Portfolio optimization for heavy-tailed assets: Extreme Risk Index vs. Markowitz

Georg Mainik Georgi Mitov Ludger Rüschendorf*

October 17, 2013

Abstract

Using daily returns of the S&P 500 stocks from 2001 to 2011, we perform a backtesting study of the portfolio optimization strategy based on the extreme risk index (ERI). This method uses multivariate extreme value theory to minimize the probability of large portfolio losses. With more than 400 stocks to choose from, our study applies extreme value techniques in portfolio management on a large scale. We compare the performance of this strategy with the Markowitz approach and investigate how the ERI method can be applied most effectively. Our results show that the annualized return of the ERI strategy is particularly high for assets with heavy tails. The comparison also includes maximal drawdown, transaction costs, portfolio concentration, and asset diversity in the portfolio. In addition to that we study the impact of an alternative tail index estimator.

JEL Classification: G11, C14

Keywords: portfolio optimization, heavy tails, extreme risk index, extreme value theory,

*corresponding author:
Phone: +49 761 203 5664, Fax: +49 761 203 5661, ruschen@stochastik.uni-freiburg.de

1 Introduction

In this paper we propose and test a portfolio optimization strategy that aims to improve the portfolio return by stabilizing the portfolio value. Minimizing the probability of large drawdowns, this strategy can help to retrieve the portfolio value as good as possible also in times of high risk in the markets. This intended performance is, of course, not a new aim in portfolio management, and it became even more vital since the default of Lehman Brothers in 2008. The following years of financial crisis have demonstrated that the technical progress of financial markets and their globalization have also brought up some new challenges. One of these challenges is the need for diversification strategies that account for strong drawdowns and increasing dependence of asset returns in crisis periods. This has raised the relevance of non-Gaussian models, tail dependence, and quantile based risk measures in portfolio optimization [23, 11, 5, 2, 4, 9, 21, 6, 13, 26, 18].

The industry standard of portfolio diversification goes back to Markowitz [19], whose fundamental results shaped the modern portfolio theory and were acknowledged with a Nobel prize in 1990. Approved and improved over decades, the mean-variance approach of Markowitz is the natural benchmark to any new development on in this field. However, despite its well-deserved high reputation, some aspects of the mean-variance portfolio optimization are worth an additional discussion. Measuring risk by variance and dependence by correlation, this method relies on the assumption of multivariate normal – or at least elliptical – distribution for the asset returns. This kind of joint distribution implies that the dependence between the assets is the same in benign and in turbulent market conditions. Unfortunately, there are several reasons to question this assumption, such as panic sales in market turmoils, structural dependence in some industries, and the procyclical effect of financial regulation.

Another issue is the estimation of parameters used to calculate the Markowitz portfolio. It is well known that the estimation of covariance matrices from historical market data can be difficult, and this bias can have a crucial effect on the investment performance. To some extent, these technical difficulties can be addressed by advanced estimation methods – see [22] and [14], just to call a few. There is, however, a general issue that is persistent by nature. With increasingly heavy tails, second moments of asset returns cease to exist. In particular, this issue affects all non-Gaussian α-stable models, which are often used to improve the modelling of return tails. In such cases, covariances estimates are not informative. Thus the mean-variance approach tends to face its limitations especially in crisis periods, when financial returns behave in extreme way. Several modifications addressing this issue have been discussed; see, e.g., [23].

The present paper comprises an empirical study of the portfolio optimization strategy proposed in [18]. It is based on the so-called Extreme Risk Index (ERI), which quantifies the impact of heavy, dependent tails of asset returns on the tail of the portfolio return. We apply this strategy to the daily return data of the S&P 500 stocks in the period from November 2007 to September 2011 and compare the investment performance with the Markowitz approach. The computation of portfolio weights utilizes the data from the six years prior to each trading day. In addition to the portfolio value we also track some other characteristics related to portfolio structure, degree of diversification, and transaction costs.

In the first round of our backtesting experiments we apply ERI optimization to all S&P 500 stocks with full history in our data set (444 out of 500). In this basic setting the ERI based algorithm slightly outperforms Markowitz in terms of annualized returns (6.8% vs. 5.8%). Both algorithms significantly outperform the S&P 500 index, which has the annualized return of -5.2%.

As next step we subdivide the stocks into three groups according to their tail characteristics. Our results show that ERI optimization is particularly useful for assets with heavy tails. On this asset group it clearly outperforms Markowitz and yields an annualized return of 11.5%. This is impressive compared to the 5.0% achieved with the Markowitz strategy, and even more so because the backtesting period includes the recent financial crisis. Tracking the portfolio turnover, we found that the ERI strategy tends to increase the transaction costs. However, the turnover of the ERI optimal portfolio for the group with heavy tails is lower than the turnover of the Markowitz portfolio in the basic experiment without grouping.

Another remarkable detail is that the ERI optimization is a strategy that ignores expected returns and constructs the portfolio only from the objective of risk minimization. Thus there may be even more space for improvement in ERI based strategies. Since there are also various improvements to the implementation of the Markowitz approach, we consider it as fair to choose very basic implementations of both methods for the first comparison. Our results suggest that ERI optimization can be a useful alternative for portfolio selection in risky asset classes. In some sense, this strategy seems to earn the reward that the economic theory promises for the higher risk of heavier tails.

The paper is organized as follows. The alternative portfolio optimization algorithm and its technical backgrounds are introduced in Section 2. In Section 3 we give an outline of the data used in the backtesting study, define the estimator for the optimal portfolio, and introduce all additional portfolio characteristics to be tracked. Detailed results of the backtesting experiments are presented and discussed in Section 4. Conclusions are given in Section 5.

2 Theoretical backgrounds

2.1 Asset and portfolio losses

Let $S_i(t)$ denote prices of assets S_i, $i = 1, \dots, N$, at times $t = 0, 1, \dots, T$. Focusing on the downside risk, let $X_i(t)$ denote the *logarithmic losses* of the assets S_i,

$$X_i(t) := -\log\left(\frac{S_i(t)}{S_i(t-1)}\right) = \log S_i(t-1) - \log S_i(t), \tag{1}$$

and let $\widetilde{X}_i(t)$ denote the corresponding *relative losses*:

$$\widetilde{X}_i(t) := \frac{S_i(t-1) - S_i(t)}{S_i(t)} = \frac{S_i(t-1)}{S_i(t)} - 1.$$

For daily stock returns, X_i and $\widetilde{X}_i$ are almost identical because $\widetilde{X}_i$ is the first-order Taylor approximation to the logarithmic loss X_i.

This approximation also extends to asset portfolios. Consider an investment strategy (static or one-period) diversifying a unit capital over the assets $S_1, \dots, S_N$. It can be represented by a vector w of portfolio weights, $w \in H_1 := \{x \in \mathbb{R}^N : \sum_{i=1}^N x_i = 1\}$. Excluding short positions, the portfolio set can be restricted to the unit simplex $\Delta^N := \{w \in [0,1]^N : \sum_{i=1}^N w_i = 1\}$. This is the portfolio set we will work with from now on. Each component $w_i \geq 0$ corresponds to the fraction of the total capital invested in S_i, and the relative portfolio loss is equal to the scalar product $w^T \widetilde{X}(t) := \sum_{i=1}^N w_i \widetilde{X}_i(t)$ of the portfolio vector w and the relative loss vector $\widetilde{X}(t) = (\widetilde{X}_1(t), \dots, \widetilde{X}_N(t))$:

$$\sum_{i=1}^N \frac{w_i}{S_i(t-1)} (S_i(t-1) - S_i(t)) = w^T \widetilde{X}(t). \tag{2}$$

Thus the scalar product $w^T X(t)$ for the logarithmic loss vector $X(t) := (X_1(t), \ldots, X_N(t))$ is the first-order Taylor approximation to $w^T \widetilde{X}$. This kind of approximation is also relevant to the Markowitz approach, which is typically applied to logarithmic returns.

2.2 Multivariate regular variation

To define the Extreme Risk Index (ERI) of the random vector $X(t)$, we recollect the notion of *multivariate regular variation* (MRV). A random vector $X = (X_1, \ldots, X_N)$ is MRV if the joint distribution of its polar coordinates $R := \|X\|_1 := \sum_{i=1}^N |X_i|$ and $Z := \|X\|_1^{-1} X$ satisfies

$$\mathcal{L}(r^{-1}R, Z | R > r) \xrightarrow{w} \rho_\alpha \otimes \Psi, \quad r \to \infty, \tag{3}$$

where Ψ is a probability measure on the 1-norm unit sphere $\mathbb{S}_1^N$ and ρ_α is the Pareto distribution: $\rho_\alpha(s, \infty) = s^{-\alpha}$, $s \geq 1$. The symbol $\xrightarrow{w}$ in (3) represents the weak convergence of probability measures. Besides (3), there are several other equivalent definitions of MRV; for more details we refer to [24]. The parameter $\alpha > 0$ is called *tail index*. It separates finite moments of R from infinite ones in the sense that $ER^\beta < \infty$ for $\beta < \alpha$ and $ER^\beta = \infty$ for $\beta > \alpha$. In the non-degenerate case, same moment explosion occurs for all components X_i of the random vector X. The measure Ψ is called *spectral* (or *angular*) *measure* of X and describes the asymptotic distribution of excess directions for the random vector X.

Intuitively speaking, MRV means that the radius R has a polynomial tail and is asymptotically (i.e., for large R) independent of the angular part Z. Moreover, if a measurable set $A \subset \mathrm{R}^N$ is sufficiently far away from the origin, i.e., if $\|x\|_1 \geq t$ for all $x \in A$ with some large t, then

$$P(X \in sA) \simeq s^{-\alpha} P(X \in A) \tag{4}$$

for $s \geq 1$ and $sA := \{sx : x \in A\}$. The scaling property (4) allows to extrapolate from large losses to extremely large ones, which even may be beyond the range of the observed data. Approximations of this kind are the key idea of the Extreme Value Theory (cf. [8]).

Many popular models are MRV. In particular, this is the case for multivariate t and multivariate α-stable distributions (cf. [12, 1]). In the latter case, the stability index α is also the tail index, and the spectral measure characterizing the multivariate stability property is a constant multiple of

Ψ from (3). In all these models, the components X_i are *tail equivalent* in the sense that $P(X_i > r)/P(X_j > r) \to c_{i,j} > 0$ as $r \to \infty$ for all $i, j \in \{1, \ldots, N\}$. This is equivalent to the following non-degeneracy condition for the angular measure Ψ:

$$\Psi\{x \in \mathbb{S}_1^N : x_i = 0\} < 1$$

for $i = 1, \ldots, N$.

It should be noted that the MRV assumption (3) is of asymptotic nature and that it is also quite restrictive. MRV models are often criticized for excluding even slightly different tail indices α_i for the components X_i. However, this criticism also affects the multivariate t and multivariate α-stable models, which are widely accepted in practice despite the resulting restriction to equal α_i. It is indeed true that, estimating the tail index α_i for each component X_i separately, one would hardly ever obtain identical values for different i. But on the other hand, the confidence intervals for α_i often overlap, so that a MRV model may be close enough to reality and provide a useful result.

The major reason why MRV models can be useful in practice is that the practical questions are non-asymptotic. In fact, it is not the restrictive asymptotic relation (3) that matters, but the scaling property (4). If (4) is sufficiently close to reality in the range that is relevant to the application, the eventual violation of (3) further out in the tails does not influence the result too much.

Practical applications often involve heuristics of this kind. In particular, if S_i are stock prices and hence non-negative, then the relative losses $\widetilde{X}_i$ are bounded by 1. Going sufficiently far out into the tail, one must observe quite different behaviours for the relative portfolio loss $w^T\widetilde{X}$ and the logarithmic approximation w^TX. However, with typical daily return values in the low percentage area and values around 10% occurring only in crisis times, relative asset losses do exhibit polynomial scaling of the type

$$\frac{P(\widetilde{X}_i > rs)}{P(\widetilde{X}_i > r)} \simeq s^{-\alpha}. \tag{5}$$

Hence we are lucky to remain in the area where X and $\widetilde{X}$ can be treated as if they both were MRV, and the approximation $w^T\widetilde{X} \simeq w^TX$ works reasonably well. Thus, even though the scaling property (5) eventually breaks down if rs gets too close to 1, it has some useful consequences in the application range. This is confirmed by our backtesting results.

2.3 Portfolio optimization via Extreme Risk Index

The MRV assumption (3) implies that

$$\lim_{r\to\infty} \frac{P(w^T X > r)}{P(\|X\|_1 > r)} = \gamma_w := \int_{\mathbb{S}_1^N} \max(0, w^T z)^\alpha \, d\Psi(z) \tag{6}$$

(cf. [18] and [16, Lemma 2.2]). This implies that for any portfolio vectors $v, w \in \Delta^N$ and large $r > 0$

$$\frac{P(v^T X > r)}{P(w^T X > r)} \simeq \frac{\gamma_v}{\gamma_w}. \tag{7}$$

Moreover, for $\lambda \leq 1$ close to 1 one obtains that

$$\frac{\mathrm{VaR}_\lambda(v^T X)}{\mathrm{VaR}_\lambda(w^T X)} \simeq \left(\frac{\gamma_v}{\gamma_w}\right)^{1/\alpha} \tag{8}$$

(cf. [18] and [17, Corollary 2.3]). Motivated by (7) and (8), the functional $\gamma_w = \gamma_w(\Psi, \alpha)$ is called *Extreme Risk Index* (ERI). Minimizing the function $w \mapsto \gamma_w$, one obtains a portfolio that minimizes the loss for large $\|X\|$, i.e., in case of crisis events. In precise mathematical terms, one minimizes $\mathrm{VaR}_\lambda(w^T X)$ for $\lambda \to 1$. The practical meaning of this procedure is the utilization of the scaling property (4) to obtain a portfolio that minimizes the downside risk during a market crash. This approximate result is not perfect, but it can be a step into the right direction.

Based on the integral representation (6), the following portfolio optimization approach was proposed in [18]:

- Estimate γ_w by plugging appropriate estimates for α and Ψ into (6);
- Estimate the optimal portfolio by minimizing the resulting estimator $\hat{\gamma}_w$ with respect to w.

The general properties of the optimization problem are discussed in [18], [15], and [17]. In particular, it is known that the function $w \mapsto \hat{\gamma}_w$ is convex for $\alpha > 1$. Thus, given that the expectations of X_i are finite, a typical optimal portfolio would diversify over multiple assets. The consistency of the plug-in estimator $\hat{\gamma}_w$ and of the resulting estimated optimal portfolio w^* in a strict theoretical sense is studied in [18, 15, 16].

3 Outline of the backtesting study

3.1 The data

The contribution of the present paper is a backtesting study of the ERI based portfolio optimization approach on real market data. Our data set comprises all constituents of the S&P 500 market index that have a full history for the period of 10 years back from 19-Oct-2011. These are 444 stocks out of 500. For each date of the backtest period 19-Oct-2007 to 19-Oct-2011 the estimation of the optimal portfolio is based on the 1500 foregoing observations – approximately 6 years of history – for all stocks back in time. For example, the optimal portfolio for 19-Oct-2007 is estimated from the stock price data for the period (19-Oct-2001 to 18-Oct-2007).

Our computations are based on the logarithmic losses $X_i(t)$ as defined in (1). As already mentioned above, we exclude short positions. This basic framework is most natural for the comparison of portfolio strategies. The asset index i varies between 1 and $N = 444$, and the time index t takes values between 1 and $T = 2509$ (1500 days history + 1009 days in the backtest period) To estimate α and Ψ, we transform the (logarithmic) loss vectors $X(t)$ into polar coordinates

$$(R(t), Z(t)) = (||X(t)||_1, ||X(t)||_1^{-1} X(t)), \quad t = 1, \ldots, T.$$

3.2 The estimators and the algorithms

We estimate α by applying the *Hill estimator* to the radial parts $R(t)$:

$$\hat{\alpha} = \frac{k}{\sum_{j=1}^{k} \log(R_{(j),t}/R_{(k+1),t})} \tag{9}$$

where $t > 1500$ and $R_{(1),t} \geq \ldots \geq R_{(1500),t}$ is the descending order statistic of the radial parts $R(t-1500), \ldots, R(t-1)$ and $k = 150$. That is, out of the 1500 data points in the historical observation window $t-1500, \ldots, t-1$ we use the 10% with largest radial parts. Going back to [10], the Hill estimator is the most prototypical approach for the estimation of the tail index α. The choice of k determines which observations are assumed to describe the tail behaviour. Another important criterion for the choice of k is the trade-off between the bias, which typically increases for large k, and the variance of the estimator, which increases for small k. In addition to the static 10%-rule we also consider the adaptive approach proposed in [20]. See [25, 7, 3] for further related methods.

As proposed in [18], we estimate Ψ by the empirical measure of the angular parts from observations with largest radial parts. More specifically, we use the same 10% data points (the so-called *tail fraction*) in the moving observation window that were used to obtain $\hat{\alpha}$. The resulting estimator γ_w is

$$\hat{\gamma}_w(t) = \frac{1}{k} \sum_{j=1}^{k} \max(0, w^T Z(i_{j,t}))^{\hat{\alpha}},$$

where $i_{j,t}$ is the sample index of the order statistic $R_{(j),t}$ in the full data set:

$$R_{(j),t} = R(i_{j,t}), \quad j = 1, \ldots, 1500, t = 1501, \ldots, T.$$

The resulting estimate of the optimal portfolio $w^*(t)$ on the trading day t is the portfolio vector $w \in \Delta^N$ that minimizes $\hat{\gamma}_w(t)$:

$$\hat{\gamma}_{w^*(t)} = \min_{w \in \Delta^N} \hat{\gamma}_w(t).$$

Finally, the estimated optimal portfolio $w^*(t)$ is used to compose the portfolio for the trading day t. The resulting (relative) portfolio return is calculated by substituting $w^*(t)$ in (2).

The procedure outlined above is repeated for all trading days $t > 1500$. For instance, the optimal portfolio for 22-Oct-2007 is based on the observation window from 29-Oct-2001 to 21-Oct-2007, whereas for 23-Oct-2007 we use the observation window from 30-Oct-2001 to 22-Oct-2007, and so on.

The benchmark for this portfolio optimization algorithm is the Markowitz approach applied to logarithmic asset returns, with the same moving observation window of 1500 points and empirical estimators for the covariance matrix. Similarly to the ERI approach, our implementation of the Markowitz approach chooses the portfolio with minimal risk, i.e. with minimal variance. There are two reasons for this choice. On the one hand, ERI minimization is also a pure risk minimization procedure, so that ignoring estimates of the expected return in the Markowitz benchmark increases the fairness of competition. On the other hand, computation of a "real" Markowitz efficient portfolio would require some target return, target risk, or target risk-return ratio. Thus the performance of this portfolio and the comparison results can be strongly influenced by the target parameters.

4 Empirical results

4.1 Basic setting without sub-groups

We start with the most crude application of the ERI minimization strategy, estimating the tail index from the radial parts of the random vector $(X_1, \ldots, X_{444})$ of all stock retruns involved in our study. The resulting estimate $\hat{\alpha} = \hat{\alpha}(t)$ varies in time, but it is applied to all $N = 444$ stocks as if their joint distribution were MRV. This is a very courageous assumption, but even in this case we see some useful results. A first impression of these results is given in Figure 1, where the value of the ERI optimal portfolio is compared to the performance of the Markowitz approach and to the S&P 500 index. The graphic suggests that the value of the ERI based portfolio is more stable during market crashes. On the other hand, the Markowitz portfolio

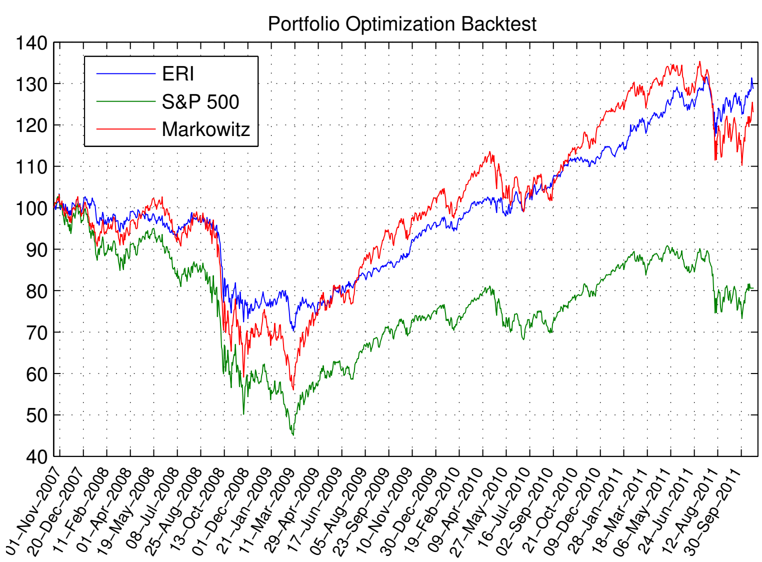


Figure 1: Portfolio optimization backtest for the ERI minimization strategy under the assumption that all all stock returns have the same tail index α. The resulting portfolio value of the ERI strategy and its peers (Markowitz approach and S&P 500) is scaled to 100 for the first date of the backtest period.

seems to catch up again during recovery periods. This may be explained by the fact that the ERI approach only looks at the potential losses, whereas the Markowitz approach also tries to assess potential gains. Surprisingly, the overall result is similar, and it seems that the ERI strategy – even in its crudest implementation – has a potential to stabilize the portfolio value in crises.

	ERI	Markowitz	S&P 500
CR = Cumulative Return	30.07 %	25.48%	-19.38%
AR = Annualized Return	6.76 %	5.81%	-5.22%
AS = Annualized Sharpe	0.4715	0.3469	-0.0462
AST = Annualized $\mathrm{STARR}_{0.95}$	0.1926	0.1410	-0.0187
MD = Max Drawdown	46.61 %	58.61%	56.34%
AC = Average Concentration Coefficient	8.69	127.22	444
AT = Average Turnover	0.0400	0.0272	N/A
PCA = First PCA factor Explained Variance	31.32%	35.48%	N/A

Table 1: Backtest statistics for the ERI minimization strategy in the basic setting (applied to all stocks at once) vs. Markowitz approach and S&P 500.

Further characteristics of the basic ERI approach compared to Markowitz and S&P 500 are shown in Table 1. The numbers show that the ERI strategy indeed outperforms Markowitz in many respects. In particular, the ERI optimal portfolio gives higher cumulative returns and a higher Sharpe ratio, whereas the maximal drawdown is lower than with the Markowitz strategy. An extension of the Sharpe ratio based on the Expected Shortfall (ES) is the STARR ratio (cf. [23]):

$$\mathrm{STARR}_\lambda(Z) := \frac{\mathrm{E}(Z - r_f)}{\mathrm{ES}_\lambda(Z - r_f)}$$

where r_f is the risk-free interest rate and λ is a confidence level close to 1. The backtested STARR is also higher for the ERI strategy than for the Markowitz approach. The computation of the Sharpe and STARR ratios is based on empirical estimators for the expectation and for the Expected Shortfall. In particular, the estimate of $\mathrm{ES}_{0.95}$ over the backtesting period of 1009 days is based on 51 largest observations of the portfolio loss. Since a risk free rate on a daily scale is both difficult to determine and negligibly small, we set $r_f = 0$.

To measure the portfolio stock concentration, we compute the Concentration Coefficient (CC). It is defined as

$$\mathrm{CC}(t) := \left(\sum_{i=1}^{n} w_i^2(t) \right)^{-1} \tag{10}$$

where $w_i(t)$ is the relative weight of the asset i in the investment portfolio at time t. Conceptually, this approach is well known in measures of industrial concentration, where it is called as the Herfindahl–Hirschman index (HHI). Brandes Institute introduced the concentration coefficient by inverting the HHI.

The CC of an equally weighted portfolio is identical with the number of assets. As the portfolio becomes concentrated on fewer assets, the CC decreases proportionally. The numbers in Table 1 indicate that the ERI strategy is quite selective, whereas the number of stocks in the Markowitz portfolio is on the same scale with the total number of assets.

To assess the level of diversification provided by each optimization algorithm, we performed Principal Component Analysis (PCA) over the returns of all stocks relevant to the corresponding portfolios. We defined relevance via portfolio weights assigned by the algorithms and restricted PCA to the stocks with portfolio weights higher than 0.01%. Then we estimated the portion of the sample variance explained by the first PCA factor and averaged these daily estimates over the backtesting period. The lower the average portion of sample variance explained by the first PCA factor, the higher is the portfolio diversification. The numbers in Table 1 are quite surprising: despite the significantly higher concentration, the diversification level of the ERI based portfolio is higher than that of the Markowitz strategy.

The only performance characteristic where ERI stays behind Markowitz is the portfolio turnover, which is a proxy to the transaction costs of a strategy. We use a definition of portfolio turnover that is based on the absolute values of the rebalancing trades:

$$\tau(t) := \sum_{i=1}^{n} |w_i(t) - w_i(t_-)|$$

where $w_i(t)$ is the (relative) portfolio weight of the asset i after rebalancing (according to the optimization strategy) at time t, and $w_i(t_-)$ is the portfolio weight of the asset i before rebalancing at time t, i.e., at the end of the trading period $t-1$. Averages of $\tau(t)$ over all t in the backtesting period are given in Table 1. The average turnover of the ERI optimal portfolio (0.0400) is higher than that of the Markowitz portfolio (0.0272).

Some technical details. For calculation of the portfolio value we used the relative returns and do not expect much difference when using logarithmic approximations. In the calculation of STARR and Sharpe ratio we do not use risk free rates since these are on a daily scale very small and thus not influence the ratio calculations. For the estimation of ES in STARR we use the average of all sample values smaller than 95% of the VaR of the sample. Our backtest period is of length 1009 and thus $n = 51$.

4.2 Grouping the stocks with similar α

In the previous section we treated all stocks as if their (logarithmic) returns X_i had the same tail index α. This simplification can influence the quantitative and qualitative results. To obtain a better insight, we divide the stocks into three different groups with respect to their individual α and compare the performance of the portfolio optimization strategies on each of these groups. Figure 2 shows the histogram of the estimates of the tail index α for different stocks on the first day of the backtesting period ($t = 1501$).

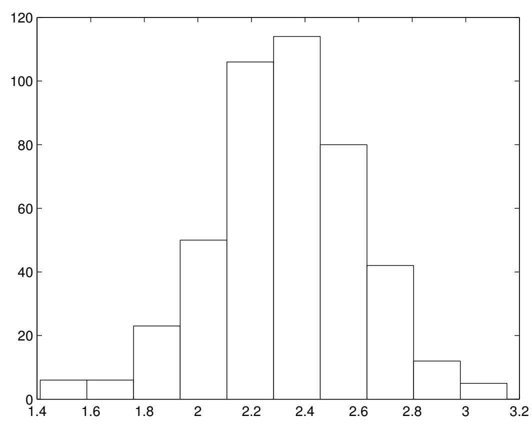


Figure 2: Estimated values of the tail index α for different stocks on the first day of the backtesting period

We consider the following groups:

1. all stocks with $\alpha \leq 2.2$
2. all stocks with $\alpha \in (2.2, 2.6)$
3. all stocks with $\alpha \geq 2.6$

The first group contains 134, the second 243, and the third one 67 stocks. These groups remained static during the backtesting period. That is, the estimated α on the first day of the backtesting period determines in which group each stock is placed.

Selection from the set of stocks with $\alpha \leq 2.2$

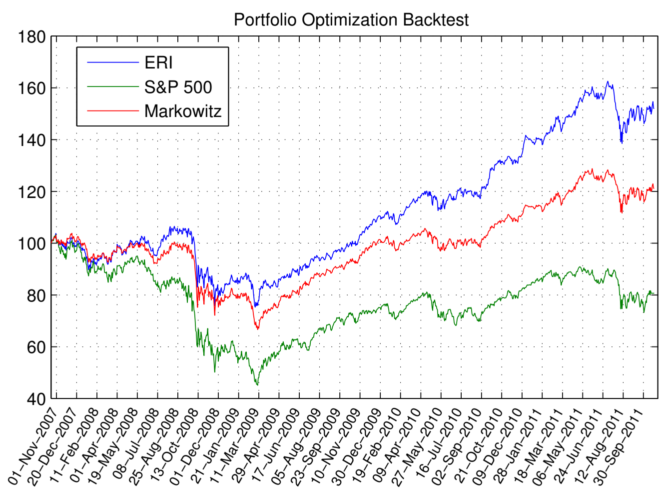


Figure 3: Portfolio optimization backtest. Stocks with $\alpha \leq 2.2$

The backtesting results on stocks with tail index $\alpha \leq 2$ are summarized in Figure 3 and Table 2. In this case ERI minimization clearly outperforms the Markowitz approach and yields an impressive annualized return of 11.48%. This is more than the double of the 4.99% achieved with the Markowitz portfolio. The Sharpe and STARR ratios of the ERI strategy are also clearly higher than with Markowitz. The concentration of both portfolios is on the same scale, but still a bit higher for the ERI based one. Similarly to the

	ERI	Markowitz	S&P 500
CR = Cumulative Return	54.70%	21.58%	-19.38%
AR = Annualized Return	11.48%	4.99%	-5.22%
AS = Annualized Sharpe	0.6623	0.3546	-0.0462
AST = Annualized $\text{STARR}_{0.95}$	0.2695	0.1430	-0.0187
MD = Max Drawdown	53.67%	48.03%	56.34%
AC = Average Concentration Coefficient	7.4499	10.9712	134
AT = Average Turnover	0.0269	0.0154	N/A
PCA = First PCA factor Explained Variance	35.02%	33.33%	N/A

Table 2: Backtest statistics. Stocks with $\alpha \leq 2.2$

basic backtesting set-up on all S&P 500 stocks, the ERI strategy produces a higher portfolio turnover (0.0269 vs. 0.0154 with Markowitz). However, both values are lower than the average turnover of the Markowitz portfolio in the basic setting (0.0272).

These results suggest that the ERI optimal strategy is particularly useful for optimizing portfolios of stocks with heavy tails, in our case of 134 out of 444 stocks. This is to be expected since the ERI methodology was developed for heavy-tailed MRV models. Beyond that, there is also a statistical reason for the inferior performance of the Markowitz approach in the present setting. Estimation of covariances becomes increasingly difficult for heavier tails, and for $\alpha < 2$ the covariances (and hence correlations) do not even exist. Thus empirical covariances in the Markowitz portfolio can push the investor into the wrong direction.

Selection from the set of stocks with $\alpha \in (2.2, 2.6)$

If the stock selection is restricted to those with α between 2.2 and 2.6, the annualized return of the ERI based portfolio (7.93%) is somewhat above the Markowitz benchmark (6.96%). While the returns are on the same scale, the volatility of the Markowitz portfolio is much higher. Thus ERI optimization clearly outperforms Markowitz in terms of Sharpe ratio, STARR (both higher

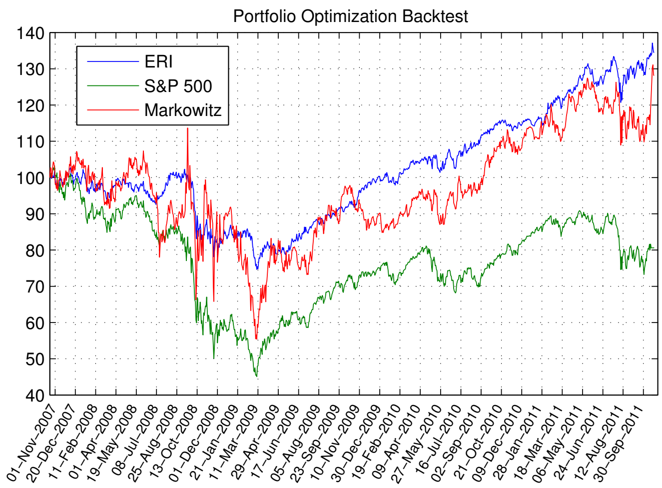


Figure 4: Portfolio optimization backtest. Stocks with $\alpha \in (2.2, 2.6)$

	ERI	Markowitz	S&P 500
CR = Cumulative Return	35.87%	31.00%	-19.38%
AR = Annualized Return	7.93%	6.96%	-5.22%
AS = Annualized Sharpe	0.5448	0.3711	-0.0462
AST = Annualized STARR$_{0.95}$	0.2306	0.1517	-0.0187
MD = Max Drawdown	45.56%	57.70%	56.34%
AC = Average Concentration Coefficient	7.3987	1.0000	243
AT = Average Turnover	0.0249	0.0000	N/A
PCA = First PCA factor Explained Variance	32.78%	100.00%	N/A

Table 3: Backtest statistics. Stocks with $\alpha \in (2.2, 2.6)$

for ERI), and maximal drawdown (lower for ERI). It is somewhat astonishing that the PCA of Markowitz is 100% i.e. the Markowitz min variance algorithm selects only one stock.

Selection from the set of stocks with $\alpha \geq 2.6$

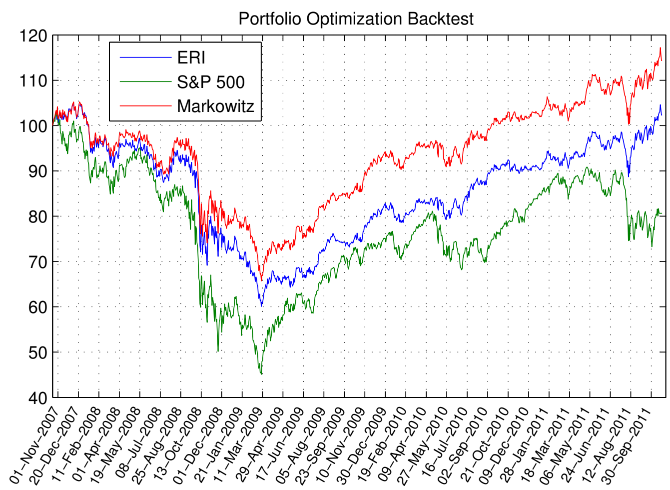


Figure 5: Portfolio optimization backtest. Stocks with $\alpha \geq 2.6$

For stocks with $\alpha > 2.6$ (and hence lightest tails), the performance of the ERI minimization strategy stays behind Markowitz in terms of annualized return, Sharpe ratio, STARR, and turnover. The maximal drawdown and the diversification level in terms of PCA is similar for both competing strategies. The portfolio concentration is on the same level, and slightly higher for the ERI optimal portfolio.

Thus the impressive advantage of the ERI minimization strategy seems to be restricted to stocks with pronounced heavy-tail behaviour. This advantage turns into near parity for stocks with moderately heavy tails. For light-tailed stocks the Markowitz strategy yields much better results. These findings perfectly accord with model assumptions underlying these two methodologies: Markowitz uses covariances, and ERI minimization is particularly applicable in cases when covariances do not exist or cannot be estimated reliably. On the other hand, ERI minimization strongly relies on the estimation of the

	ERI	Markowitz	S&P 500
CR = Cumulative Return	3.47%	13.62%	-19.38%
AR = Annualized Return	0.85%	3.23%	-5.22%
AS = Annualized Sharpe	0.1397	0.2636	-0.0462
AST = Annualized $\mathrm{STARR}_{0.95}$	0.0581	0.1114	-0.0187
MD = Max Drawdown	42.58%	43.43%	56.34%
AC = Average Concentration Coefficient	3.4817	4.4715	67
AT = Average Turnover	0.0165	0.0091	N/A
PCA = First PCA factor Explained Variance	54.92%	52.43%	N/A

Table 4: Backtest statistics. $\alpha \geq 2.6$

tail index α, which is known to become increasingly difficult for lighter tails – see, e.g., [8].

4.3 Backtesting with an alternative estimator for α

To assess the suitablity of the estimator we used for α, we repeated our backtesting experiments with another estimation approach. The Hill estimator in (9) uses the tail fraction size k as a parameter. The foregoing results are based on a static 10% rule, i.e. $k = 150$. It is well known that the choice of the tail fraction size k can have a strong influence on the resulting estimates – see, e.g., [8]. Thus, as an alternative to the static 10% rule, we tried the recent adaptive approach by Nguyen and Samorodnitsky [20], which involves sequential statistical testing for polynomial tails. The results of this backtesting study are outlined below.

Optimization over the entire set of stocks

Figure 6 and Table 5 represent the basic setting without grouping the stocks according to the estimated tail index α. It is a bit surprising that the adaptive choice of the tail fraction size k does not improve the performance of the ERI based strategy. The annualized return is significantly lower than with the static 10% rule. The overall result clearly stays behind the Markowitz

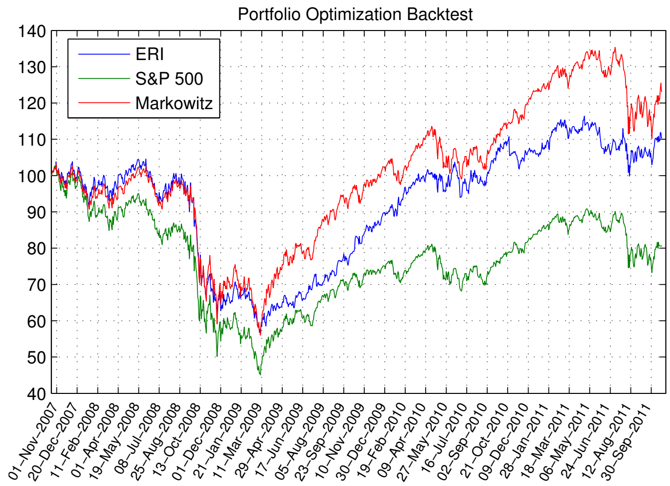


Figure 6: Alternative estimator $\hat{\alpha}$: portfolio optimization backtest in the basic set-up (on all S&P 500 stocks)

	ERI	Markowitz	S&P 500
CR = Cumulative Return	11.76%	25.48%	-19.38%
AR = Annualized Return	2.81%	5.81%	-5.22%
AS = Annualized Sharpe	0.2360	0.3469	-0.0462
AST = Annualized $\text{STARR}_{0.95}$	0.0939	0.1410	-0.0187
MD = Max Drawdown	51.39%	58.61%	56.34%
AC = Average Concentration Coefficient	64.33	127.22	444
AT = Average Turnover	0.0381	0.0272	N/A
PCA = First PCA factor Explained Variance	46.48%	35.48%	N/A

Table 5: Alternative $\hat{\alpha}$: backtest statistics in the basic set-up.

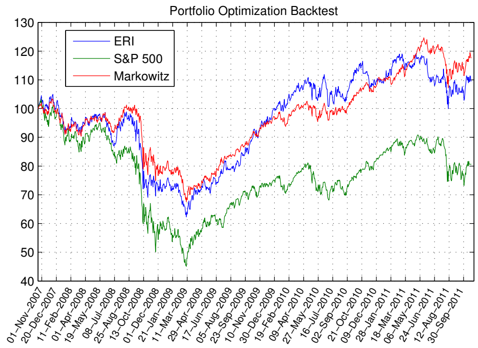
benchmark. The only aspect where ERI is still better is the maximal drawdown, but it cannot compensate for the lower overall return. The reason for this outcome is the lower value of the tail fraction size k that is selected by the adaptive approach. Typical values are about 25, and all values are lower than 150 that come from the static 10% rule. Thus the adaptive approach looks too far in the tail, where the scaling of excess probabilities may already be different from the scaling in the application range.

Grouping the stocks according to the estimated α

As next step, we grouped the stocks according to their estimates. On average, the Nguyen–Samorodnitsky estimator gave higher values of α, i.e., it indicated lighter tails than the static 10% rule. Therefore we chose a different grouping of the α values: $\alpha \leq 2.7$, $\alpha \in (2.7, 4.5)$, and $\alpha \geq 4.5$. The backtesting results are presented in Table 6.

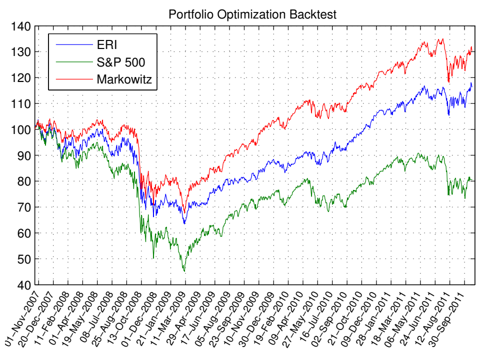
In all three cases the annualized return of the ERI strategy is lower than that of the Markowitz portfolio. Interestingly, the worst performance of the ERI based strategy occurs in the middle group, and not in the group with lightest tails. Possible explanations here may be the different composition of the three groups (heavy, moderate, or light tails) and also the different values of α used in the portfolio optimization algorithm.

All in all we can conclude that the adaptive choice of the tail fraction size k can be problematic in real applications. This can be explained by the tail orientation of the Nguyen–Samorodnitsky approach. Roughly speaking, it tests for polynomial tails and chooses the largest value of k for which the test is still positive. While this is perfectly reasonable for data from an exact MRV model, there are at least two reasons why this method can fail on real data. First, if the data fails to satisfy the MRV assumption far out in the tail, the subsequent testing for small values of k can be misleading. The second reason was already discussed in Section 2.2: If the polynomial scaling changes for different severities, then the scaling behaviour of the distribution in the application area can differ from what is suggested by the true, but too asymptotic tail index. Our backtesting results show that these issues are highly relevant in practice.



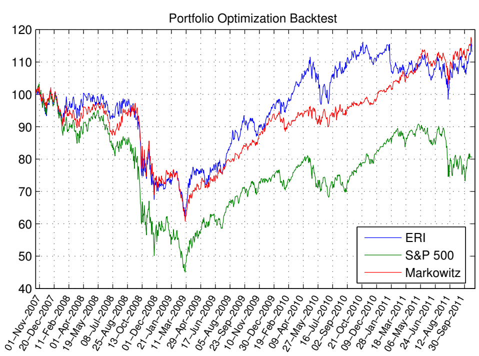
	ERI	Markowitz	S&P 500
CR	11.33%	18.75%	-19.38%
AR	2.71%	4.37%	-5.22%
AS	0.2317	0.3260	-0.0462
AST	0.0958	0.1351	-0.0187
MD	47.76%	45.79%	56.34%
AC	9.0194	10.0857	107
AT	0.0385	0.0148	N/A
PCA	43.65%	30.87%	N/A

(a) $\alpha \leq 2.7$



	ERI	Markowitz	S&P 500
CR	17.67%	31.97%	-19.38%
AR	4.14%	7.15%	-5.22%
AS	0.3015	0.4578	-0.0462
AST	0.1202	0.1852	-0.0187
MD	46.26%	49.93%	56.34%
AC	48.1646	28.0871	252
AT	0.0384	0.0201	N/A
PCA	48.29%	32.56%	N/A

(b) $\alpha \in (2.7, 4.5)$



	ERI	Markowitz	S&P 500
CR	15.23%	17.02%	-19.38%
AR	3.59%	3.99%	-5.22%
AS	0.2699	0.2991	-0.0462
AST	0.1155	0.1233	-0.0187
MD	46.43%	48.29%	56.34%
AC	6.7252	3.9777	85
AT	0.0326	0.0113	N/A
PCA	58.04%	45.34%	N/A

(c) $\alpha \geq 4.5$

Table 6: Alternative $\hat{\alpha}$: backtest statistics on stocks (CR = Cumulative Return; AR = Annualized Return; AS = Annualized Sharpe; AST = Annualized $\text{STARR}_{0.95}$; MD = Max Drawdown; AC = Average Concentration Coefficient; AT = Average Turnover; PCA = First PCA factor Explained Variance)

4.4 Behaviour of portfolio characteristics over time

We conclude our analysis by a comparison of the ways the competing portfolios behave over time. This allows for deeper insight and allows to discover some more points of difference.

Concentration and portfolio composition

We start with the development of the concentration coefficient (CC) introduced in (10). Its behaviour over time in the basic set-up (no grouping of stocks according to α) is shown in Figure 7. This graphic shows that the number of stocks in the Markowitz portfolio is permanently about 10 times higher than in the ERI optimal portfolio. The CC oscillation pattern suggests that the ERI based portfolio is more volatile in the crisis and much less volatile in benign periods.

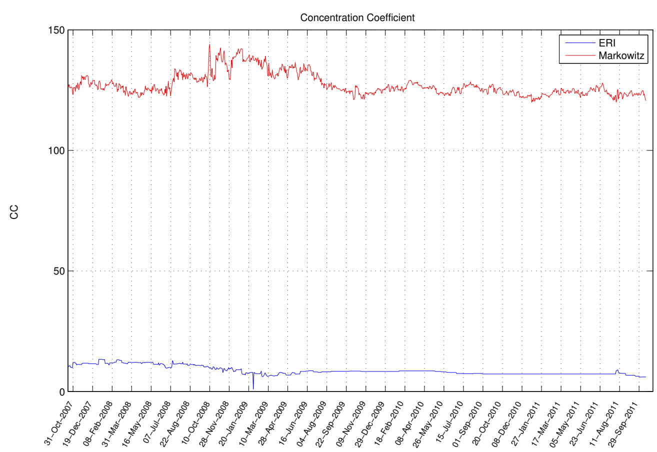


Figure 7: Concentration Coefficient in the backtesting experiment on all S&P 500 stocks. Total set of stocks with 10% threshold alpha estimation.

This impression is confirmed by Figure 8. The dynamics of the Markowitz Portfolio in Figure 9 is similar, but the difference between the crisis and recovery period is somewhat weaker. All in all it seems that the Markowitz portfolio undergoes many small changes, whereas the changes in the ERI optimal portfolio are less but much stronger.

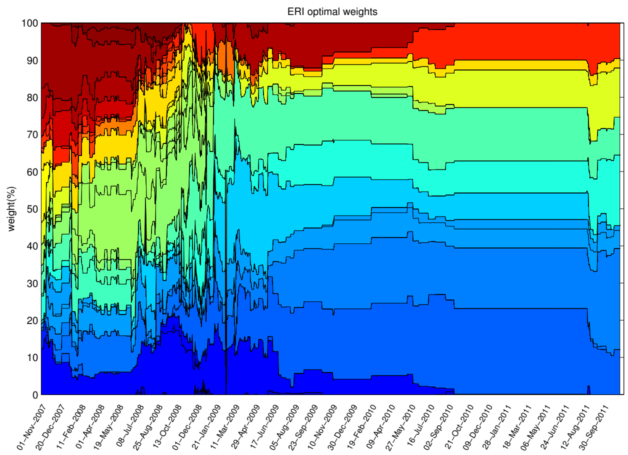


Figure 8: ERI optimal weights in backtesting on all S&P 500 stocks

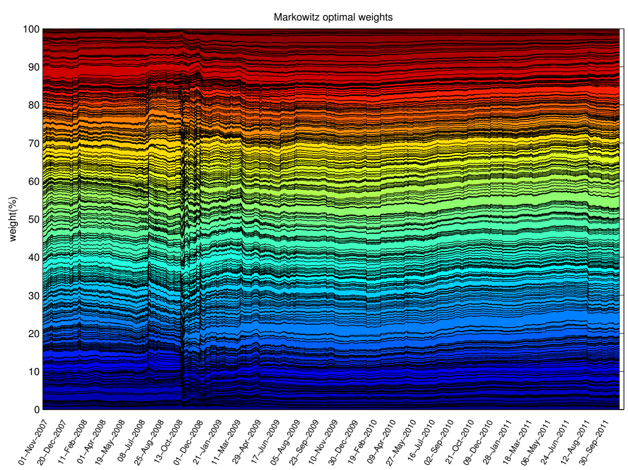


Figure 9: Weights of the Markowitz portfolio in backtesting on all S&P 500 stocks

Turnover

The impression about stronger changes in the ERI portfolio accords with the findings on the average portfolio turnover in Sections 4.1 and 4.2. The development of the turnover coefficient over time is shown in Figure 10. The larger the spikes in the turnover pattern, the greater the instantaneous portfolio shift. The difference between the ERI minimization and the Markowitz portfolio in the crisis period is remarkable. The turnover pattern of the Markowitz portfolio points to a lot of small portfolio changes that lead to permanent, but moderate trading activity. The pattern of the ERI based portfolio has a lower level of basic activity, but much greater spikes corresponding to large portfolio shifts. Thus, if carried out immediately, the restructuring of the ERI optimal portfolio requires more liquidity in the market. This disadvantageous feature can be tempered by splitting the transactions and distributing them over time. The tradeoff between fast reaction to events in the market and liquidity constraints is an interesting topic for further research.

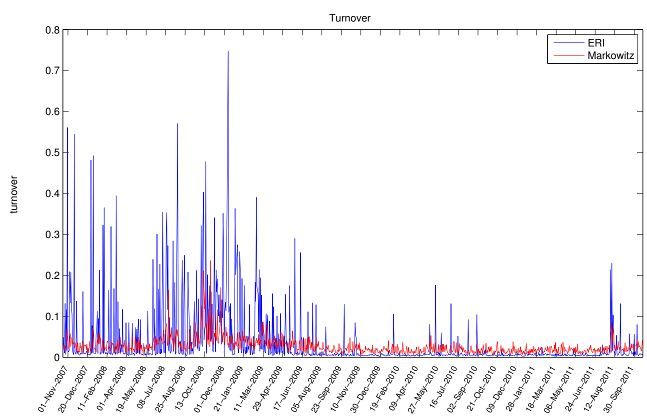


Figure 10: Portfolio turnover

Diversification measured by PCA

The development of the first PCA factor over time is shown in Figure 11. The amount of portfolio variance that can be explained by the first PCA factor increases in the months after the default of Lehman Brothers to a new level. This shows that the recent financial crisis has changed the perception

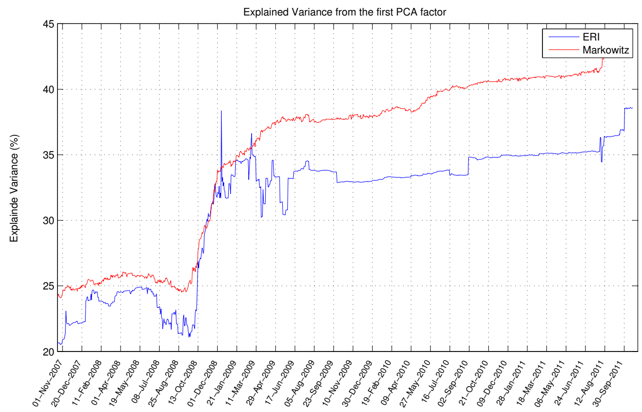


Figure 11: Variance explained by the first PCA factor

of dependence in the market and thus increased the dependence between the stocks. Ranging below 25% before the crisis, the first PCA factors of both strategies are typically above 35% afterwards. This chart indicates a change in the intrinsic market dynamics. The stronger co-movements of S&P 500 stocks reflect the new perception of systemic risk. As a consequence, the diversification potential in the after-crisis period is lower than in the time before the crisis.

Most of the time, the first PCA factor of the ERI optimal portfolio ranges somewhat below that of the Markowitz portfolio. Thus we can conclude that ERI optimization brings more diversity into the portfolio than the mean-variance approach. To the use of PCA: There is one exception to this rule: in February 2009, the first PCA factor of the ERI optimal portfolio peaks out to 100%. It corresponds to a single day when the ERI strategy selects only one stock for the investment portfolio. On this remarkable day, the first PCA factor is obviously identical with the investment portfolio. Recalculations let to slightly different weights but to almost identical portfolio returns. Results of this kind can be avoided in practice by appropriate bounds on portfolio restructuring.

Tail index estimates

In addition to the backtesting studies where the tail index α is estimated for the radial part of the random vector X, we also estimated α for each

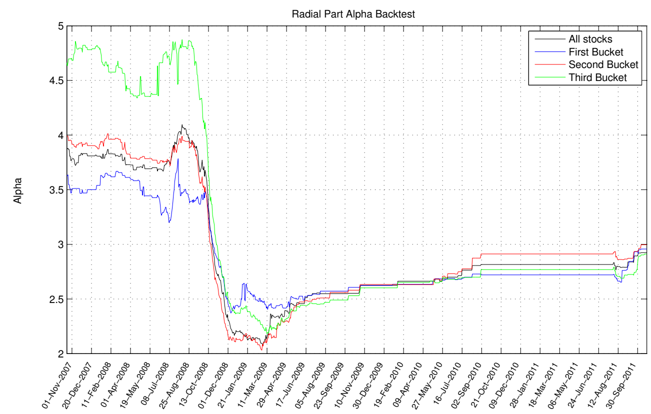


Figure 12: Estimated tail index α for the radial part of S&P 500 stocks and three subgroups

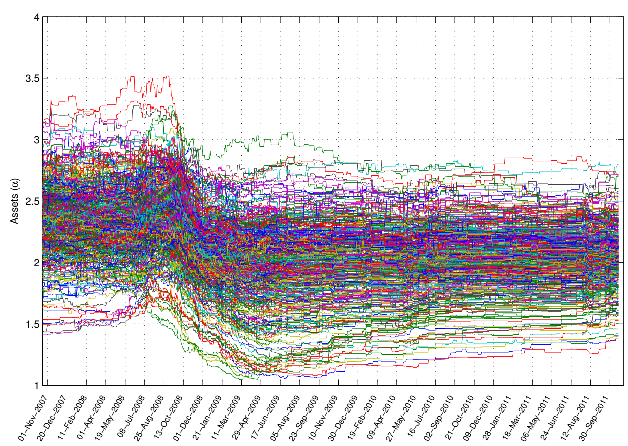


Figure 13: Estimated tail index α for the S&P 500 stocks

stock separately. The estimated values of α for the radial parts are shown in Figure 12, and the development of α estimates for single stocks is shown in Figure 13. Beginning in summer 2008, there is a common downside trend for all stocks, i.e. all return tails become heavier in the crisis time. This trend stops in spring 2009. The missing recovery since then can be explained by the width of the estimation window. Based on the foregoing 1500 trading days, our estimators remain influenced by the crisis for 6 years. This effect is visibile in both figures. In addition to that, Figure 12 shows that after the crisis the estimated values of α in all three sub-groups are very close to each other and even change their ordering compared to the pre-crisis period: the group with lowest α before crisis does not give the lowest α after the crisis. These effects may be explained by the strong influence of extremal events during the crisis on the estimates in the after-crisis period. As the historical observation window includes 1500 days, the crisis events do not disappear from this window until the end of the backtesting period. It seems that the estimated values of α tend to ignore the recovery of the stocks in the after-crisis period. This may be one more explanation to the different performance of the ERI strategy in the different stock groups. This effect can be tempered by downweighting the observations in the historical window when they move away from the present time. The choice of this weighting rule goes beyond the scope of this paper and should be studied separately.

5 Conclusions

Our backtesting results suggest that the Extreme Risk Index (ERI) could be useful in practice. Comparing very basic implementations of the ERI methodology and the Markowitz approach, we obtained very promising results for stocks with heavy tails. Tailored to such assets, the ERI optimal portfolio not only outperforms Markowitz, but it also yields an annualized return of 11.5% over 4 years including the crisis. This should outweigh the higher transaction costs caused by the ERI based approach. Thus, taking into account the special nature of diversifications for heavy-tailed asset returns, the ERI strategy may increase the reward for the corresponding risks.

Our study also shows that the Markowitz approach can catch up with ERI optimization in some cases, especially when applied to stocks with lighter tails. Therefore a combined algorithm switching between Markowitz and ERI depending on the volatility of the assets may also be a good choice. Other improvements of the ERI methodology may be achieved by downweighting the crisis events when they reach the far end of the historical observation window and by smoothing the pattern of trading activities. All these questions should be subject to further research.

Acknowledgement

The authors thank Svetlozar T. Rachev who suggested to undertake a backtesting study of the ERI methodology and for his help to organize this study.

References

[1] A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, 1980.

[2] L. Chollete, V. de la Peña, and C.-C. Lu. International diversification: An extreme value approach. *Journal of Banking & Finance*, 36(3):871–885, 2012.

[3] J. Daníelsson, L. de Haan, L. Peng, and C. G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248, 2001.

[4] F. Desmoulins-Lebeault and K.-R. Cécile. Non-gaussian diversification: When size matters. *Journal of Banking & Finance*, 36(7):1987–1996, 2012.

[5] F. J. DiTraglia and J. R. Gerlach. Portfolio selection: An extreme value approach. *Journal of Banking & Finance*, 37(2):305–323, 2013.

[6] T. Doganoglu, C. Hartz, and S. Mittnik. Portfolio optimization when risk factors are conditionally varying and heavy tailed. *Computational Economics*, 29(3-4):333–354, 2007.

[7] H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75(2):149–172, 1998.

[8] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*. Springer, 1997.

[9] X. D. He and X. Y. Zhou. Portfolio choice via quantiles. *Mathematical Finance*, 21(2):203–231, 2011.

[10] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

[11] W. Hu and A. N. Kercheval. Portfolio optimization for student t and skewed t returns. *Quantitative Finance*, 10(1):91–105, 2010.

[12] H. Hult and F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34 (3):587–608, 2002.

[13] N. Hyung and C. G. de Vries. Portfolio selection with heavy tails. *Journal of Empirical Finance*, 14(3):383–400, 2007.

[14] P. Jorion. Bayesian and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking & Finance*, 15(3):717–727, 1991.

[15] G. Mainik. *On Asymptotic Diversification Effects for Heavy-Tailed Risks*. PhD thesis, University of Freiburg, 2010. URL `http://www.freidok.uni-freiburg.de/volltexte/7510/pdf/thesis.pdf`.

[16] G. Mainik. Estimating asymptotic dependence functionals in multivariate regularly varying models. *Lithuanian Mathematical Journal*, 52(3): 259–281, 2012.

[17] G. Mainik and P. Embrechts. Diversification in heavy-tailed portfolios: properties and pitfalls. *Annals of Actuarial Science*, 7:26–45, 2013.

[18] G. Mainik and L. Rüschendorf. On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, 14:593–623, 2010.

[19] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[20] T. Nguyen and G. Samorodnitsky. Tail inference: where does the tail begin? *Extremes*, 15(4):437–461, 2012.

[21] S. Ortobelli, S. T. Rachev, and F. J. Fabozzi. Risk management and dynamic portfolio selection with stable Paretian distributions. *Journal of Empirical Finance*, 17(2):195–211, 2010.

[22] M. Potters, J.-P. Bouchaud, and L. Laloux. Financial Applications of Random Matrix Theory: Old Laces and New Pieces. *Preprint*, 2005. URL `http://arxiv.org/abs/physics/0507111`.

[23] S. T. Rachev, C. Menn, and F. J. Fabozzi. *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Wiley, 2005.

[24] S. I. Resnick. *Heavy-Tail Phenomena*. Springer, 2007.

[25] S. I. Resnick and C. Stărică. Smoothing the Hill estimator. *Advances in Applied Probability*, 29(1):271–293, 1997.

[26] C. Zhou. Dependence structure of risk factors and diversification effects. *Insurance: Mathematics and Economics*, 46(3):531–540, 2010.