, (\hat{G}_{k,T}^{-1})^{1/2} \right)^\top \in \mathbb{R}^T$,
- $\mathbf{Y}_{\bullet,k} = (Y_{k,1}, \dots, Y_{k,T})^\top \in \mathbb{R}^T$.

Then, we solve the regression problem

$$\arg \min_{\mu_k, \mathbf{L}_{k,\bullet}} \left\| \hat{\boldsymbol{\kappa}}_k \odot \left(\mathbf{Y}_{\bullet,k} - \mathbf{B}_k \begin{pmatrix} \mu_k, \mathbf{L}_{k,\bullet}^\top \end{pmatrix}^\top \right) \right\|_2^2 + \left\| \boldsymbol{\eta}_k \odot \begin{pmatrix} \mu_k, \mathbf{L}_{k,\bullet}^\top \end{pmatrix}^\top \right\|_1,$$

with $\|\mathbf{x}\|_1 = \sum_{j=1}^k |x_j|$ the ℓ_1 -norm on $\mathbb{R}^k$ and $\odot$ the Hadamard product.

Second, we want to emphasize the possibility to impose additional structure on $\mathbf{L}$. This can easily be done in the ECME algorithm by excluding some potential regressors in the CM1 step. Consequently, a condition such as $\mathbf{L}^{(f)} = \mathbf{I}$ can be enforced, by regressing factors only on a constant and not on other factors. This corresponds to using prior knowledge to induce parsimony and decrease estimation error. As such, it is equivalent to ℓ_1 shrinkage, with the selected $\eta_{j,k}$ going to infinity. We add this formally, by simply considering $\mathbf{L}_{k,\bullet} \odot \boldsymbol{\Omega}_k$, where $\boldsymbol{\Omega}_k$ is a vector of 0's and 1's of size $k-1$. For instance, in the case

$\mathbf{L}^{(f)} = \mathbf{I}$, we take

$$\begin{aligned}\Omega_2 &= 0 \\ &\vdots \\ \Omega_{K_f} &= \mathbf{0}_{K_f-1}^\top \\ \Omega_{K_f+1} &= \mathbf{1}_{K_f}^\top \\ &\vdots \\ \Omega_K &= \mathbf{1}_{K-1}^\top.\end{aligned}$$

This leads to the following algorithm:

E-step: For $k = 1, \dots, K, t = 1, \dots, T$, calculate

$$\hat{G}_{k,t}^{-1} = \mathbb{E} \left[G_{k,t}^{-1} \mid v_{k,t} \right] = k_{\lambda_k-1/2-1} \left(\delta_k^2 + v_{k,t}^2/c_{kk}, \alpha_k^2 \right) / k_{\lambda_k-1/2} \left(\delta_k^2 + v_{k,t}^2/c_{kk}, \alpha_k^2 \right).$$

CM1-step: Update θ_P by solving

$$\arg \min_{\mu_k, (\mathbf{L}_{k,\bullet} \odot \Omega_k)} \left\| \hat{\kappa}_k \odot \left(\mathbf{Y}_{\bullet,k} - \mathbf{B}_k \left(\mu_k, (\mathbf{L}_{k,\bullet} \odot \Omega_k)^\top \right)^\top \right) \right\|_2^2 + \left\| \boldsymbol{\eta}_k \odot \left(\mu_k, (\mathbf{L}_{k,\bullet} \odot \Omega_k)^\top \right)^\top \right\|_1, \quad (36)$$

and calculating

$$c_{kk} = \frac{1}{T} \sum_{t=1}^T \hat{G}_{k,t}^{-1} v_{k,t}^2, \quad (37)$$

for $k = 1, \dots, K$.

CM2-step: Given the CM1-step updates of θ_P , obtain new updates of θ_D by maximizing the incomplete data log-likelihood function:

$$\arg \max_{\theta_D} \left\{ \sum_{t=1}^T \ln L_{\mathbf{Y}_t}(\theta_D, \hat{\theta}_P) \right\} = \sum_{k=1}^K \sum_{t=1}^T \ln f_{\text{GHyp}}(v_{k,t}; \theta_D, \hat{\theta}_P). \quad (38)$$

Iterate the above steps until convergence.

A.3 Hybrid Models

We repeat the formalization from the main text: We assume there exists $j = 1, \dots, d$ independent random variables $\mathbf{G}_{1,t}, \dots, \mathbf{G}_{d,t}$ such that either $\mathbf{G}_{j,t} \sim \text{GIG}(\lambda_j, \alpha_j, \delta_j)$ for all t or $\mathbf{G}_{j,t} = 1$ for all t . Let for the following $\mathcal{K}_0, \dots, \mathcal{K}_d$ be defined as in (24) and define the vector $\mathbf{v}_j = (v_{k,t})_{k \in \mathcal{K}_j}$ for $j = 0, \dots, d$.

This leads to the following algorithm:

E-step: For $j = 0, \dots, d, t = 1, \dots, T$, calculate $(\hat{G}_{k,t}^{-1})_{k \in \mathcal{K}_j} = \mathbb{E} \left[(G_{k,t}^{-1})_{k \in \mathcal{K}_j} \mid (v_{k,t})_{k \in \mathcal{K}_j} \right]$, where

- $\mathbb{E} \left[(G_{k,t}^{-1})_{k \in \mathcal{K}_j} \mid (v_{k,t})_{k \in \mathcal{K}_j} \right] = \mathbf{1}_{|\mathcal{K}_j|} \text{ if } j = 0$
- $\mathbb{E} \left[(G_{k,t}^{-1})_{k \in \mathcal{K}_j} \mid (v_{k,t})_{k \in \mathcal{K}_j} \right] = \mathbf{1}_{|\mathcal{K}_j|} \cdot \mathbb{E} \left[\mathbf{G}_{j,t}^{-1} \mid (v_{k,t})_{k \in \mathcal{K}_j} \right] = \mathbf{1}_{|\mathcal{K}_j|} \cdot \mathbf{k}_{\lambda_{\mathcal{K}_j}-1} \left(\delta_{\mathcal{K}_j}^2, \alpha_j^2 \right) / \mathbf{k}_{\lambda_{\mathcal{K}_j}} \left(\delta_{\mathcal{K}_j}^2, \alpha_j^2 \right),$

with $\lambda_{\mathcal{K}_j} = \lambda_j - |\mathcal{K}_j|/2$, $\delta_{\mathcal{K}_j} = \sqrt{\delta_j + \mathbf{v}_j^\top \mathbf{C}^{-1} \mathbf{v}_j}$.

CM1-step: Update $\boldsymbol{\theta}_P$ by solving

$$\arg \min_{\mu_k, \mathbf{L}_{k,\bullet}} \left\| \hat{\mathbf{k}}_k \odot \left(\mathbf{Y}_{\bullet,k} - \mathbf{B}_k \left(\mu_k, \mathbf{L}_{k,\bullet}^\top \right)^\top \right) \right\|_2^2 + \left\| \boldsymbol{\eta}_k \odot \left(\mu_k, \mathbf{L}_{k,\bullet}^\top \right)^\top \right\|_1, \quad (39)$$

and calculating

$$c_{kk} = \frac{1}{T} \sum_{t=1}^T \hat{G}_{k,t}^{-1} v_{k,t}^2, \quad (40)$$

for $k = 1, \dots, K$.

CM2-step: Given the CM1-step updates of $\boldsymbol{\theta}_P$, obtain new updates of $\boldsymbol{\theta}_D$ by maximizing the incomplete data log-likelihood function:

$$\begin{aligned} \sum_{t=1}^T \ln L_{\mathbf{Y}_t}(\boldsymbol{\theta}_D, \hat{\boldsymbol{\theta}}_P) &= \sum_{t=1}^T \left(\sum_{k \in \mathcal{K}_0} \ln f_N(v_{k,t}; \boldsymbol{\theta}_D, \hat{\boldsymbol{\theta}}_P) \right. \\ &\quad \left. + \sum_{j=1}^d \ln f_{\text{MGHyp}}((v_{k,t})_{k \in \mathcal{K}_j}; \boldsymbol{\theta}_D, \hat{\boldsymbol{\theta}}_P) \right), \end{aligned} \quad (41)$$

where f_N, f_{MGHyp} are the densities of the Gaussian and MGHyp distributions respectively.

B Proofs

Lemma 1 Let $\mathbf{X}, \mathbf{Y}$ be random vectors in $\mathbb{R}^{K_f}$ and $\mathbb{R}^{K_r}$ respectively, such that $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L})$, with $K = K_f + K_r$. Then the mgf of $\mathbf{Y}$ is given as

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) = \exp(\mathbf{u}^\top \boldsymbol{\mu}^{(\mathbf{Y})}) \prod_{i=1}^K \left(\frac{\alpha_i^2}{\alpha_i^2 - \tilde{u}_i^2} \right)^{\lambda_i/2} \frac{K_{\lambda_i}(\delta_i \sqrt{\alpha_i^2 - \tilde{u}_i^2})}{K_{\lambda_i}(\delta_i \alpha_i)}, \quad (20)$$

with, for $i = 1, \dots, K$,

$$\tilde{u}_i = \begin{cases} c_{ii,t}^{1/2} \sum_{\ell=k+1}^K u_\ell q_{\ell i}, & \text{if } i < k+1, \\ c_{ii,t}^{1/2} \left(u_i + \sum_{\ell=i+1}^K u_\ell q_{\ell i} \right), & \text{if } i \geq k+1. \end{cases}$$

PROOF First, it is well-known that

$$\begin{aligned}\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) &= \mathbb{E} [\exp(\mathbf{u}^\top \mathbf{Y})] \\ &= \mathbb{E} [\exp(\mathbf{u}^\top \mathbf{A}_2 \mathbf{Z})] \\ &= \mathbb{M}_{\mathbf{Z}}(\mathbf{A}_2^\top \mathbf{u}).\end{aligned}$$

Moreover, $\mathbb{M}_{\mathbf{Z}}$ was given in [Näf et al. \(2019\)](#), as

$$\mathbb{M}_{\mathbf{Z}}(\mathbf{v}) = \exp(\mathbf{v}^\top \boldsymbol{\mu}) \prod_{i=1}^K \left(\frac{\alpha_i^2}{\alpha_i^2 - \tilde{v}_i^2} \right)^{\lambda_i/2} \frac{K_{\lambda_i}(\delta_i \sqrt{\alpha_i^2 - \tilde{v}_i^2})}{K_{\lambda_i}(\delta_i \alpha_i)},$$

where for $\mathbf{v} = (v_1, \dots, v_K) \in \mathbb{R}^K$

$$\tilde{v}_i = c_{ii,t}^{1/2} \left(v_i + \sum_{k=i+1}^K v_k q_{ki} \right), \quad i = 1, \dots, K.$$

The result then follows by using $\mathbf{v} = (\mathbf{0}_{K_f \times 1}, \mathbf{u})$ in $\mathbb{M}_{\mathbf{Z}}$.

Lemma 2 Define for $\mathbf{v} = (v_1, \dots, v_K) \in \mathbb{R}^K$,

$$\tilde{v}_i = c_{ii}^{1/2} \left(v_i + \sum_{k=i+1}^K v_k q_{ki} \right), \quad i = 1, \dots, K,$$

and

$$\omega_j = \left(\sum_{i \in \mathcal{K}_j} \tilde{v}_i^2 \right)^{1/2}. \quad (25)$$

The moment generating function of $\mathbf{Z} \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L}, \mathcal{K})$ is given by

$$\mathbb{M}_{\mathbf{Z}}(\mathbf{v}) = \exp(\mathbf{v}^\top \boldsymbol{\mu}) \exp(\omega_0^2/2) \cdot \prod_{j=1}^d \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right)^{\lambda_j/2} \frac{K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2})}{K_{\lambda_j}(\delta_j \alpha_j)}, \quad (26)$$

for $\mathbf{v}$ such that $\omega_j \in (-\alpha_j, \alpha_j)$ for all j . This leads to the cumulant generating function

$$\mathbb{K}_{\mathbf{Z}}(\mathbf{v}) = \mathbf{v}^\top \boldsymbol{\mu} + \frac{\omega_0^2}{2} + \sum_{j=1}^d \left\{ \frac{\lambda_j}{2} \ln \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right) + \ln K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2}) - \ln K_{\lambda_j}(\delta_j \alpha_j) \right\}.$$

PROOF It holds that $\mathbf{v}^\top (\mathbf{Z} - \boldsymbol{\mu}) = \mathbf{v}^\top \mathbf{L} \mathbf{C}^{1/2} \mathbf{D}^{1/2} \boldsymbol{\epsilon}$. Then, first

$$\mathbf{D}^{1/2} \boldsymbol{\epsilon} = \begin{pmatrix} G_{11}^{1/2} \epsilon_1 \\ G_{22}^{1/2} \epsilon_2 \\ G_{33}^{1/2} \epsilon_3 \\ \vdots \\ G_{KK}^{1/2} \epsilon_K \end{pmatrix},$$

and

$$\begin{aligned} \mathbf{v}^\top \mathbf{L} \mathbf{C}^{1/2} &= \begin{pmatrix} v_1 & \dots & v_K \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ q_{21} & 1 & 0 & \dots & 0 \\ q_{31} & q_{32} & 1 & \dots & 0 \\ & \ddots & & & 0 \\ q_{K1} & q_{K2} & q_{K3} & \dots & 1 \end{pmatrix} \begin{pmatrix} c_{11}^{1/2} & 0 & 0 & \dots & 0 \\ 0 & c_{22}^{1/2} & 0 & \dots & 0 \\ 0 & 0 & c_{33}^{1/2} & \dots & 0 \\ & \ddots & & & 0 \\ 0 & 0 & 0 & \dots & c_{KK}^{1/2} \end{pmatrix} \\ &= \left(c_{11}^{1/2} \left(v_1 + \sum_{k=2}^K v_k q_{k1} \right) \quad c_{22}^{1/2} \left(v_2 + \sum_{k=3}^K v_k q_{k2} \right) \quad c_{33}^{1/2} \left(v_3 + \sum_{k=4}^K v_k q_{k3} \right) \quad \dots \quad c_{KK}^{1/2} v_K \right). \end{aligned}$$

It follows, with

$$\tilde{v}_i = c_{ii}^{1/2} \left(v_i + \sum_{k=i+1}^K v_k q_{ki} \right), \quad i = 1, \dots, K,$$

$$\begin{aligned} \mathbf{v}^\top (\mathbf{Z} - \boldsymbol{\mu}) &= \sum_{i=1}^K \tilde{v}_i G_i^{1/2} \epsilon_i \\ &= \sum_{i \in \mathcal{K}_0} \tilde{v}_i \epsilon_i + \sum_{j=1}^d \mathbf{G}_j^{1/2} \sum_{i \in \mathcal{K}_j} \tilde{v}_i \epsilon_i \\ &\stackrel{d}{=} \left(\sum_{i \in \mathcal{K}_0} \tilde{v}_i^2 \right)^{1/2} N(0, 1) + \sum_{j=1}^d \left(\sum_{i \in \mathcal{K}_j} \tilde{v}_i^2 \right)^{1/2} N(0, \mathbf{G}_j) \\ &:= \boldsymbol{\omega}^\top \boldsymbol{\xi}, \end{aligned}$$

where $\stackrel{d}{=}$ denotes equality in distribution. Thus $\mathbf{v}^\top \mathbf{Z} - \mathbf{v}^\top \boldsymbol{\mu}$ has the same distribution as a weighted sum of $|\mathcal{K}_0|$ independent Gaussians and d independent GHyp random variables, with weights

$$\omega_j = \left(\sum_{i \in \mathcal{K}_j} \tilde{v}_i^2 \right)^{1/2}, \quad j = 0, \dots, d.$$

The mgf of a single GHyp random variable, $Z \sim \text{GHyp}(\lambda, \alpha, \delta, \mu)$ is given as

$$\mathbb{M}_Z(u) = \mathbb{E}[\exp(iuZ)] = e^{\mu iu} \left(\frac{\alpha^2}{\alpha^2 - u^2} \right)^{\lambda/2} \frac{K_\lambda(\delta \sqrt{\alpha^2 - u^2})}{K_\lambda(\delta \alpha)}. \quad (42)$$

It follows from the above that the mgf is given by

$$\begin{aligned}\mathbb{M}_{\mathbf{Z}}(\mathbf{v}) &= \mathbb{E}[\exp(\mathbf{v}^\top \mathbf{Z})] = \exp(\mathbf{v}^\top \boldsymbol{\mu}) \mathbb{E}[\exp(\boldsymbol{\omega}^\top \boldsymbol{\xi})] \\ &= \exp(\mathbf{v}^\top \boldsymbol{\mu}) \exp(\omega_0^2/2) \cdot \prod_{j=1}^d \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right)^{\lambda_j/2} \frac{K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2})}{K_{\lambda_j}(\delta_j \alpha_j)},\end{aligned}$$

for $\mathbf{w}$ such that $\omega_j \in (-\alpha_j, \alpha_j)$ for all j .

Lemma 3 *Let $\mathbf{X}, \mathbf{Y}$ be random vectors in $\mathbb{R}^{K_f}$ and $\mathbb{R}^{K_r}$ respectively, such that $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \sim \text{HGH}(\boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{L}, \mathcal{K})$, with $K = K_f + K_r$. Then the mgf of $\mathbf{Y}$ is given as*

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) = \exp(\mathbf{u}^\top \boldsymbol{\mu}) \exp(\omega_0^2/2) \cdot \prod_{j=1}^d \left(\frac{\alpha_j^2}{\alpha_j^2 - \omega_j^2} \right)^{\lambda_j/2} \frac{K_{\lambda_j}(\delta_j \sqrt{\alpha_j^2 - \omega_j^2})}{K_{\lambda_j}(\delta_j \alpha_j)},$$

where for $j = 0, \dots, d$, ω_j is defined as in (25) with, for $i = 1, \dots, K$,

$$\tilde{u}_i = \begin{cases} c_{ii,t}^{1/2} \sum_{\ell=k+1}^K u_\ell q_{\ell i}, & \text{if } i < k+1, \\ c_{ii,t}^{1/2} \left(u_i + \sum_{\ell=i+1}^K u_\ell q_{\ell i} \right), & \text{if } i \geq k+1. \end{cases} \quad (27)$$

PROOF As before, since

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{u}) = \mathbb{M}_{\mathbf{Z}}(\mathbf{A}_2^\top \mathbf{u}),$$

with $\mathbb{M}_{\mathbf{Z}}$ given in (26), the result follows by using $\mathbf{v} = \mathbf{A}_2^\top \mathbf{u} = (\mathbf{0}, \mathbf{u})$ as an input in $\mathbb{M}_{\mathbf{Z}}$.

C Further Performance Statistics

Figure 2: Gross percentage long-only cumulative portfolio returns for the 41 cryptocurrencies (excluding PAX). The cryptocurrencies contain hourly observations from 01-01-2021 until 16-06-2021. The Factor-HGH model is estimated with $L^{(r)}$ equal to the identity matrix.

Table 3: Comparison of the gross performance for mean-expected shortfall portfolios for *Classical Models*, *Gaussian-Cholesky Models* and *Factor-HGH Models*. * in the top row denotes annualized quantities. FF-3 - Fama-French three factors; + MOM - additional momentum factor; FF-5 - Fama-French five factors; $\square$ denotes an identity matrix, $\blacksquare$ denotes lower triangular matrix. All portfolios are long-only. Dataset: daily returns of 49 industry portfolios from 03-01-2000 until 28-05-2021. Rebalancing every 5 days (weekly).

	<i>Return*</i>	<i>Volatility*</i>	<i>Total Return</i>	<i>Max. Drawdown</i>	<i>Turnover*</i>	<i>Sharpe*</i>	<i>Sortino*</i>	<i>STARR_{98.5%}%*</i>	<i>ES_{98.5%}*</i>
<i>Classical Models</i>									
Intercept-Only	17.34	14.96	2710.22	-45.73	943.05	1.16	1.65	5.27	328.72
FF-3	17.57	14.09	2922.93	-47.84	788.28	1.25	1.76	5.50	319.67
FF-3 + MOM	16.87	14.17	2513.80	-49.48	778.02	1.19	1.67	5.24	322.12
FF-5	17.87	13.89	3135.25	-44.84	790.88	1.29	1.83	5.66	315.93
FF-5 + MOM	17.38	14.01	2811.60	-45.98	791.58	1.24	1.75	5.45	318.95
<i>Gaussian-Cholesky Models</i>									
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	17.86	15.14	3011.24	-45.38	953.25	1.18	1.68	5.33	334.92
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	17.83	15.12	2991.01	-44.38	977.84	1.18	1.68	5.33	334.25
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	17.49	14.98	2797.57	-47.23	874.60	1.17	1.66	5.31	329.39
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.34	14.96	2710.82	-45.76	943.87	1.16	1.65	5.27	328.80
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	19.41	15.39	4143.04	-41.56	1035.90	1.26	1.80	5.69	341.20
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	19.13	15.36	3913.76	-41.52	1080.78	1.25	1.78	5.62	340.24
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	17.58	14.98	2849.67	-46.53	893.21	1.17	1.67	5.34	329.33
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.34	14.96	2710.82	-45.76	943.87	1.16	1.65	5.27	328.80
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	18.16	15.30	3189.44	-45.79	965.92	1.19	1.69	5.34	339.88
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	17.93	15.29	3039.44	-44.60	987.86	1.17	1.67	5.28	339.30
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	17.66	14.97	2901.16	-46.86	881.78	1.18	1.68	5.37	328.67
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.34	14.96	2710.76	-45.76	943.88	1.16	1.65	5.27	328.79
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	19.64	15.49	4335.41	-40.70	1041.55	1.27	1.81	5.69	345.00
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	19.27	15.49	4007.00	-40.97	1076.14	1.24	1.78	5.59	344.92
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	17.74	14.97	2947.96	-46.38	893.79	1.18	1.69	5.40	328.62
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.34	14.96	2710.82	-45.76	943.87	1.16	1.65	5.27	328.80
<i>Factor-HGH Models</i>									
No-Factors	16.79	18.04	2152.58	-60.30	2013.54	0.93	1.31	4.66	360.42
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	17.21	14.74	2650.67	-50.96	925.94	1.17	1.64	5.14	334.77
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	17.49	15.04	2791.68	-51.71	2149.21	1.16	1.64	5.10	343.06
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	16.87	14.52	2483.27	-51.66	832.68	1.16	1.63	5.13	328.69
FF-3 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.54	14.75	2848.12	-51.78	2007.26	1.19	1.67	5.23	335.78
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	18.24	14.82	3295.26	-48.55	1008.56	1.23	1.72	5.36	340.51
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	17.48	15.30	2761.54	-51.55	2232.99	1.14	1.60	4.97	351.54
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	16.90	15.13	2451.38	-53.61	896.00	1.12	1.56	5.02	336.41
FF-3 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	17.29	15.27	2651.20	-52.60	2077.93	1.13	1.59	5.01	345.22
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	17.24	15.13	2634.25	-52.97	1014.17	1.14	1.59	4.99	345.43
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	16.70	15.33	2334.40	-52.30	2243.28	1.09	1.52	4.75	351.62
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	16.73	14.52	2410.52	-52.11	895.51	1.15	1.62	5.10	328.24
FF-5 & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	16.48	14.89	2259.49	-52.48	2029.01	1.11	1.54	4.85	339.96
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \square)$	18.01	15.10	3110.50	-48.67	1113.58	1.19	1.66	5.15	349.45
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\square, \blacksquare)$	17.24	15.48	2608.06	-54.00	2394.55	1.11	1.55	4.83	356.88
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \square)$	16.95	15.20	2474.79	-53.58	926.18	1.12	1.55	5.03	337.30
FF-5 + MOM & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\blacksquare, \blacksquare)$	16.57	15.35	2267.23	-57.20	2163.83	1.08	1.50	4.76	347.82

Table 4: Comparison of the gross performance for mean-expected shortfall portfolios for *Classical Models*, *Gaussian-Cholesky Models* and *Factor-HGH Models*. * in the top row denotes annualized quantities. CMkt - return on the market factor; $\begin{smallmatrix} \square \\ \square \end{smallmatrix}$ denotes an identity matrix, $\begin{smallmatrix} \square \\ \blacksquare \end{smallmatrix}$ denotes lower triangular matrix. All portfolios are long-only. Dataset: hourly returns of 42 cryptocurrencies from 01-01-2021 until 16-06-2021. Rebalancing every 5 hours. Note: because we only have one factor, the cases where $L^{(f)} = \begin{smallmatrix} \square \\ \square \end{smallmatrix}$ and $L^{(f)} = \begin{smallmatrix} \square \\ \blacksquare \end{smallmatrix}$ are identical. So the lower triangular one is dropped from the table.

	Return*	Volatility*	Total Return	Max. Drawdown	Turnover*	Sharpe*	Sortino*	STARR _{98.5%} %*	ES _{98.5%} *
<i>Classical Models</i>									
Intercept-Only	12.32	8.55	5.30	-5.62	104.35	1.44	1.93	1.35	26.28
CAPM	15.46	10.71	6.64	-8.98	128.84	1.44	1.87	1.33	33.35
<i>Gaussian-Cholesky Models</i>									
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix})$	12.15	10.00	5.16	-8.47	124.12	1.21	1.58	1.17	29.81
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \blacksquare \end{smallmatrix})$	18.18	8.60	8.00	-5.62	110.98	2.11	2.85	2.00	26.12
<i>Factor-HGH Models</i>									
No-Factors	21.16	9.72	9.35	-3.11	99.52	2.18	3.28	2.40	25.32
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix})$	25.50	10.04	11.40	-2.90	80.63	2.54	3.94	2.79	26.28
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \blacksquare \end{smallmatrix})$	19.10	9.66	8.38	-3.12	97.52	1.98	2.95	2.15	25.51

Table 5: Comparison of the gross performance for mean-expected shortfall portfolios for *Classical Models*, *Gaussian-Cholesky Models* and *Factor-HGH Models*. * in the top row denotes annualized quantities. CMkt - return on the market factor; $\begin{smallmatrix} \square \\ \square \end{smallmatrix}$ denotes an identity matrix, $\begin{smallmatrix} \square \\ \blacksquare \end{smallmatrix}$ denotes lower triangular matrix. All portfolios are long-only. Dataset: hourly returns of 41 cryptocurrencies (excluding PAX) from 01-01-2021 until 16-06-2021. Rebalancing every 5 hours. Note: because we only have one factor, the cases where $L^{(f)} = \begin{smallmatrix} \square \\ \square \end{smallmatrix}$ and $L^{(f)} = \begin{smallmatrix} \square \\ \blacksquare \end{smallmatrix}$ are identical. So the lower triangular one is dropped from the table.

	Return*	Volatility*	Total Return	Max. Drawdown	Turnover*	Sharpe*	Sortino*	STARR _{98.5%} %*	ES _{98.5%} *
<i>Classical Models</i>									
Intercept-Only	51.30	88.62	5.04	-75.44	965.43	0.58	0.75	0.47	312.76
CAPM	186.19	65.80	103.44	-51.06	749.78	2.83	3.76	2.00	268.43
<i>Gaussian-Cholesky Models</i>									
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix})$	73.43	89.72	15.08	-74.82	975.09	0.82	1.05	0.65	322.83
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \blacksquare \end{smallmatrix})$	51.37	88.63	5.06	-75.45	965.95	0.58	0.75	0.47	312.84
<i>Factor-HGH Models</i>									
No-Factors	270.12	86.72	173.21	-47.04	977.95	3.12	4.62	2.54	305.69
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix})$	263.51	90.01	162.26	-46.56	599.67	2.93	4.36	2.39	317.44
CMkt & $(\mathbf{L}^{(f)}, \mathbf{L}^{(r)}) = (\begin{smallmatrix} \square & \square \\ \square & \blacksquare \end{smallmatrix})$	242.96	85.93	143.64	-49.17	841.54	2.83	4.14	2.29	305.37

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public-private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.