Risk Parity Portfolio Optimization under Heavy-Tailed Returns and Time-Varying Volatility

Marc S. Paolella[a,b] Paweł Polak[c,d] Patrick S. Walker[a,e*‡]

[a] *Department of Banking and Finance, University of Zurich, Zurich, Switzerland*

[b] *Swiss Finance Institute, Zurich, Switzerland*

[c] *Department of Applied Mathematics and Statistics, Stony Brook University, United States*

[d] *Institute for Advanced Computational Science, Stony Brook University, United States*

[e] *OLZ AG, Zurich, Switzerland*

December 6, 2023

Abstract

Risk parity portfolio optimization, using expected shortfall as the risk measure, is investigated when asset returns are fat-tailed and heteroscedastic. The conditional return distribution is modeled by an elliptical multivariate generalized hyperbolic distribution, allowing for fast parameter estimation, via an expectation-maximization algorithm and a semi-closed form of the risk contributions. The efficient computation of non-Gaussian risk parity weights sidesteps the need for numerical simulations or Cornish-Fisher-type approximations. Accounting for fat-tailed returns, the risk parity allocation is less sensitive to volatility shocks, thereby generating lower portfolio turnover, in particular during market turmoils such as the global financial crisis. Although risk parity portfolios are surprisingly robust to the misuse of the Gaussian distribution, a more realistic model for conditional returns and time-varying volatilities can improve risk-adjusted returns, reduces turnover during periods of market stress and enables the use of a holistic risk model for portfolio and risk management.

Keywords: Elliptical Distributions; GARCH; Heavy-Tails; Multivariate Generalized Hyperbolic Distribution; Risk Parity.

JEL Classification: C51; C53; C55; G11.

*Corresponding author's address: Plattenstrasse 14, CH-8032 Zurich, Switzerland; patrick.walker@bf.uzh.ch

†This research was partially conducted while Walker was a visiting scholar at the Statistics Department at Columbia University. He is grateful for the financial support by the Swiss National Science Foundation (SNSF) through the Doc.Mobility fellowship no. P1ZHP1171760. The funding source had no involvement in this work.

‡We wish to thank Simon Broda, Gianluca De Nard, Danqiao Guo, Claudio Morana and Karl Schmedders, as well as the participants of the 2nd International Conference on Econometrics and Statistics in Hong Kong for valuable feedback and questions on the paper. Particular gratitude goes to Antonio Polino, who helped greatly with the conception and implementation of the risk parity optimization algorithm.

1 Introduction

The seminal mean-variance portfolio of Markowitz (1952, 1959) is known to suffer when the unknown population parameters are replaced by statistical estimates, see, among other, the works of Barry (1974), Best and Grauer (1991), Frankfurter et al. (1971), Frost and Savarino (1986, 1988), and Michaud (1989). Small estimation errors can translate into large deviations of the estimated portfolio weights from the true, optimal ones, ultimately leading to mediocre performance. This fragility makes real world applications of the mean-variance portfolio delicate. Various approaches have been proposed to improve the quality and stability of optimized mean-variance portfolios. Among them are constraining the portfolio weights (DeMiguel et al., 2009a; Frost and Savarino, 1988; Jagannathan and Ma, 2003); robust optimization (Goldfarb and Iyengar, 2003; Tütüncü and Koenig, 2004); resampling methods (Michaud and Michaud, 2008a,b); Bayesian methods (Frost and Savarino, 1986; Pástor and Stambaugh, 2000; Polson and Tew, 2000), as well as shrinkage estimation of the covariance matrix (Engle et al., 2017; De Nard et al., 2022; Ledoit and Wolf, 2003, 2004a,b, 2020).

A drastic but successful solution to the precarious estimation of expected returns is to forego it completely and to focus on minimizing the portfolio risk. The minimum-variance (short: min-var) solution constitutes the only portfolio on the efficient frontier that does not require the imputation of expected returns. In addition to avoiding a tricky estimation problem, interest in low-risk strategies has increased over the years due to their attractive ex-post risk-return profile; see, e.g., Ang et al. (2006), Ang (2014), Baker et al. (2011), Clarke et al. (2006), Haugen and Baker (1996), and Frazzini and Pedersen (2014). Despite this track record of high risk-adjusted returns, min-var optimization also faces criticism. Firstly, the portfolio weights still exhibit considerable sensitivity to changes in the covariance estimates, this problem often being addressed by imposing constraints in the solution space of the optimization problem. When no such ex-ante bounds are imposed on the weights, the min-var portfolio can exhibit large concentration in single assets that might not be acceptable for investors (Maillard et al., 2010). This issue becomes most pressing in times of financial distress, when asset correlations increase and diversification potential disappears. An unconstrained min-var portfolio then simply selects a small set of stocks with the lowest volatility. Secondly, a related limitation of mean- and min-variance portfolios, as noted in Paolella et al. (2021), is that when the portfolio is rebalanced, the variation in weights due to updated data often causes excessive turnover and intolerably high transaction costs.

In light of these challenges, other more robust objective functions have been proposed for portfolio optimization. One such approach, the concept of *risk parity portfolio allocation*, has been proposed in Maillard et al. (2010) and Qian (2005) as an allocation scheme less sensitive to estimation errors. The fundamental idea behind risk parity is to allocate risk across assets rather than capital, i.e., the goal is a portfolio allocation in which all assets contribute equally to the overall risk. Hence the strategy is also called equal risk contribution. In its simplest form, risk parity weighs each asset proportionally to the inverse of its volatility. While giving larger weights to less risky assets has been applied since decades by investors as a tool for position sizing and risk management, the risk parity approach has gained popularity in both the academic literature and among practitioners since the global financial crisis, see, e.g., Roncalli (2013), Unger (2014), and Qian (2011, 2016). When applied to multiple asset classes, such as stocks, bonds and commodities, risk parity has become a widely popular multi-asset strategy that offers an alternative to classical allocation rules, such as the simple 60/40 equity-bond portfolio.

One reason for the popularity of risk parity-based investment solutions is that this approach to forming a portfolio allocation is supposed to perform well in various market environments, notably during financial market crashes, such as the global financial crisis (Hurst et al., 2010). Against this backdrop, it might be surprising that the majority of studies on risk parity make use of the standard deviation as the risk measure and the multivariate normal distribution for modeling asset returns. These overly simplistic assumptions are violatead by the stylized facts of asset returns—in particular the exponentially decaying tails are too thin to adequately capture the occurrence of large price movements. Building a portfolio to provide protection during financial turmoil on the restrictive normality assumption might not be optimal. Therefore, the incorporation of a heavy-tailed distribution and the focus on an adequate tail risk measure when constructing the risk parity portfolio is a crucial task.

Studies that investigate risk parity under non-Gaussian return distributions are, among others, Mercuri and Rroji (2014), using the mixed tempered stable distribution class, and Stefanovits (2010), using the multivariate symmetric Student-t distribution to account for excess kurtosis. As a coherent and homogeneous tail risk measure (Acerbi and Tasche, 2002; Tasche, 2002), expected shortfall (short: ES, also known as CVaR) can be incorporated in the risk parity optimization framework, as is investigated, among others, in Cesarone and Colucci (2018), Jurczenko and Teiletche (2019), and Roncalli (2013). However, these studies rely on historical simulation, making the computation of optimal risk parity weights cumbersome or even infeasible in high dimensions. A Cornish-Fisher type approximation of ES is used in Boudt et al. (2012), whereas Mausser and Romanko (2018) present an efficient method to compute risk parity allocations in the case of a discrete scenario distribution. Recently, tail risk parity under a generalized Pareto distribution with a Vine-copula has been proposed in Gava et al. (2021). Again, the authors use Monte-Carlo simulation in computing the risk parity allocation and only consider a low-dimensional set of asset returns.

We propose to use a powerful parametric model that can capture the most important stylized facts of asset returns for risk parity optimization. Hence, we employ the COMFORT model of Paolella and Polak (2015a) for conditional return modeling under semi-heavy tails and time-varying volatility. As shown by these authors, this model class achieves superior predictive ability in forecasting the multivariate distribution of future returns, leading to improvements in risk prediction and tail-risk based portfolio optimization. The model is estimated via an iterative expectation-maximization (EM) algorithm that enables us to estimate the correlation structure on filtered, Gaussian data, thereby, ultimately, providing a more accurate and robust risk parity allocation.

The application of non-Gaussian distributions and tail risk measures in risk parity portfolios crucially hinges on the computational costs of determining the risk contributions. Numerical optimization of the objective function requires computing the risk contributions of each asset in every possible candidate portfolio, making this object the computational bottleneck. For non-Gaussian distributions, computation of the ES is often done by simulation and the calculation of risk contributions requires costly numerical gradient approximations. A major contribution of our work is an improved method to determine the optimal risk parity portfolio weights when the returns are modeled by an elliptical multivariate generalized hyperbolic distribution. By exploiting a closed form gradient expression together with a numerical Hessian approximation and a formula for fast evaluation of the ES from Paolella and Polak (2015b), the risk contributions and risk parity weights can be computed almost instantaneously, giving way to a highly efficient algorithm for non-Gaussian risk parity portfolio construction.

The remaining part of the paper is structured as follows. Section 2 introduces the COMFORT framework for conditional return modeling. Section 3 reviews the definition of risk contributions and the risk parity portfolio optimization problem. Next, Section 4 outlines the fast computation of the expected shortfall and the risk contributions under elliptical distributions, and how the numerical optimization of the risk parity objective function can be accelerated. Finally, Section 5 presents our empirical out-of-sample portfolio optimization analysis, whereas Section 6 concludes this work. Two Appendices gather technical details and further information on computational speed and accuracy.

2 Fat-Tailed Conditional Returns with Regime-Switching Conditional Correlations

Risk parity optimization under a symmetric Student-t distribution is investigated in Stefanovits (2010), showing favorable properties of this distribution for risk parity modeling. Our analysis extends the case to the more general class of elliptical multivariate generalized hyperbolic (MGHyp) distributions, which includes not just the aforementioned multivariate Student-t distribution but also the Laplace (or Variance Gamma) and the normal inverse Gaussian (NIG) distributions, both often used in modeling financial asset returns; see Eberlein and Keller (1995) and McNeil et al. (2015). Moreover, instead of assuming returns to be i.i.d., we employ a powerful model for conditional returns with allows for conditional heteroskedasticity via a multivariate GARCH model as in Paolella and Polak (2015a).

2.1 Multivariate Generalized Hyperbolic Distributions

A random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ has a MGHyp distribution, if it is a normal mean-variance mixture of the form

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\gamma} G + \sqrt{G} \boldsymbol{A} \boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^n$; $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a matrix of positive scalars; $\boldsymbol{Z} \stackrel{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_n)$; and $G \sim \mathrm{GIG}(\lambda, \chi, \psi)$ is a univariate mixing random variable independent of $\boldsymbol{Z}$ and has a generalized inverse Gaussian (GIG) density given by

$$f_G(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda} \left(\sqrt{\chi\psi}\right)^{\lambda}}{2\mathcal{K}_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0,$$

with constants $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. Here $\mathcal{K}_\lambda(x)$ is the modified Bessel function of the third kind, given by

$$\mathcal{K}_\lambda(x) = \frac{1}{2} \int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0.$$

We denote the distribution by $\boldsymbol{Y} \sim \mathrm{MGHyp}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ is the dispersion matrix. Due to the normal mixture characterization, the conditional distribution of $\boldsymbol{Y}$ given G is multivariate normal,

$$\boldsymbol{Y} \mid (G = g) \sim \mathrm{MN}(\boldsymbol{\mu} + \boldsymbol{\gamma} g, g\boldsymbol{\Sigma}).$$

This property makes the MGHyp distribution suitable for fast parameter estimation by algorithms of the expectation-maximization (EM) type; see McNeil et al. (2015), Protassov (2004), and the references therein.

2.2 Elliptical Distributions

Another general class of distributions that allow for excess kurtosis are elliptical probability distributions; see Fang et al. (1990), McNeil et al. (2015), and Paolella (2019, App. C) for general references. First, a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ is said to have a spherical distribution if, for every orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{n \times n}$, it holds $\boldsymbol{U}\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{Z}$. This is equivalent to $\boldsymbol{Z}$ having characteristic function

$$\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{Z})) = \psi(\boldsymbol{t}'\boldsymbol{t}) = \psi(t_1^2 + \cdots + t_n^2), \ \boldsymbol{t} \in \mathbb{R}^n,$$

for a scalar function ψ called the generator of $\boldsymbol{Z}$. We write $\boldsymbol{Z} \sim S_n(\psi)$. Then, a random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ is called elliptically distributed if it is a multivariate affine transformation of a spherically distributed random vector $\boldsymbol{Z}$, i.e.

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z},$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{Z} \sim S_n(\psi)$. The characteristic function of $\boldsymbol{Y}$ is of the form

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{Y})) = \mathbb{E}(\exp(i\boldsymbol{t}'(\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z})) = \exp(i\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}),$$

for $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ and we write $\boldsymbol{Z} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. The class of elliptical distributions is closed under linear combinations and all univariate marginal distributions are symmetric around their mean.

For $\boldsymbol{\gamma} = \mathbf{0}$, the MGHyp distribution in (2) is elliptical. In particular, all marginal distributions are symmetric in this case. The univariate generalized hyperbolic distribution $\text{GHyp}(\lambda, \chi, \psi, \mu, \gamma, \sigma)$, popular for modeling univariate asset returns, is attained for $n = 1$. Several special cases of the MGHyp distribution can be obtained when some of the GIG parameters are fixed prior to estimation. For $\lambda = 1$, $\psi > 0$ and $\chi = 0$, we get the multivariate asymmetric Laplace (MALap) distribution; for $\lambda = -\nu/2$, $\chi = \nu$, and $\psi = 0$, we have the asymmetric or skewed t-distribution (MAt) with ν degrees of freedom; and for $\lambda = -1/2$ and $\psi = 1$, the distribution is called normal inverse Gaussian (NIG). The symmetric variants are abbreviated MLap, Mt, and SNIG, respectively.

2.3 The COMFORT-CCC Model

Let $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$ denote a return vector of K financial assets at time t, for $t = 1, 2, \ldots, T$. The equally spaced realization of the return vector is denoted $\mathbf{Y} = [\mathbf{Y}_1 \mid \mathbf{Y}_2 \mid \cdots \mid \mathbf{Y}_T]$ and the information set at time t, defined as the sigma algebra generated by the history of returns $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_t\}$, is denoted $\mathbf{\Phi}_t$. In what follows, we assume that $\mathbf{Y}_t$ has a time varying conditional distribution with the COMFORT representation, from Paolella and Polak (2015a), given by

$$\begin{aligned} \mathbf{Y}_t \mid \mathbf{\Phi}_{t-1} &\stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \\ \boldsymbol{\varepsilon}_t &= \mathbf{H}_t^{1/2}\sqrt{G_t}\mathbf{Z}_t, \end{aligned} \tag{2}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $\mathbf{H}_t$ is a symmetric, positive definite dispersion matrix of order K; $\mathbf{Z}_t \stackrel{\text{iid}}{\sim} \text{N}(\mathbf{0}, \mathbf{I}_K)$ is a sequence of independent and identically distributed (i.i.d.) normal random variables; and $G_t \sim \text{GIG}(\lambda, \chi, \psi)$ is an i.i.d. mixing random variable, independent of $\mathbf{Z}_t$, with the generalized inverse Gaussian (GIG) density given by

$$f_G(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2\mathcal{K}_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0;$$

$\mathcal{K}_\lambda(x)$ is the modified Bessel function of the third kind, given by

$$\mathcal{K}_\lambda(x) = \frac{1}{2}\int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0;$$

and $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. The same MGHyp distribution arises from the parameter constellation $(\lambda, \chi/c, c\psi, \boldsymbol{\mu}, c\mathbf{H}_t, c\boldsymbol{\gamma})$ for any $c > 0$; hence for identification purposes, we fix either χ or ψ. The model assumes $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$, as well as the GIG parameters (λ, χ, ψ) to be time invariant. For the GIG distribution to be well defined, $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ and $\mathbb{E}\left[G_t^2 \mid \mathbf{\Phi}_{t-1}\right]$ have to be positive (Paolella and Polak, 2015a).
The symmetric, positive definite, conditional dispersion matrix $\mathbf{H}_t$ is decomposed as

$$\mathbf{H}_t = \mathbf{S}_t\mathbf{\Gamma}\mathbf{S}_t, \tag{3}$$

where $\mathbf{S}_t$ is a diagonal matrix holding the strictly positive conditional scale terms $s_{k,t}$, $k = 1, \ldots, K$, and $\mathbf{\Gamma}$ is a positive definite dependency matrix. The univariate scale terms $s_{k,t}$ are each modeled by a GARCH-type process. To describe the volatility clustering we use the GARCH(1,1) model defined by

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{4}$$

where $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$, is the kth element of the $\boldsymbol{\varepsilon}_t$ vector in (2), and $\omega_k > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, for $k = 1, 2, \ldots, K$.

In the COMFORT-CCC model, $\boldsymbol{\mu}$ and $\mathbf{H}_t$ are the location vector and the dispersion matrix of the conditional distribution of $\mathbf{Y}_t$, respectively, while the mean and the covariance matrix are given by

$$\begin{aligned} \mathbb{E}\left[\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}\right] &= \boldsymbol{\mu} + \mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right] \boldsymbol{\gamma}, \\ \operatorname{Cov}\left(\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}\right) &= \mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right] \mathbf{H}_t + \mathbb{V}\left(G_t \mid \mathbf{\Phi}_{t-1}\right) \boldsymbol{\gamma}\boldsymbol{\gamma}', \end{aligned}$$

where $\mathbb{V}\left(G_t \mid \mathbf{\Phi}_{t-1}\right) = \mathbb{E}\left[G_t^2 \mid \mathbf{\Phi}_{t-1}\right] - \left(\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]\right)^2$. Furthermore, by definition of the mean-variance mixture distribution of $\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}$, the matrix $\mathbf{\Gamma}$ is a correlation matrix, conditionally on the realization of the mixing process G_t. For convenience and to make the analogy to the Gaussian literature, we use the names dependency matrix and correlation matrix interchangeably for $\mathbf{\Gamma}$.

3 Risk Parity Portfolio Optimization

The main concept of risk parity portfolio allocation is to distribute risk budgets, instead of capital budgets, equally across different assets or asset classes. For this purpose a total risk budget, quantified using a given risk measure, is split equally between the assets. The risk measure has to satisfy the assumptions of Euler's decomposition theorem, i.e., must be a continuously differentiable, homogeneous function of degree one, in order to ensure that the sum of risk contributions of all assets equals the total portfolio risk. Two popular such risk measures are the variance and expected shortfall (ES). In practice, the volatility is almost exclusively used. However, using a tail risk measure such as ES, might be preferable in order to focus on the relevant tails of the return distribution. Applications of ES in risk parity portfolio construction are given, e.g., in Cesarone and Colucci (2018) and Roncalli (2013). While we focus on ES in our analysis, we state the problem of risk parity optimization for a general risk measure $\mathcal{R}$.

3.1 Risk Contributions

The construction of the risk parity portfolio requires that the overall risk can be split into the sum of the so-called risk contributions of the individual assets. If the risk measure $\mathcal{R}$ fulfills the assumptions of the Euler theorem, then the decomposition

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=1}^n x_i \frac{\partial \mathcal{R}(\boldsymbol{x})}{\partial x_i} = \sum_{i=1}^n \mathrm{RC}_i(\boldsymbol{x}), \quad \mathrm{RC}_i(\boldsymbol{x}) := x_i \frac{\partial \mathcal{R}(\boldsymbol{x})}{\partial x_i}$$

holds, where $\mathrm{RC}_i(\boldsymbol{x})$, $i = 1, \ldots, n$, is the so-called risk contribution of asset i to the portfolio with asset weights $\boldsymbol{x} = (x_1, \ldots, x_n)$. The risk contribution $\mathrm{RC}_i(\boldsymbol{x})$ therefore measures the marginal contribution of asset i to the total risk $\mathcal{R}(\boldsymbol{x})$ of the portfolio, multiplied by its weight. We consider only linear portfolios, so that, for a vector of log-returns $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ and a weight vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$, the profit or loss of the portfolio is given by $P(\boldsymbol{x}) := P := \boldsymbol{x}'\boldsymbol{Y}$.

In the special case of a multivariate normal return distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, the risk contributions are given in closed form as

$$\mathrm{RC}_i(\boldsymbol{x}) = x_i \frac{\partial \sigma(\boldsymbol{x})}{\partial x_i} = \frac{x_i (\mathbf{\Sigma}\boldsymbol{x})_i}{\sigma(\boldsymbol{x})}, \tag{5}$$

where $\mathcal{R}(\boldsymbol{x}) = \sigma(\boldsymbol{x}) = \sqrt{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}$ is the portfolio volatility and $(\boldsymbol{\Sigma}\boldsymbol{x})_i$ is the covariance between the portfolio return and the return on asset i.

3.2 Risk Parity Portfolios

The risk parity portfolio $\boldsymbol{x}_{\mathrm{RP}}$ is then defined by having equal risk contributions from all assets, i.e., the risk contributions must satisfy

$$\mathrm{RC}_i(\boldsymbol{x}_{\mathrm{RP}}) = \mathrm{RC}_j(\boldsymbol{x}_{\mathrm{RP}}),\ i, j = 1, \ldots, n. \tag{6}$$

In the Gaussian case with risk contributions (5), the risk parity portfolio must thus fulfill

$$x_i(\boldsymbol{\Sigma}\boldsymbol{x})_i = x_j(\boldsymbol{\Sigma}\boldsymbol{x})_j,\ i, j = 1, \ldots, n,$$

with the conditions $x_i \geq 0$ and $\sum_{i=1}^n x_i = 1$ encoding the no-shortselling and the fully-invested constraints. If the covariance matrix $\boldsymbol{\Sigma}$ is positive definite, then the Gaussian risk parity problem has a unique solution $\boldsymbol{x}^*_{\mathrm{RP}}$; see Maillard et al. (2010) and Spinu (2013).

Remarks:

1. An analytical solution of the risk parity weights is known only in the special case of a Gaussian distribution with all assets having identical correlations, i.e., $\rho_{ij} = \rho$ for all $i, j = 1, \ldots, n$. As shown in Maillard et al. (2010), the unique risk parity weights are then given by

$$\boldsymbol{x}^*_{\mathrm{RP},i} = \frac{\sigma_i^{-1}}{\sum_{j=1}^n \sigma_j^{-1}},\ i = 1, \ldots, n. \tag{7}$$

   In this simple case (also called naive risk parity), the allocation to each asset is proportional to its volatility. Incorporation of arbitrary distributions or other risk measures in naive risk parity is straightforward and fast as it does not require any numerical optimization.

2. Furthermore, it is shown in Maillard et al. (2010) that the risk parity portfolio is mean-variance efficient, only if in addition to the assumptions above, all assets have identical Sharpe ratios; see also Lee (2011). As these restrictive assumptions are not fulfilled in most real world applications, the risk parity allocation will usually not be an efficient portfolio.

By contrast, computing the optimal risk parity weights in the more general case requires optimization methods so that numerical efficiency is pivotal for large-dimensional asset universes. Furthermore, the existence and uniqueness of the risk parity portfolio is often not guaranteed, e.g., when deviating from the multivariate normal distribution or when the no-shortselling constraint is relaxed—see Kolm et al. (2014). In this case, one can find an approximate risk parity portfolio that fulfills the risk parity conditions more or less closely (Maillard et al., 2010; Roncalli, 2013). Such an approximation is attained by solving the optimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n_+ : |\boldsymbol{x}| = 1} f(\boldsymbol{x}), \tag{8}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is an auxiliary objective function that has a global minimum in $\boldsymbol{x}_{\mathrm{RP}}$. A numerical solution of problem (8) delivers an approximate risk parity portfolio whose risk contributions are close to parity. Two popular choices for the auxiliary function $f(\boldsymbol{x})$ are

$$f_1(\boldsymbol{x}) = \sum_{i,j=1}^n \left(\mathrm{RC}_i(\boldsymbol{x}) - \mathrm{RC}_j(\boldsymbol{x})\right)^2, \tag{9}$$

and

$$f_2(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\mathrm{RC}_i(\boldsymbol{x}) - \frac{\mathrm{ES}_\alpha(\boldsymbol{x})}{n} \right)^2, \tag{10}$$

see Maillard et al. (2010) and Roncalli (2013) for a discussion, as well as Cesarone and Colucci (2018) and Kolm et al. (2014) for further choices of the objective function. In the following, we use the objective $f_1(\boldsymbol{x})$ because of slightly faster convergence of the optimization routine. Solving this constrained non-linear optimization problem can be done, among other methods, with a sequential quadratic programming (SQP) algorithm because the objective function and the constraints are twice continuously differentiable. Nevertheless, the computations quickly become demanding when the risk contributions are expensive to evaluate and the dimension n is not small. For this reason, the fast computation of risk contributions is pivotal and limits the choice of non-Gaussian distributions for modeling returns in a risk parity strategy.

In the following Section 4, we consider elliptical distributions for which the risk contributions are given in closed form. Since elliptical MGHyp distributions offer a good trade-off between statistical flexibility and computational ease, we will consider this distributional setting for our analysis below.

4 Fast Computation of Tail Risk Parity Portfolio under Elliptical Distributions

4.1 Tail Risk Contributions under Elliptical Returns

We now turn to the computation of the risk contributions using ES as the risk measure. Denote by $P_i := x_i Y_i$ the return due to the i-th asset of portfolio $\boldsymbol{x} = (x_1, \ldots, x_n)$. In Tasche (2002) an expression for $\mathrm{RC}_i(\boldsymbol{x})$ under ES is given when the asset returns are continuously distributed. For any portfolio vector $\boldsymbol{x}$ it holds

$$\mathrm{RC}_i(\boldsymbol{x}) = \mathbb{E}(P_i \mid P(\boldsymbol{x}) \leq \mathrm{VaR}_\alpha(\boldsymbol{x})). \tag{11}$$

This result is intuitive as it states that the risk contribution of asset i is given by the asset's expected loss in the event that the portfolio return is below the VaR. Using the linearity of conditional expectations, we can easily check that

$$\sum_{i=1}^{n} \mathbb{E}(P_i \mid P(\boldsymbol{x}) \leq \mathrm{VaR}_\alpha(\boldsymbol{x})) = \mathbb{E}(P(\boldsymbol{x}) \mid P(\boldsymbol{x}) \leq \mathrm{VaR}_\alpha(\boldsymbol{x})) = \mathrm{ES}_\alpha(\boldsymbol{x}) = \mathcal{R}(\boldsymbol{x}).$$

Earlier, Gourieroux et al. (2000) showed an analogous result when $\mathcal{R}$ is the VaR, in which case

$$\mathrm{RC}_i(\boldsymbol{x}) = \mathbb{E}(P_i \mid P(\boldsymbol{x}) = \mathrm{VaR}_\alpha(\boldsymbol{x})).$$

Expression (11) requires the computation of the expected loss of the position x_i in asset i, conditional on the event that the portfolio return falls below the threshold $\mathrm{VaR}_\alpha(\boldsymbol{x})$. This expression is not easy to evaluate and usually simulation is used; see, among others, Roncalli (2013) and Stefanovits (2010). Fortunately, analytical formulas for the risk contribution have been derived for several distributions.

Starting with the case of multivariate normality, i.e., for return vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \sim \mathrm{MN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, $\boldsymbol{\Sigma} = (\Sigma_{ij})_{i,j=1,\ldots,n}$ positive definite and a portfolio vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$, we have a closed form for the portfolio ES

$$\mathrm{ES}_\alpha(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\mu} + \sqrt{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}} \cdot \frac{\phi(\Phi^{-1}(1-\alpha))}{\alpha}, \tag{12}$$

and computing the risk contributions amounts to differentiating this with respect to x_i

$$\mathrm{RC}_i(\boldsymbol{x}) = x_i \cdot \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} = x_i \left(\mu_i + \frac{(\boldsymbol{\Sigma x})_i}{\sigma(\boldsymbol{x})} \cdot \frac{\phi(\Phi^{-1}(1-\alpha))}{\alpha} \right), \tag{13}$$

with $\sigma(\boldsymbol{x}) = \sqrt{\boldsymbol{x}'\boldsymbol{\Sigma x}}$ the portfolio volatility and $(\boldsymbol{\Sigma x})_i$ the covariance between the return on asset i and the portfolio $\boldsymbol{x}$. This formula is reported, for example, in Dhaene et al. (2008, Ex. 2.5), Landsman and Valdez (2003), and Roncalli (2013, p. 80), with the original contribution due to Panjer (2001).

In the more flexible setting of elliptical distributions that allows for excess kurtosis there exists a close form solution for the risk contributions, originally derived in the context of aggregating risks within insurance firms, see Landsman and Valdez (2003). This formula is extremely useful for efficient computation of risk parity portfolios under elliptical returns because it only requires to compute the total ES of the portfolio, which vastly accelerates the evaluation of the objective function (8). Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \sim \mathrm{E}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ be the vector of asset returns and $\boldsymbol{x} = (x_1, \ldots, x_n)'$ the portfolio weights, then for $i = 1, \ldots, n$, and $\alpha \in (0,1)$, the risk contribution can be expressed as

$$\mathrm{RC}_i(\boldsymbol{x}) = x_i \left(\mu_i + \frac{(\boldsymbol{\Sigma x})_i}{\sigma^2(\boldsymbol{x})} (\mathrm{ES}_\alpha(\boldsymbol{x}) - \mu_P) \right), \tag{14}$$

where $\mu_P = \boldsymbol{x}'\boldsymbol{\mu}$ is the mean of the portfolio return $P = \boldsymbol{x}'\boldsymbol{Y}$. A short and elegant proof for this result can be found in Dhaene et al. (2008).

Remarks:

1. In case of the general MGHyp distribution, an expression of the risk contribution using the ES is given in Hellmich and Kassberger (2011, Prop. 1). Noting that the joint distribution of asset i and the portfolio loss P is a two-dimensional MGHyp distribution, the risk contribution for $i = 1, \ldots, n$ and $\alpha \in (0,1)$ can be written as

$$\mathrm{RC}_i(\boldsymbol{x}) = -\frac{1}{1-\alpha} \int_{\mathbb{R}} \int_{-\infty}^{\mathrm{GH}_1^{-1}(1-\alpha)} y_1 \cdot f_{\mathrm{GH}_2}(y_1, y_2) dy_2 dy_1. \tag{15}$$

   Unfortunately, this formula requires calculating the VaR, given by the quantile $\mathrm{GH}_1^{-1}(1-\alpha)$ of the univariate GHyp distribution of portfolio returns, and to numerically evaluate a two-dimensional integral for every candidate portfolio. This is computationally demanding, making it prohibitive to use the risk contributions under the general MGHyp distribution for risk parity portfolio optimization.

2. The assumption of elliptically distributed return vectors is less restrictive than it might appear. In an empirical analysis in McNeil et al. (2015), the elliptical special cases of the MGHyp distribution do not deliver a statistically inferior in-sample fit for daily returns in a formal likelihood-ratio test. Furthermore, as discussed in Paolella and Polak (2015b) and Paolella et al. (2019), the out-of-sample forecasting ability of the elliptical MGHyp models are at par with, if not superior to, their asymmetric counterparts due to the inherently difficult task of forecasting the asymmetry parameter vector.

3. It is well-known that under elliptically distributed returns, the minimization of standard deviation, ES, or any other coherent risk measure delivers the same portfolio weights; see Embrechts et al. (2002) and Hu and Kercheval (2010). As we exclusively consider elliptical models in our empirical section, the min-var and the min-ES portfolios always coincide, and in the remainder we use the

general term risk minimization (RM) for this portfolio strategy. Notice that this does not imply that the estimated optimal portfolios coincide when a Gaussian or an elliptical non-Gaussian distribution are used because the parameter estimates of the mean vector and the covariance matrix will generally differ under different distributional assumptions; see Hu and Kercheval (2010).

4. As shown in Stefanovits (2010), the risk parity weights under elliptically distributed returns coincide for variance and ES, only if all assets have the same expected return, i.e., in the case $\mu_i = \mu_j$ for all i, j. Finally, even under the assumption of normally distributed returns, the risk parity weights for variance and ES do not generally coincide, see Jurczenko and Teiletche (2019).

4.2 Fast Computation of the ES under Elliptical MGHyp Distributions

To combine the flexibility of modeling excess kurtosis by the MGHyp distribution with the fast computation of the risk contributions under elliptical distributions, we make use of an expression from Paolella and Polak (2015b) for the ES under the elliptical MGHyp distribution, i.e., for $\boldsymbol{\gamma} = \mathbf{0}$. The application of this formula in (14) is pivotal for fast, efficient computation of the risk parity weights under a heavy-tailed non-Gaussian distribution.

Let the portfolio return $P = \boldsymbol{x}'\boldsymbol{Y}$ be univariate symmetrically generalized hyperbolical distributed $P \sim \text{GHyp}(\lambda, \chi, \psi, \mu, 0, \sigma)$. Then, for $\alpha \in (0, 1)$, Paolella and Polak (2015b, Prop. 3.1) show

$$\text{ES}_\alpha(P) = -\mu + \frac{\sigma}{\alpha\sqrt{2\pi}} C, \tag{16}$$

with the constant C given by

$$C = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2\mathcal{K}_\lambda\left(\sqrt{\chi\psi}\right)} \frac{2\mathcal{K}_{\tilde{\lambda}}\left(\sqrt{\tilde{\chi}\psi}\right)}{\tilde{\chi}^{-\tilde{\lambda}}\left(\sqrt{\tilde{\chi}\psi}\right)^{\tilde{\lambda}}},$$

where

$$\tilde{\lambda} = \lambda + \frac{1}{2}, \qquad \tilde{\chi} = \chi + \frac{(\text{VaR}_\alpha(P) + \mu)^2}{\sigma^2}.$$

Use of this formula is many times faster than standard evaluation of the expected shortfall, and hence is subsequently used in the computation of the risk contributions (14).

Next, we outline a method to accelerate and stabilize the numerical solution of the optimization problem (8). For solving this non-linear constrained optimization problem we use, as in Roncalli (2013) and Stefanovits (2010), the SQP algorithm (implemented in MATLAB), a popular and efficient algorithm for this task. It iteratively solves a sequence of subproblems that consist of optimizing quadratic approximations of the original problem. However, this requires the gradient of the objective function, which is numerically approximated if it is not available in closed form. As demonstrated in Appendix B, this comes at the cost of increased computational time and lowered precision. Instead we provide a closed expression for the gradient of the risk parity objective function in (8).

4.3 Closed Form Expression of the Gradient and Hessian approximation

We now derive the closed form expression of the gradient together with a Hessian approximation for the objective function $f_1(\boldsymbol{x})$ used in (8). The analogous result for $f_2(\boldsymbol{x})$ is given in Appendix A. We use general risk budgets $b_i \geq 0$, obtaining the risk parity case for $b_i = 1/n$ for all $i = 1, \ldots, n$. Our risk parity objective function is

$$f_1(\boldsymbol{x}) = \sum_{i,j=1}^{n} \left(\frac{x_i}{b_i} \frac{\partial \text{ES}_\alpha(\boldsymbol{x})}{\partial x_i} - \frac{x_j}{b_j} \frac{\partial \text{ES}_\alpha(\boldsymbol{x})}{\partial x_j} \right)^2,$$

so taking partial derivative amounts to

$$\frac{\partial f_1(\boldsymbol{x})}{\partial x_k} = 2\sum_{i,j=1}^{n}\left(\frac{x_i}{b_i}\frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} - \frac{x_j}{b_j}\frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_j}\right)\cdot$$
$$\cdot\left(\frac{\delta_{ik}}{b_i}\frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} + \frac{x_i}{b_i}\frac{\partial^2 \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i \partial x_k} - \frac{\delta_{jk}}{b_j}\frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_j} - \frac{x_j}{b_j}\frac{\partial^2 \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_j \partial x_k}\right),$$

where δ_{ik} is the Kronecker delta. It is useful to express the partial derivatives in matrix notation; after some algebra, it can be shown that

$$\nabla_{f_1}(\boldsymbol{x}) = 4n\left(\boldsymbol{x}\odot\left(\frac{\nabla_{\mathrm{ES}}(\boldsymbol{x})}{\boldsymbol{b}}\right)^2\right) + 4H_{\mathrm{ES}}(\boldsymbol{x})\left(n\boldsymbol{m} - \frac{l\boldsymbol{x}}{\boldsymbol{b}}\right) - 4l\frac{\nabla_{\mathrm{ES}}(\boldsymbol{x})}{\boldsymbol{b}},$$

with $H_{\mathrm{ES}}(\boldsymbol{x})$ the Hessian matrix of the ES function, $\nabla_{\mathrm{ES}}(\boldsymbol{x})$ the gradient of the ES function, and

$$l = \frac{\nabla_{\mathrm{ES}}(\boldsymbol{x})}{\boldsymbol{b}}\cdot\boldsymbol{x}, \quad \boldsymbol{m} = \left(\frac{\boldsymbol{x}}{\boldsymbol{b}}\right)^2\odot\nabla_{\mathrm{ES}}(\boldsymbol{x}).$$

Here $\odot$ represents the elementwise product between two vectors, division between vectors is intended element-wise and $\cdot$ indicates the scalar product.
While this closed form expression is appealing, in practice it is too slow. The Hessian matrix contains n^2 elements so that computing it will slow down the algorithm considerably because the gradient needs to be called at every iteration of the optimization algorithm. To overcome this bottleneck, notice that the Hessian only enters in the above equation when multiplied by a vector, so we can use a central difference approximation to compute it more efficiently. Specifically, the first order Taylor approximation of the gradient of ES_α at a point $\boldsymbol{x}\in\mathbb{R}^n$ is

$$\nabla_{\mathrm{ES}}(\boldsymbol{x}+r\boldsymbol{l}) = \nabla_{\mathrm{ES}}(\boldsymbol{x}) + rH_{\mathrm{ES}}(\boldsymbol{x})\boldsymbol{l} + \mathcal{O}(r^2\|\boldsymbol{l}\|^2),$$

for arbitrary $r\in\mathbb{R}$ and $\boldsymbol{l}\in\mathbb{R}^n$, hence

$$H_{\mathrm{ES}}(\boldsymbol{x})\boldsymbol{l} = \frac{\nabla_{\mathrm{ES}}(\boldsymbol{x}+r\boldsymbol{l}) - \nabla_{\mathrm{ES}}(\boldsymbol{x})}{r} + \mathcal{O}(r\|\boldsymbol{l}\|^2).$$

In this work we use a central difference quotient for higher order approximation

$$H_{\mathrm{ES}}(\boldsymbol{x})\boldsymbol{l} = \frac{\nabla_{\mathrm{ES}}(\boldsymbol{x}+r\boldsymbol{l}) - \nabla_{\mathrm{ES}}(\boldsymbol{x}-r\boldsymbol{l})}{2r} + \mathcal{O}(r^2\|\boldsymbol{l}\|^2). \tag{17}$$

Therefore, to compute the product $H_{\mathrm{ES}}(\boldsymbol{x})\boldsymbol{l}$ we just need to compute the gradient of the ES function twice, requiring only a linear number of calls to the ES function. Together with the closed-form expression (16) delineated above, this makes the computation of the gradient much faster than the direct finite difference approximation.

Initializing the Algorithm

The speed and accuracy of the SQP algorithm can be further improved by a starting point in the numerical solver that is close to the optimal solution. For the risk parity problem (8) there exist two obvious candidate portfolios. First, one can use the $1/n$ portfolio weights because it has been empirically observed that the risk parity weights are often considerably close to the equally weighted portfolio, especially when the assets are homogeneous in their level of risk. The second approach is to use the NRP weights, analogous to (7) but computed with the same risk measure as used for the risk parity strategy, i.e., the ES in our setting. This approach is investigated in Cesarone and Colucci (2018) and is found to be a good approximation of the risk parity weights. The NRP vector $\boldsymbol{x}_{\mathrm{NRP}}$ is computed as

$$x_{\mathrm{NRP},i} := \frac{\mathrm{ES}_{\alpha}^{-1}(\boldsymbol{e}_i)}{\sum_{j=1}^{n} \mathrm{ES}_{\alpha}^{-1}(\boldsymbol{e}_j)},$$

where $\boldsymbol{e}_i$ is the ith unit vector and $\mathrm{ES}_{\alpha}(\boldsymbol{e}_i)$ is the ES of the portfolio consisting of one unit of asset i. The NRP strategy is exceptionally fast to compute because it requires only n evaluations of the ES function.

5 Empirical Analysis

An extensive out-of-sample portfolio analysis is performed on two different data sets. We start with the Dow Jones 30 (DJ30) as a pure stock market data set where the risk characteristics of the various assets are rather homogeneous. In the literature, such as Cesarone and Colucci (2018), Roncalli (2013), and Unger (2014), the RP strategy is often regarded as an alternative to the min-var portfolio. For this reason, we investigate the RP strategy in an asset universe that is typical for the analysis of the min-var strategy and compare it extensively to the latter. Most real-world applications of RP strategies consider the allocation of capital into different asset classes, as opposed to assets of one class. For this reason we consider a data set of liquidly traded ETFs that cover multiple asset classes.

To quantify the impact of different levels of turnover on the portfolio performance, we assume proportional transaction cost as in DeMiguel et al. (2009b) and compute the turnover in the same way. Furthermore, we follow Cesarone and Colucci (2018) in ignoring portfolio rebalancing that is solely due to intraday prices changes. Thus for the passive $1/n$ strategy we report a turnover of zero, although in reality it does have some turnover. This approximation does not change the findings of our analysis and even makes the $1/n$ portfolio more competitive under transaction fees.

5.1 Dow Jones 30 Data

The data set consists of the daily log-returns of the $n = 30$ components of the DJ30 index, spanning the period from 03.01.1996 to 30.12.2016. We include all companies that were part of the DJ30 index at some point in time and that were listed on the New York Stock Exchange during the entire sample period. Using 1000 data points for parameter estimation, the out-of-sample trading period starts on 17.12.1999.

The results are presented in Table 1, with the best value for each category given in bold. For the RM strategy the highest risk-adjusted returns are attained for the SNIG, followed by the MLap and Mt distributions. The Gaussian model delivers the lowest Sharpe ratio and the highest turnover but still clearly outperforms the $1/n$ strategy. The Sharpe ratios net of transaction costs are the highest for the SNIG distribution under all fee levels. Even using the high value of 50bp, the RM strategy, under all distributions, outperforms the $1/n$ portfolio. The latter has the highest volatility and maximum drawdown, i.e., equal capital diversification provides an inferior loss protection.

The SNIG distribution performs best also for the RP strategy. Notably, the performance differences between using a Gaussian and a non-Gaussian elliptical MGHyp distribution is much smaller for the RP than for the RM portfolio. This can be interpreted as robustness of the RP portfolio towards distributional misspecification. The RP strategy outperforms the equally weighted portfolio in terms of returns, risk-adjusted returns, volatility and maximum drawdown. However, it must be noted that the RP strategy is clearly outperformed by the RM strategy in all performance measures, except for turnover, and under any distributional assumption. This supports the intuition of RP as being a trade-off

Risk Minimization								
Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
SNIG	**0.0254**	**0.9120**	**109.0773**	**35.8788**	0.6395	**0.4429**	**0.6304**	**0.0071**
MLap	**0.0254**	0.9123	108.7343	36.0652	0.6651	0.4413	0.6278	**0.0071**
Mt	0.0252	0.9127	107.9117	36.2944	0.6806	0.4378	0.6230	0.0070
MN	0.0231	0.9092	99.2368	43.3728	0.7120	0.4041	0.5720	0.0065
1/n	0.0243	1.2075	104.2089	56.9816	**0.0000**	0.3196	0.4490	0.0049

Distr.	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
SNIG	**0.4429**	**0.4417**	**0.4373**	**0.4317**	**0.4206**	**0.3872**
MLap	0.4413	0.4402	0.4356	0.4298	0.4182	0.3835
Mt	0.4378	0.4366	0.4319	0.4260	0.4141	0.3786
MN	0.4041	0.4029	0.3979	0.3917	0.3793	0.3420
1/n	0.3196	0.3196	0.3196	0.3196	0.3196	0.3196

Risk Parity								
Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
SNIG	**0.0249**	1.1196	**106.5539**	50.4878	0.1203	**0.3524**	**0.4954**	**0.0054**
MLap	**0.0249**	1.1197	106.5536	50.5082	0.1213	**0.3524**	**0.4954**	**0.0054**
Mt	0.0248	1.1192	106.4459	50.4841	0.1229	0.3522	0.4951	**0.0054**
MN	0.0245	**1.1155**	105.0241	**50.2743**	0.1379	0.3486	0.4903	**0.0054**
1/n	0.0243	1.2075	104.2089	56.9816	**0.0000**	0.3196	0.4490	0.0049

Distr.	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
SNIG	**0.3524**	**0.3522**	**0.3516**	**0.3507**	**0.3490**	**0.3439**
MLap	**0.3524**	**0.3522**	**0.3515**	0.3507	0.3489	0.3438
Mt	0.3522	0.3520	0.3513	0.3504	0.3487	0.3435
MN	0.3486	0.3484	0.3477	0.3467	0.3447	0.3388
1/n	0.3196	0.3196	0.3196	0.3196	0.3196	0.3196

Table 1: DJ30 stocks traded from 17.12.1999 until 30.12.2016; using 1000 data points for estimation; returns and volatilities are daily and measured in percent; *Top Panel*: Comparison of the RM strategy under various distributions; *Second Panel*: Sharpe ratios net of transaction costs of the RM strategy under various levels of transaction fees; *Third Panel*: Comparison of the RP strategy under various distributions; *Bottom Panel*: Sharpe ratios net of transaction costs of the RP strategy under various levels of transaction fees

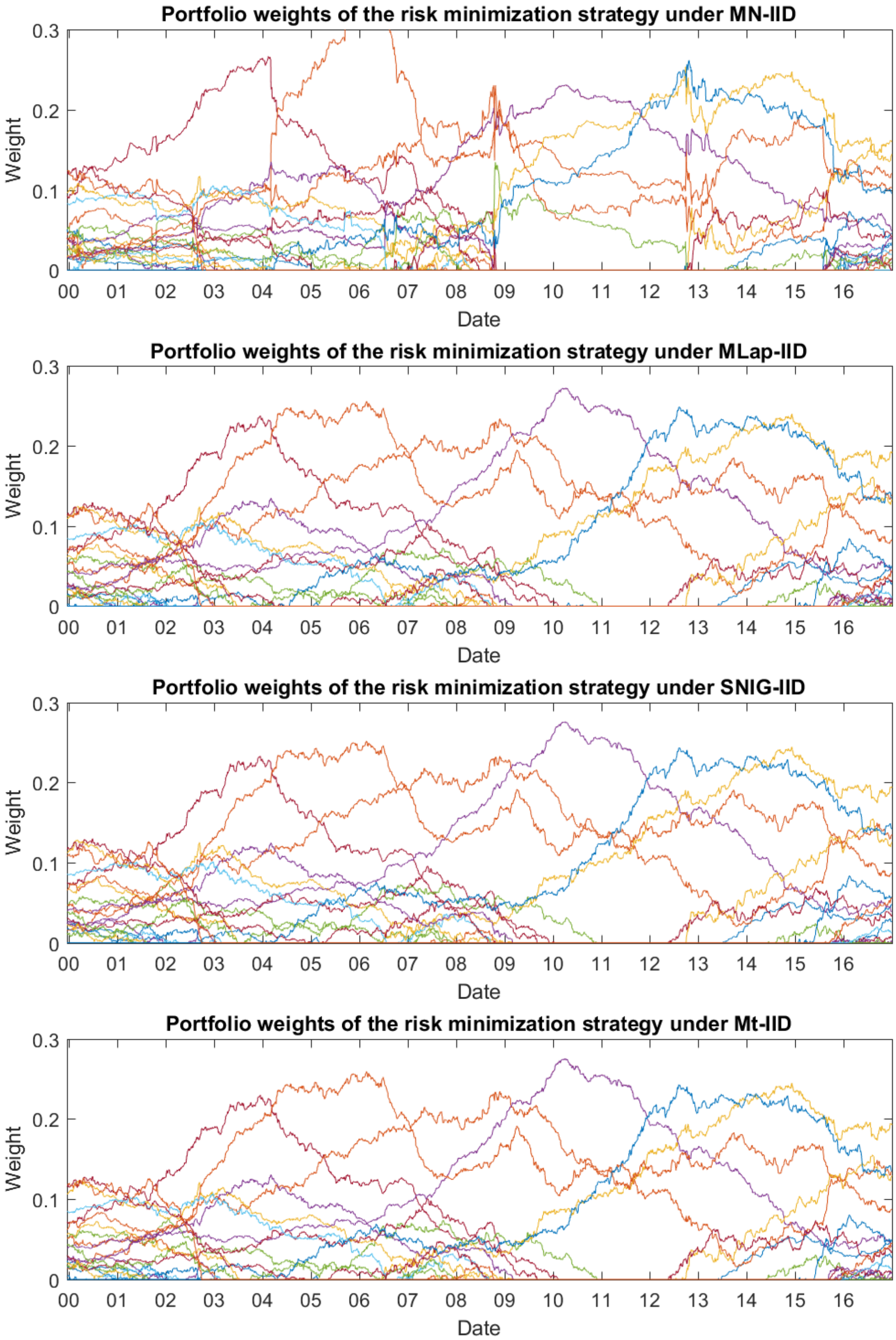

Figure 1: Portfolio weights of the RM strategy with DJ30 stocks; out-of-sample trading period from 17.12.1999 until 30.12.2016; MN-IID (top), MLap-IID (second), SNIG-IID (third), and Mt-IID model (bottom).

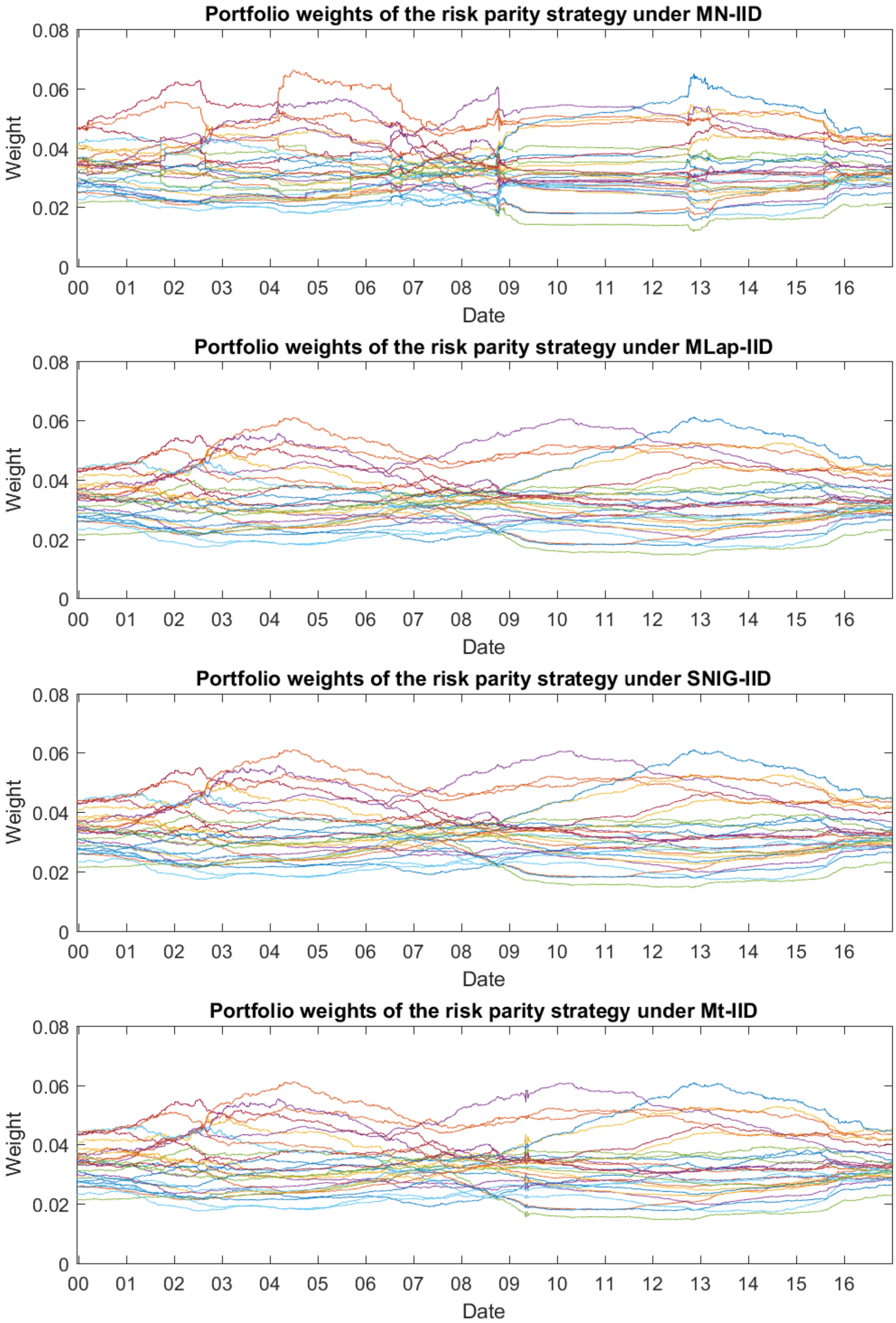

Figure 2: Portfolio weights of the RP strategy with DJ30 stocks; out-of-sample trading period from 17.12.1999 until 30.12.2016; MN-IID (top), MLap-IID (second), SNIG-IID (third), and Mt-IID model (bottom).

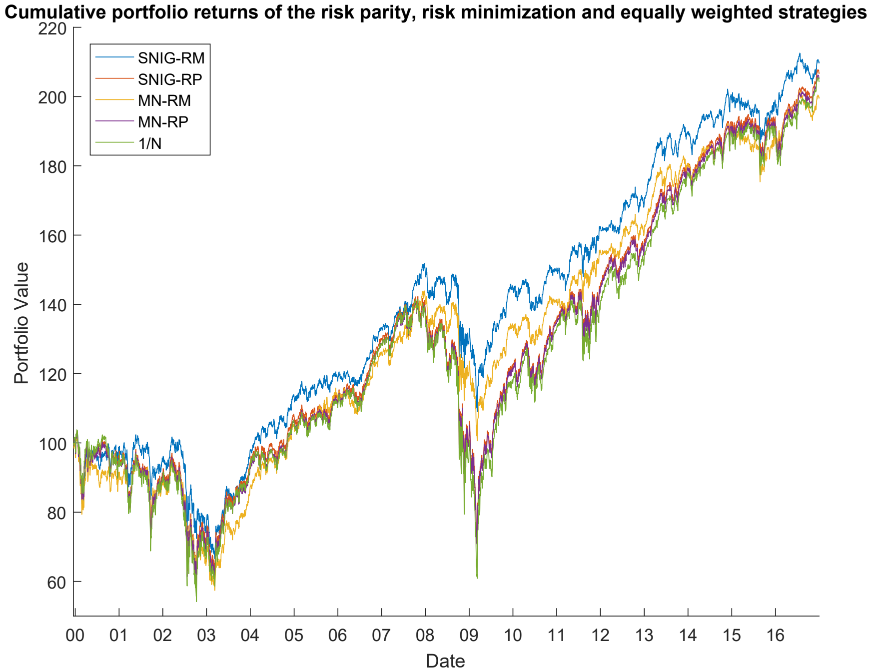

Figure 3: Cumulative returns of the RM, RP and equally weighted strategies; trading DJ30 stocks; out-of-sample trading period from 17.12.1999 until 30.12.2016; only the best performing MGHyp model (SNIG) and the Gaussian (MN) model are plotted.

between the RM and the $1/n$ portfolio, but it also questions the meaningfulness of the RP approach for the stocks-only application as it neither provides optimal downside protection nor better risk-adjusted returns. The maximum drawdown of the RP portfolio is only slightly lower than for the $1/n$, but it is much higher than the drawdown of the RM strategy. This is of particular importance because leverage is often applied to RP portfolio to meet return targets.

On the other hand, the turnover of the RP portfolio is vastly lower than for RM, with the lowest amount of rebalancing attained under the SNIG distribution. The reduction in turnover appears to be the biggest advantage of using a non-Gaussian elliptical MGHyp distribution over the Gaussian one for RP investing. The increased stability of portfolio weights can also be seen from Figures 1 and 2. Use of a heavy-tailed distribution leads to smoother weights for both the RM and the RP strategies. Most notably, at the beginning of the financial crisis in fall 2008, the non-Gaussian distributions cause less rebalancing, which can be of utmost importance during a market crash. The weights under the SNIG, MLap and Mt distributions are almost indistinguishable. As seen from Figure 2, the RP strategy, when applied to a set of homogeneous assets (in terms of similar risk and high correlations), leads to an allocation is closer to the equally weighted than the RM portfolio. In particular it avoids extreme concentrations in individual assets.

The cumulative return plots in Figure 3 shows that the RM strategy under the SNIG distribution outperforms all other portfolios during the entire trading period. The equally weighted portfolio suffers the heaviest losses in all market crashes, most visibly during the burst of the dotcom bubble, the turmoils following the 9/11 attacks, the Iraq war, and the financial crisis 2008-2009. The performance of the RP follows more closely that of the equally weighted than the RM strategy, leading to reduced loss protection in crisis times. Furthermore, the difference between the Gaussian and the MGHyp-based RP strategies is very small, especially compared to the case of the RM strategy.

5.2 Multi-Asset ETF Data

We cover a spectrum of asset classes and geographic locations by using eight liquid iShares ETFs issued by BlackRock Inc. and traded on the New York Stock Exchange. These are the MSCI BRIC ETF (BKF) covering equity markets in BRIC countries; the MSCI Eurozone ETF (EZU) covering large- and mid-cap equities from the Euro zone; the Core S&P Total US Stock Market ETF (ITOT) tracking a broad-based index of US equities; the S&P GSCI Commodity-Indexed Trust ETF (GSG) tracking a broad investment in futures contracts on energy, agriculture and industrial metals; the Gold Trust ETF (IAU) following the price of gold; the US Real Estate ETF (IYR) investing in listed US real estate; the 1-3 Year Treasury Bond ETF (SHY) mirroring US treasury bonds with maturities between one and three years; and similarly the 10-20 Year Treasury Bond ETF (TLH) for US treasury bonds with maturities between ten and twenty years. The sample spans the period from 03.01.2008 until 29.12.2017. Using 750 data points for model calibration, the trading period starts on 23.12.2010. Due to the anticipated concentration of the RM allocation in bonds, we additionally consider the mean-variance (MV) portfolio with a required annual return of 5%.

The results are given in Table 2. The RM approach delivers again, under any return distribution, a higher Sharpe ratio, lower volatility and lower drawdown than the RP strategy. Moreover, the risk-adjusted return of the $1/n$ strategy is far below those of the RM and RP portfolios. However, the total return under both RM and RP are very low, which is due to both portfolios being heavily invested in short-term U.S. treasury bonds; see Figures 4 and 5. The RM approach is extremely concentrated in this asset, while the RP portfolio also holds a significant amount of long-term U.S. bonds. Nevertheless, also the RP approach holds only small amounts of the other six assets and the asset allocation resembles more closely the RM approach than the equally weighted portfolio. This is due to the vastly different risk characteristics of the assets in this investment universe. In order to elevate returns, we also consider the MV portfolio with 5% minimum expected annual return. As seen from the last two panels of Table 2, the total returns are strongly improved and exceed those of the $1/n$ portfolio. The same holds true for the risk-adjusted returns, which outperform all other strategies considered herein. The MV portfolio's risk estimates exceed those of the RM strategy but are much lower than for RP. As seen from Figures 5 and 6, the RP allocation resembles somewhat that of the MV portfolio but the weights are smoother and show less sensitivity with respect to the distributional assumption.

5.3 Portfolio Optimization under Heteroskedasticity: Dow Jones 30 Data

Finally, in this section, we account for the heteroskedasticity of returns by fitting the GARCH-dynamics of the COMFORT-CCC model, instead of the i.i.d. case. For this exampple, we return to the data set consisting of the $n = 30$ components of the DJ30 index, spanning the period from 03.01.1996 to 30.12.2016. As in Section 5.1 we use 1000 data points for estimation, the out-of-sample trading period starts on 17.12.1999.

The results are shown in Table 3. Incorporating the GARCH-CCC covariance dynamics improves the returns, risk-adjusted returns and risk characteristics of both the RM and the RP strategies. In particular, for the RP strategy, the impact of modeling the covariance dynamics with a simple MGARCH structure is larger than that of accounting for heavy tails with an elliptical MGHyp distribution. However, this comes at the price of vastly increased turnover. When transaction fees are taken into account, the GARCH-CCC structure quickly becomes unfavorable because of the heavy increase in transaction costs. The RP still exhibits lower turnover than the RM portfolio but, for both strategies, the performance improvement of the GARCH-CCC model is already offset by higher trading costs for low and modest levels of transaction fees.

Risk Minimization

Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
MLap	0.0031	0.0527	5.4776	0.7451	0.4310	**0.9334**	**1.3703**	**0.0164**
Mt	0.0029	**0.0515**	5.2061	0.7032	0.2698	0.9090	1.3268	0.0161
SNIG	0.0029	0.0525	5.1746	**0.6873**	0.2107	0.8854	1.2904	0.0155
MN	0.0027	0.0536	4.8385	0.7067	0.5870	0.8113	1.1824	0.0142
1/n	**0.0112**	0.6180	**19.7273**	19.2632	**0.0000**	0.2869	0.3944	0.0046

Distribution	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
MLap	**0.9334**	**0.9207**	**0.8693**	0.8047	0.6743	0.2812
Mt	0.9090	0.9008	0.8677	**0.8261**	0.7425	0.4898
SNIG	0.8854	0.8791	0.8537	0.8220	**0.7581**	**0.5646**
MN	0.8113	0.7940	0.7246	0.6377	0.4636	-0.0575
1/n	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869

Risk Parity

Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
MLap	0.0059	0.1319	10.3613	**3.6663**	0.1019	**0.7060**	**1.0132**	**0.0130**
SNIG	0.0059	0.1319	10.3369	3.6876	0.1008	0.7046	1.0110	**0.0130**
Mt	0.0058	0.1315	10.1723	3.7585	0.0980	0.6954	0.9972	0.0128
MN	0.0057	**0.1302**	10.0687	3.7827	0.1097	0.6950	0.9979	0.0127
1/n	**0.0112**	0.6180	**19.7273**	19.2632	**0.0000**	0.2869	0.3944	0.0046

Distribution	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
MLap	**0.7060**	**0.7048**	**0.6999**	**0.6938**	**0.6816**	**0.6448**
SNIG	0.7046	0.7034	0.6985	0.6925	0.6804	0.6440
Mt	0.6954	0.6942	0.6895	0.6836	0.6718	0.6363
MN	0.6950	0.6937	0.6884	0.6817	0.6684	0.6283
1/n	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869

Mean-ES

Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
Mt	0.0114	0.1448	20.0531	2.4043	0.8085	**1.2445**	**1.8002**	**0.0220**
MN	**0.0172**	0.2204	**30.3970**	3.2286	1.8405	1.2395	1.7967	0.0212
MLap	0.0093	**0.1207**	16.3966	**1.9778**	0.5668	1.2209	1.7700	0.0222
SNIG	0.0098	0.1276	17.3118	2.1189	0.6171	1.2196	1.7661	0.0219
1/n	0.0112	0.6180	**19.7273**	19.2632	**0.0000**	0.2869	0.3944	0.0046

Distribution	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
Mt	**1.2445**	**1.2356**	**1.2003**	**1.1561**	1.0677	0.8021
MN	1.2395	1.2263	1.1733	1.1071	0.9746	0.5768
MLap	1.2209	1.2135	1.1838	1.1467	**1.0724**	**0.8488**
SNIG	1.2196	1.2119	1.1814	1.1431	1.0665	0.8363
1/n	0.2869	0.2869	0.2869	0.2869	0.2869	0.2869

Table 2: Eight multi-asset, multi-country ETFs traded from 23.12.2010 until 29.12.2017; using 750 data points for estimation; returns and volatilities are daily and measured in percent, *Top Panel*: Comparison of the RM strategy under various distributions; *Second Panel*: Sharpe ratios net of transaction costs of the RM strategy under various levels of transaction fees; *Third Panel*: Comparison of the RP strategy under various distributions; *Bottom Panel*: Sharpe ratios net of transaction costs of the RP strategy under various levels of transaction fees; *Fifth Panel*: Comparison of the mean-ES strategy with 5% minimum expected annual return under various distributions; *Bottom Panel*: Sharpe ratios net of transaction costs of the mean-ES strategy with 5% minimum expected annual return under various levels of transaction fees.

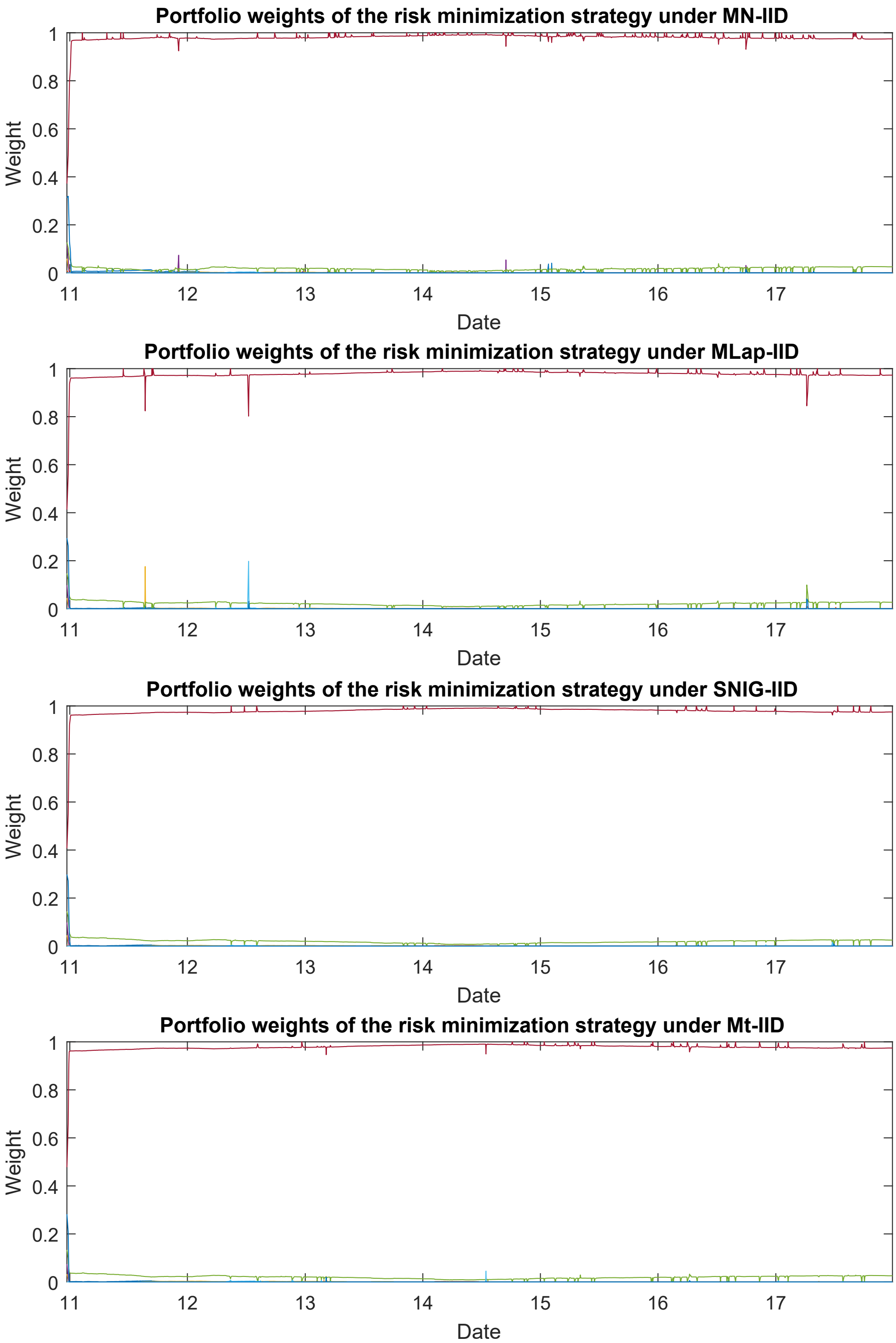

Figure 4: Portfolio weights of the RM strategy using eight multi-asset multi-country ETFs; out-of-sample trading period from 23.12.2010 until 29.12.2017 under the MN-IID (top), the MLap-IID (second), the SNIG-IID (third), and the Mt-IID model (bottom).

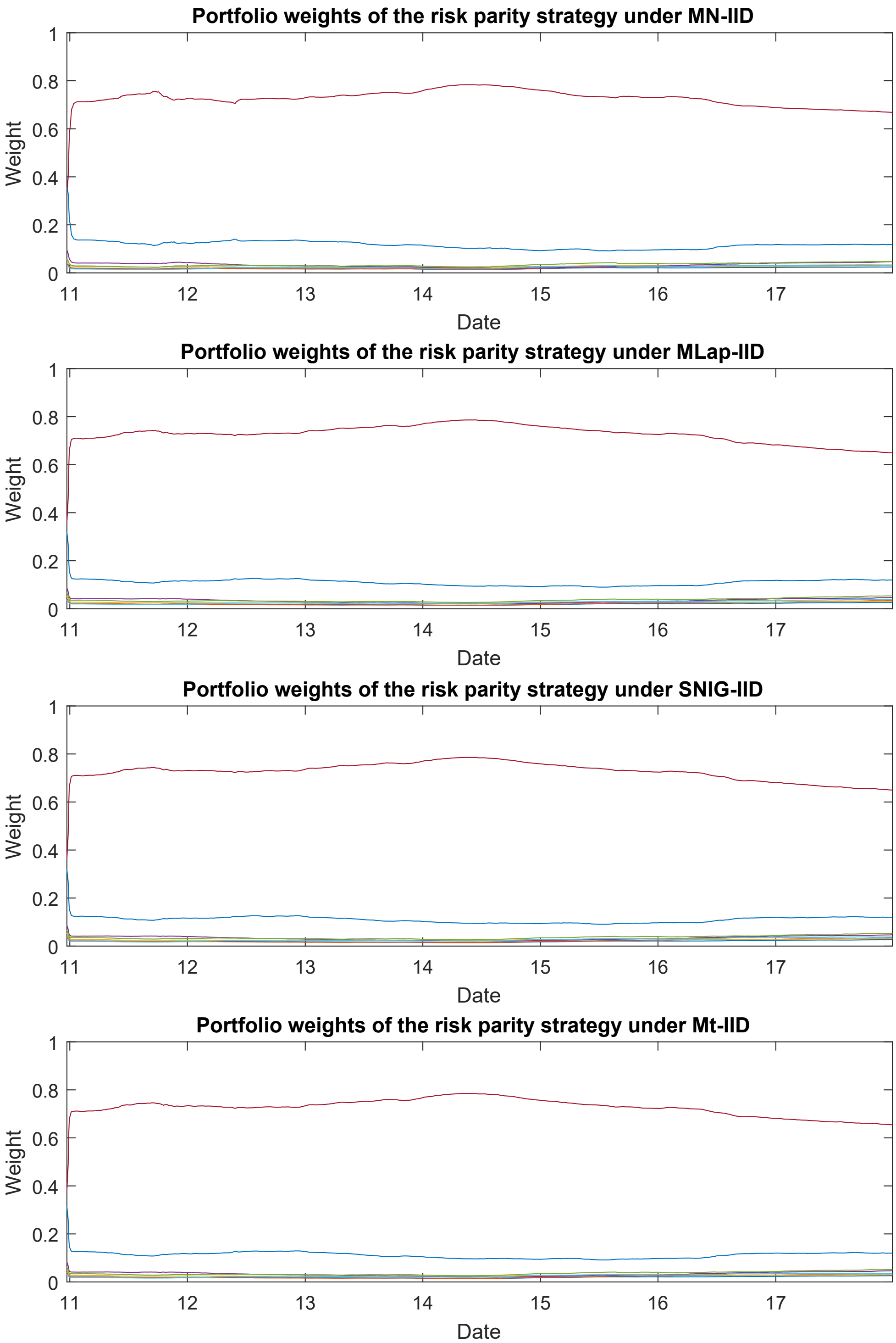

Figure 5: Portfolio weights of the RP strategy using eight multi-asset multi-country ETFs; out-of-sample trading period from 23.12.2010 until 29.12.2017 under the MN-IID (top), the MLap-IID (second), the SNIG-IID (third), and the Mt-IID model (bottom).

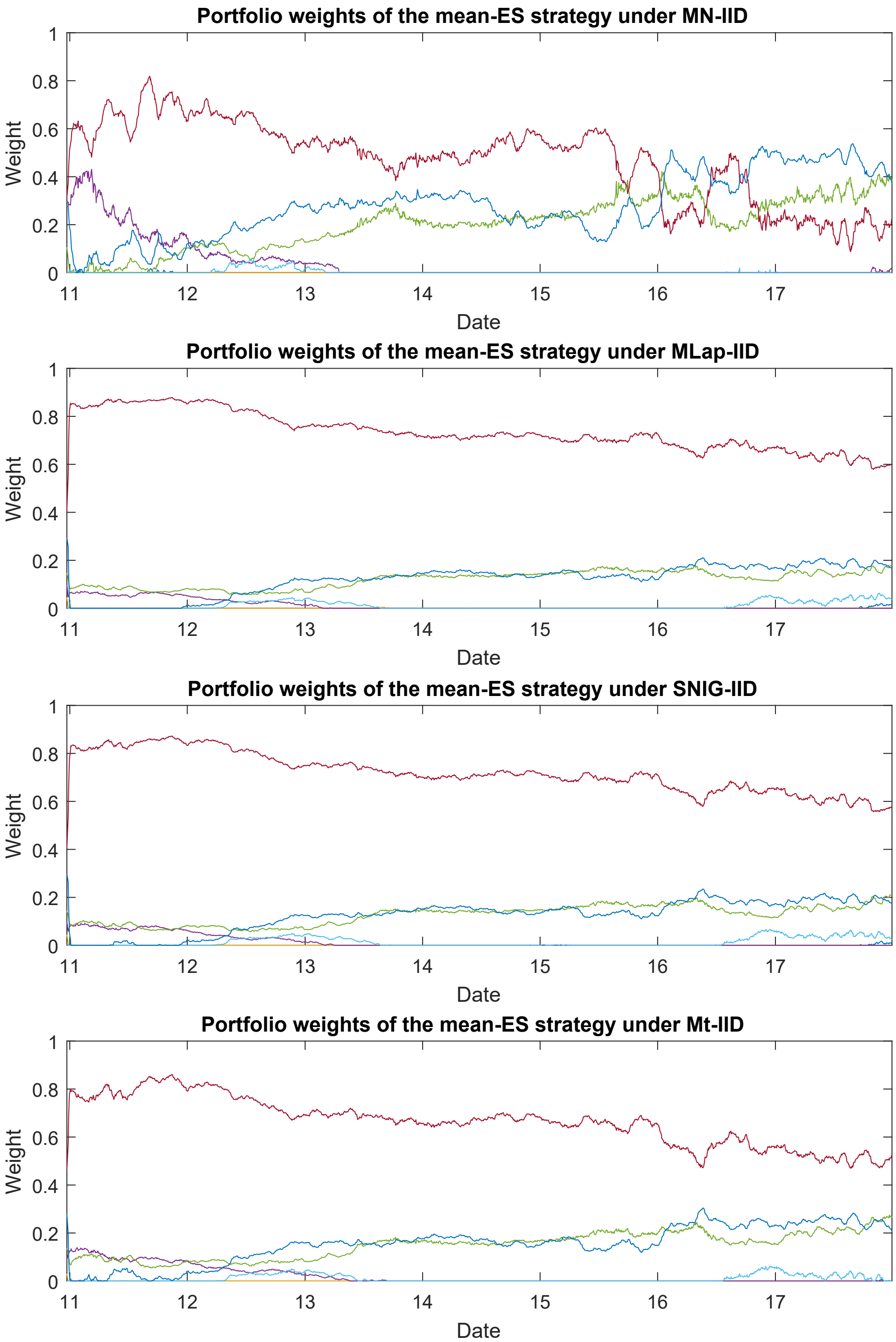

Figure 6: Portfolio weights of the MV strategy with 5% minimum expected annual return using eight multi-asset multi-country ETFs; out-of-sample trading period from 23.12.2010 until 29.12.2017 under the MN-IID (top), the MLap-IID (second), the SNIG-IID (third), and the Mt-IID model (bottom).

Risk Minimization								
Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
MLap-CCC	**0.0263**	**0.9061**	**112.7441**	**34.3972**	12.4719	**0.4608**	**0.6488**	**0.0073**
Mt-CCC	0.0257	0.9063	110.0678	34.9879	12.4703	0.4497	0.6325	0.0071
SNIG-CCC	0.0253	0.9068	108.5527	35.2704	12.7526	0.4433	0.6246	0.0070
MN-CCC	0.0221	0.9071	94.6365	37.1326	16.0321	0.3863	0.5425	0.0061
1/n	0.0243	1.2075	104.2089	56.9816	**0.0000**	0.3196	0.4490	0.0049

Distribution	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
MLap-CCC	**0.4608**	**0.4389**	**0.3515**	0.2422	0.0236	-0.6317
Mt-CCC	0.4497	0.4279	0.3405	0.2312	0.0128	-0.6420
SNIG-CCC	0.4433	0.4210	0.3321	0.2208	-0.0016	-0.6680
MN-CCC	0.3863	0.3583	0.2460	0.1057	-0.1749	-1.0148
1/n	0.3196	0.3196	0.3196	**0.3196**	**0.3196**	**0.3196**

Risk Parity								
Distr.	*Exp. Daily Ret*	*Daily Vola*	*Total Return*	*Max. Drawdown*	*Avg. Turnover*	*Sharpe*	*Sortino*	*Starr*
MLap-CCC	**0.0255**	1.1022	**109.3189**	**49.3641**	2.5488	**0.3673**	**0.5159**	**0.0057**
SNIG-CCC	0.0254	1.1024	108.9990	49.3930	2.5900	0.3661	0.5142	**0.0057**
Mt-CCC	0.0254	1.1024	108.7422	49.5167	2.6342	0.3653	0.5130	**0.0057**
MN-CCC	0.0247	**1.0980**	106.0248	50.0960	3.6627	0.3576	0.5021	0.0056
1/n	0.0243	1.2075	104.2089	56.9816	**0.0000**	0.3196	0.4490	0.0049

Distribution	*0bp*	*1bp*	*5bp*	*10bp*	*20bp*	*50bp*
MLap-CCC	**0.3673**	**0.3636**	**0.3489**	**0.3306**	0.2938	0.1837
SNIG-CCC	0.3661	0.3624	0.3475	0.3288	0.2915	0.1796
Mt-CCC	0.3653	0.3615	0.3463	0.3273	0.2894	0.1756
MN-CCC	0.3576	0.3523	0.3311	0.3046	0.2517	0.0928
1/n	0.3196	0.3196	0.3196	0.3196	**0.3196**	**0.3196**

Table 3: Comparison of portfolio strategies under various COMFORT-CCC models; DJ30 stocks traded from 17.12.1999 until 30.12.2016; using 1000 data points for estimation; returns and volatilities are daily and measured in percent; *Top Panel*: Comparison of the RM strategy under Gaussian- and MGHyp-based GARCH-CCC; *Second Panel*: Sharpe ratios net of transaction costs of the RM strategy under various levels of transaction fees; *Third Panel*: Comparison of the RP strategy under Gaussian- and MGHyp-based GARCH-CCC; *Bottom Panel*: Sharpe ratios net of transaction costs of the RP strategy under various levels of transaction fees.

6 Conclusion

We proposed a new method for efficient computation of risk parity (RP) portfolios under heavy-tailed returns. Modeling asset returns with an elliptical MGHyp distribution allows for both a fast estimation of model parameters and a semi-closed form expression of the risk contributions. Further exploitation of several numerical shortcuts, along with using a result of Paolella and Polak (2015b) for the expected shortfall under ellipticity, leads to exceptionally fast and precise computation of the RP portfolio weights. In line with Stefanovits (2010), who uses an elliptical multivariate Student-t distribution, we find for the different special cases of the MGHyp distribution investigated herein, that using a fat-tailed, elliptical distribution instead of the Gaussian for computing the RP portfolio has a surprisingly small impact on the portfolio performance.

Our empirical investigation shows that, while accounting for heavy tails in RP allocations leads only to mild improvements of portfolio performance, the major advantage is a smoother transition of asset weights over time and hence a reduced portfolio turnover. In the financial crisis of 2008, i.e., during times of severe market distress, the RP portfolio under the MGHyp distribution requires significantly less rebalancing than the Gaussian-based model as the market shocks are better captured by a return distribution that permits excess kurtosis. While the differences between the various non-Gaussian special cases of the MGHyp distribution are small, overall, the symmetric Laplace distribution delivers the best portfolio results on most data sets considered and always outperforms the Student-t case which has been studied in literature before. Notably, the negative effect of using a Gaussian return distribution, i.e., the potential harm of ignoring heavy tails, is smaller for the RP strategy than for the classical minimum variance portfolio. Thus the RP method can be regarded as more robust to distributional assumptions and ultimately less prone to model risk.

Aiming to account for the most important stylized facts of asset returns, we also consider the impact of modeling heteroskedasticity with a GARCH-CCC model, both under Gaussian and heavy tailed returns. The risk characteristics and realized returns are moderately improved when the covariance matrix forecasts are generated from a GARCH-CCC structure, but this comes at the cost of increased portfolio turnover. Further studies could investigate methods to dampen this side effect, e.g., by means of regularization of the objective function or by use of a more robust dynamic covariance matrix framework, such as that of Paolella et al. (2021).

Generally, we observe in our empirical analysis that the RP approach is dominated by risk minimization (RM), not only in terms of lower risk, but also in terms of higher risk-adjusted returns. The weak performance of RP has been reported earlier in Cesarone and Colucci (2018), Roncalli (2013), Scherer (2014), and Unger (2014), among others. In light of the fact that the RP portfolio is mean-variance efficient, only if all assets have identical Sharpe ratios and correlations, these empirical findings seem consistent. Nevertheless, it should be noted that the equally weighted portfolio is outperformed by RP in all our tests and even under realistic trading fees. This contrasts the general notion that the equally weighted portfolio is notoriously difficult to outperform by optimized portfolios (DeMiguel et al., 2009b). May o our surprise, the

While Roncalli (2013) highlights the robustness of asset weights and the avoidance of highly concentrated portfolios as core strengths of RP investing, we find that this portfolio is not necessarily devoid of highly concentrated weights. When an asset with significantly lower volatility is added to the asset universe its weight will dominate. On our data set consisting of eight multi-asset, cross-country ETFs, more than 80% of the capital is invested in the ETF on 1-3 year US treasury bills, because its volatility is close to zero. The weights of the RP portfolio in this example resemble closely those of the mean-variance portfolio. In fact, the selection of the asset universe is more crucial for RP than for min- or mean-variance strategies, because non-negligible positions are typically held in all assets. Additionally, of practical interest to investors is the insight that the equally weighted allocation is most strongly outperformed by optimized portfolios in the multi-asset case, i.e. naive diversification in a multi-asset context is shown to be heavily suboptimal.

Regarding the risk characteristics of the various portfolios, our study suggests that the RP strategy does not provide optimal protection against portfolio drawdowns. Even the simplest minimum-variance portfolio under the Gaussian returns delivers a lower out-of-sample volatility and a vastly reduced maximum drawdown than the RP strategy on the data sets considered herein.

Finally, the strengths of RP optimization arguably lie in the stability of portfolio weights and the low portfolio turnover compared to RM and general Markowitz optimization. A contribution of our study is to unveil that another advantage of RP is its robustness with respect to a misspecification of the return distribution.

In future work, one could entertain a regularization method that penalizes portfolio rebalancing with the effect that the assets weights are further stabilized and transaction costs are reduced. We conjecture that the weights of the RM portfolio can be robustified such that the resulting portfolio has a turnover level similar to RP, while maintaining a better risk-return profile. In the same spirit and in line with our empirical results, Bai et al. (2016) calls the RP allocation approach a compromise between risk minimization and the equally weighted portfolio. This raises the question whether one could build simple combinations of the minimum-variance and the static $1/n$ portfolio weights, similar to Suh (2016) and Tu and Zhou (2011), such that largely homogeneous risk contributions and low turnover are achieved, while still retaining parts of the increased risk-adjusted returns of the RM approach. Further studies could investigate this question.

Appendices

A Gradient Expression for the Risk Parity Objective Function $f_2(\boldsymbol{x})$

We presented the closed-form gradient expression for objective function $f_1(\boldsymbol{x})$ in Section 4.3. We now consider the other specification of the risk parity objective function, namely $f_2(\boldsymbol{x})$, given by

$$f_2(\boldsymbol{x}) = \sum_{i=1}^{n} \left(x_i \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} - b_i \mathrm{ES}_\alpha(\boldsymbol{x}) \right)^2 .$$

Taking the partial derivative with respect to x_k gives

$$\frac{\partial f_2(\boldsymbol{x})}{\partial x_k} = 2 \sum_{i=1}^{n} \left(x_i \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} - b_i \mathrm{ES}_\alpha(\boldsymbol{x}) \right) \left(-b_i \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_k} + x_i \frac{\partial^2 \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_k \partial x_i} + \delta_{ik} \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_i} \right).$$

This can be written in matrix form as

$$\nabla_{f_2}(\boldsymbol{x}) = 2 H_{\mathrm{ES}}(\boldsymbol{x})(\boldsymbol{x} \odot \boldsymbol{m}) + (l + y)\nabla_{\mathrm{ES}}(\boldsymbol{x}) + 2\nabla_{\mathrm{ES}}(\boldsymbol{x}) \odot \boldsymbol{m},$$

where $H_{\mathrm{ES}}(\boldsymbol{x})$ is the hessian matrix of the ES and

$$\boldsymbol{m} = \boldsymbol{x} \odot \nabla_{\mathrm{ES}}(\boldsymbol{x}) - \mathrm{ES}_\alpha(\boldsymbol{x})\boldsymbol{b}, \qquad l = -2 \sum_{k=1}^{n} x_k b_k \frac{\partial \mathrm{ES}_\alpha(\boldsymbol{x})}{\partial x_k}, \qquad y = 2\mathrm{ES}_\alpha(\boldsymbol{x}) \sum_{k=1}^{n} b_k^2.$$

We observe a slightly slower convergence of the numerical optimization under this objective function than for $f_1(\boldsymbol{x})$ and therefore used the latter in our empirical analysis.

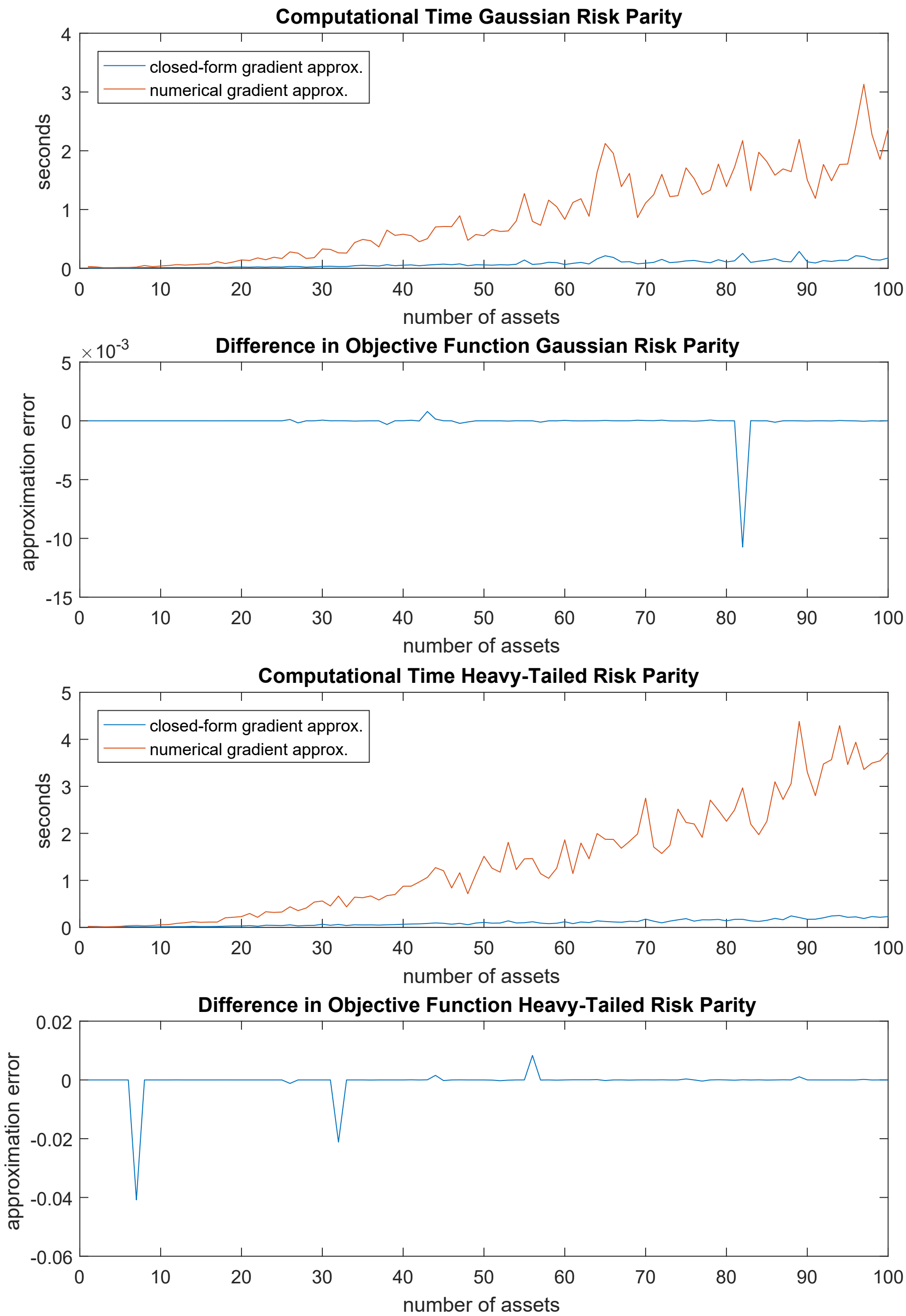

Figure 7: Computational time and approximation error of the closed-form versus numerical gradient approximation for Gaussian- and MGHyp-based risk parity portfolios for varying number of assets

B Computational Speed

The closed-form gradient expression of Section 4.3 delivers a major reduction of computational time compared to brute-force numerical approximation. In Figure 7, we report for both methods the computational time and the differences in numerical optima of the risk parity objective function, for varying numbers of assets and based on simulated asset returns. Our closed-form gradient expression, together with the Hessian approximation, becomes much faster than the basic numerical approximation when the number of assets grows. At the same time, the precision of our method is on par with the numerical approximation method. The reduction in computational time is most pronounced for the case of a non-Gaussian MGHyp distribution.

Declarations of Interest: No potential conflict of interest was reported by the authors.

References

Acerbi, C. and Tasche, D. (2002). Expected Shortfall: A Natural Coherent Alternative to Value at Risk. *Economic Notes*, 31(2):379–388.

Ang, A. (2014). *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61(1):259–299.

Bai, X., Scheinberg, K., and Tütüncü, R. (2016). Least-Squares Approach to Risk Parity in Portfolio Selection. *Quantitative Finance*, 16(3):357–376.

Baker, M., Bradley, B., and Wurgler, J. (2011). Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. *Financial Analysts Journal*, 67(1):40–54.

Barry, C. B. (1974). Portfolio Analysis under Uncertain Means, Variances, and Covariances. *The Journal of Finance*, 29(2):515–522.

Best, M. J. and Grauer, R. R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *The Review of Financial Studies*, 4(2):315–342.

Boudt, K., Carl, P., and Peterson, B. G. (2012). Asset Allocation with Conditional Value-at-Risk Budgets. *Journal of Risk*, 15(3):39–68.

Cesarone, F. and Colucci, S. (2018). Minimum Risk versus Capital and Risk Diversification Strategies for Portfolio Construction. *Journal of the Operational Research Society*, pages 1–18.

Clarke, R., De Silva, H., and Thorley, S. (2006). Minimum-Variance Portfolios in the US Equity Market. *Journal of Portfolio Management*, 33(1):10.

De Nard, G., Engle, R. F., Ledoit, O., and Wolf, M. (2022). Large Dynamic Covariance Matrices: Enhancements Based on Intraday Data. *Journal of Banking & Finance*, 138:106426.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science*, 55:798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22(5):1915–1953.

Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A., and Vanduffel, S. (2008). Some Results on the CTE-based Capital Allocation Rule. *Insurance: Mathematics and Economics*, 42(2):855 – 863.

Eberlein, E. and Keller, U. (1995). Hyperbolic Distributions in Finance. *Bernoulli*, 1:281–299.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and Dependence in Risk Management: Properties and Pitfalls. *Risk Management: Value at Risk and Beyond*.

Engle, R. F., Ledoit, O., and Wolf, M. (2017). Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics*, 0(0):1–13.

Fang, K., Kotz, S., and Ng, K. (1990). *Symmetric Multivariate and Related Distributions*. Monographs on Statistics and Applied Probability. Chapman and Hall.

Frankfurter, G. M., Phillips, H. E., and Seagle, J. P. (1971). Portfolio Selection: The Effects of Uncertain Means, Variances, and Covariances. *Journal of Financial and Quantitative Analysis*, 6(5):1251–1262.

Frazzini, A. and Pedersen, L. H. (2014). Betting Against Beta. *Journal of Financial Economics*, 111(1):1–25.

Frost, P. A. and Savarino, J. E. (1986). An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Frost, P. A. and Savarino, J. E. (1988). For Better Performance: Constrain Portfolio Weights. *The Journal of Portfolio Management*, 15(1):29–34.

Gava, J., Guevara, F., and Turc, J. (2021). Turning Tail Risks into Tailwinds. *The Journal of Portfolio Management*.

Goldfarb, D. and Iyengar, G. (2003). Robust Portfolio Selection Problems. *Mathematics of Operations Research*, 28(1):1–38.

Gourieroux, C., Laurent, J.-P., and Scaillet, O. (2000). Sensitivity Analysis of Values at Risk. *Journal of Empirical Finance*, 7(3):225–245.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the Determinants of Expected Stock Returns. *Journal of Financial Economics*, 41(3):401–439.

Hellmich, M. and Kassberger, S. (2011). Efficient and Robust Portfolio Optimization in the Multivariate Generalized Hyperbolic Framework. *Quantitative Finance*, 11(10):1503–1516.

Hu, W. and Kercheval, A. N. (2010). Portfolio Optimization for Student t and Skewed t Returns. *Quantitative Finance*, 10(1):91–105.

Hurst, B., Johnson, B., and Ooi, Y. (2010). Understanding Risk Parity. *AQR Capital Management White Paper*.

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.

Jurczenko, E. and Teiletche, J. (2019). Expected Shortfall Asset Allocation: A Multi-Dimensional Risk Budgeting Framework. *The Journal of Alternative Investments*, 22:jai.2019.1.078.

Kolm, P. N., Tütüncü, R., and Fabozzi, F. J. (2014). 60 Years of Portfolio Optimization: Practical Challenges and Current Trends. *European Journal of Operational Research*, 234(2):356–371.

Landsman, Z. M. and Valdez, E. A. (2003). Tail Conditional Expectations for Elliptical Distributions. *North American Actuarial Journal*, 7(4):55–71.

Ledoit, O. and Wolf, M. (2003). Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance*, 10(5):603 – 621.

Ledoit, O. and Wolf, M. (2004a). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, 88(2):365 – 411.

Ledoit, O. and Wolf, M. (2004b). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2020). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*.

Lee, W. (2011). Risk-Based Asset Allocation: A New Answer to an Old Question? *Journal of Portfolio Management*, 37(4):11.

Maillard, S., Roncalli, T., and Teïletche, J. (2010). The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, 36(4):60–70.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficent Diversification of Investments*. J. Wiley.

Mausser, H. and Romanko, O. (2018). Long-Only Equal Risk Contribution Portfolios for CVaR under Discrete Distributions. *Quantitative Finance*, 18(11):1927–1945.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Mercuri, L. and Rroji, E. (2014). Parametric Risk Parity. *arXiv preprint avilable at arXiv:1409.7933*.

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is Optimized Optimal? *Financial Analysts Journal*, 45(1):31–42.

Michaud, R. O. and Michaud, R. (2008a). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press.

Michaud, R. O. and Michaud, R. (2008b). Estimation Error and Portfolio Optimization: A Resampling Solution. *Journal of Investment Management*, 6(1):8–28.

Panjer, H. H. (2001). Measurement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates. *Institute of Insurance and Pension Research, University of Waterloo, Research Report*, 01(15).

Paolella, M. S. (2019). *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. John Wiley & Sons, Chichester.

Paolella, M. S. and Polak, P. (2015a). COMFORT: A Common Market Factor Non-Gaussian Returns Model. *Journal of Econometrics*, 187(2):593 – 605.

Paolella, M. S. and Polak, P. (2015b). Portfolio Selection with Active Risk Monitoring. Technical report, Swiss Finance Institute Research Paper Series No.15-17.

Paolella, 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.

Paolella, M. S., Polak, P., and Walker, P. S. (2021). A Non-Elliptical Orthogonal GARCH Model for Portfolio Selection under Transaction Costs. *Journal of Banking & Finance*, 125:106046.

Pástor, L. and Stambaugh, R. F. (2000). Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics*, 56(3):335–381.

Polson, N. G. and Tew, B. V. (2000). Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996. *Journal of Business & Economic Statistics*, 18(2):164–173.

Protassov, R. (2004). EM-based Maximum Likelihood Parameter Estimation for Multivariate Generalized Hyperbolic Distributions with Fixed Lambda. *Statistics and Computing*, 14:1:67–77.

Qian, E. (2005). Risk Parity Portfolios: Efficient Portfolios Through True Diversification. Technical report, Panagora Asset Management.

Qian, E. (2011). Risk Parity and Diversification. *Journal of Investing*, 20(1):119.

Qian, E. (2016). *Risk Parity Fundamentals*. CRC Press.

Roncalli, T. (2013). *Introduction to Risk Parity and Budgeting*. CRC Press.

Scherer, B. (2014). Book Review: Introduction to Risk Parity and Budgeting. *Quantitative Finance*, 14(5):771–772.

Spinu, F. (2013). An Algorithm for Computing Risk Parity Weights. *Available at SSRN 2297383*.

Stefanovits, D. (2010). Equal Contributions to Risk and Portfolio Construction. *Master Thesis, Department of Mathematics, ETH Zurich*.

Suh, S. (2016). A Combination Rule for Portfolio Selection with Transaction Costs. *International Review of Finance*, 16(3):393–420.

Tasche, D. (2002). Expected Shortfall and Beyond. *Journal of Banking & Finance*, 26(7):1519–1533.

Tu, J. and Zhou, G. (2011). Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies. *Journal of Financial Economics*, 99(1):204–215.

Tütüncü, R. H. and Koenig, M. (2004). Robust Asset Allocation. *Annals of Operations Research*, 132(1-4):157–187.

Unger, A. (2014). *The Use of Risk Budgets in Portfolio Optimization*. Springer.