Foster–Hart optimal portfolios ☆

Abhinav Anand [a], Tiantian Li [b], Tetsuo Kurosaki [c], Young Shin Kim [d],∗

[a] Financial Mathematics and Computation Cluster, Michael Smurfit Graduate Business School, University College Dublin, Ireland
[b] Department of Applied Mathematics and Statistics, Stony Brook University, USA
[c] Bank of Japan, Japan
[d] College of Business, Stony Brook University, Stony Brook, NY 11794-3775, USA

A R T I C L E   I N F O

A B S T R A C T

We reinvestigate the classic portfolio optimization problem where the notion of portfolio risk is captured by the "Foster–Hart risk"—a new, bankruptcy-proof, reserve based measure of risk, extremely sensitive to left tail events (Foster and Hart, 2009). To include financial market frictions induced by market microstructure, we employ a general, ex-ante transaction cost function with fixed, linear and quadratic penalty terms in the objective function. We represent the US equity market by the Dow Jones Industrial Average (DJIA) index and study the performance of the Foster–Hart optimal DJIA portfolio. In order to capture the skewed and leptokurtotic nature of real life stock returns, we model the returns of the DJIA constituents as an ARMA–GARCH process with multivariate "normal tempered stable" innovations. We demonstrate that the Foster–Hart optimal portfolio's performance is superior to those obtained under several techniques currently in use in academia and industry.

1. Introduction

The tradeoff between risk and return is one of the most central aspects of research in financial economics. While economic agents prefer high returns (all else being equal) they display high levels of risk aversion as well. From the point of view of an investor, this problem can be expressed as a constrained optimization program.

The mean–variance framework established in Markowitz (1952) is the bedrock of modern portfolio theory. It diverted the attention of investment analysis away from individual security selection towards the concept of portfolio diversification and emphasized the risk return tradeoff that modern academics and practitioners take for granted today.

However, most researchers now agree that the mean–variance framework is insufficient to capture the aforementioned risk return tradeoff. Objections have been raised regarding the inadequacy of variance (or equivalently the standard deviation) as a proxy for portfolio risk (acknowledged as early as in Markowitz (1959)) as well as the assumption of Gaussianity of asset returns (Mandelbrot, 1963a,b). In order to accommodate such concerns, the current approach relies on 'mean risk analysis', in which a suitable definition of portfolio risk is utilized, along with a more realistic portfolio return distribution.

It is in this tradition that we propose to measure portfolio risk by the Foster–Hart risk—a new measure of risk proposed in Foster and Hart (2009) which computes the minimal level of wealth an agent must possess in order to render an arbitrary

sequence of future investments bankruptcy-proof. The Foster–Hart risk depends only on the distribution of the payoffs; on the assumption that the agent prefers non-bankruptcy to bankruptcy; and does not impose any probabilistic structure on the future sequence of possible investments (or 'gambles' as referred to in their paper). Indeed, it is a highly useful way to think about financial risk and boasts of several highly desirable properties that experts in the field deem important. It also has a very satisfying interpretation as a "minimal reserve" which renders a risky investment 'safe' to undertake.[1]

On the related front of the statistical modeling of asset returns, we choose to employ the multivariate "normal tempered stable" (NTS) distribution to capture the skewed and leptokurtotic nature of asset returns. The family of tempered stable distributions is derived from the older, more well known Stable distribution family by 'tempering' the thickness of tails far from the center of the distribution. Such tempered stable families have been used to model statistical features of asset returns with high accuracy in a recent sequence of papers.[2]

We include practitioners' concerns about the costs incurred in frequent rebalancing of portfolios by incorporating an ex-ante, general transaction cost function with fixed, linear and quadratic penalty terms in the objective function. This ensures that optimal allocation of portfolio weights will accommodate concerns about portfolio rebalancing costs.

For our empirical study, we investigate the Dow Jones Industrial Average portfolio and use it to represent the US equity market. We set up an optimization program which maximizes risk adjusted returns and includes the transaction cost function referred to above. In order to account for temporal dependence among asset returns, we model DJIA constituents' return process as an ARMA–GARCH stochastic process with multivariate normal tempered stable innovations.

In order to establish that our union of Foster–Hart risk modeling with fat-tailed statistical modeling produces optimal portfolios with superior performances than those obtained by current techniques, we compare optimal portfolio performances obtained under different definitions of risk (such as standard deviation, Value at Risk and Average Value at Risk), as well as under different statistical models of assets returns (such as those with multivariate normal and multivariate Student's T distributions respectively).

By a variety of statistical techniques including in-sample analysis, out-of-sample analysis and backtesting, we demonstrate that our statistical model fits DJIA return data more accurately than its competitors. This justifies our use of the ARMA–GARCH–NTS model to compare portfolio performances under different risk measures, as well as different (out-of-sample) criteria of portfolio performance measures. Finally, we show that our results retain their characteristics even as risk appetites of investors vary.

This paper is organized as follows. In Section 2, we discuss some preliminary theory of risk modeling and statistical modeling, Section 3 describes our dataset, Section 4 discusses the estimation methodology while Section 5 outlines the details of the scope of our empirical study. Section 6 presents and analyzes results while Section 7 offers our concluding thoughts. At the end, we describe additional results and tables in the appendices.

---

[1] As may be easily observed, a trivial solution for bankruptcy-proofing future sequences of investments is to not undertake any such investment! Or equivalently, by imposing the minimal wealth required to undertake an investment to be infinite, one can trivially ensure non-bankruptcy—no matter what the future sequence of payoff distributions might be. However, the true beauty of the result is to show that agents could have a finite level of wealth, as computed by the Foster–Hart risk and still be almost surely bankruptcy-proof in the face of arbitrary sequences of payoff distributions.

[2] See in particular Kim et al. (2010), Bianchi et al. (2010), Kim et al. (2011), Kim et al. (2012) and Choi et al. (2015) as prominent examples.

2. Preliminaries

In this section we revisit some important concepts that form the basis of our study in portfolio optimization.

In the now classic Markowitz (1952), variance (or equivalently, the standard deviation) of portfolio returns was chosen to reflect the risk afflicting a portfolio.

For a portfolio of n assets with weights $w_1, \ldots, w_n$ the mean–variance program is:

$$\max(w^\top \mu) \tag{1}$$

subject to

$$w^\top \Omega w \leqslant \sigma_*^2$$
$$1^\top w = 1$$
$$w \geqslant 0^\top \tag{2}$$

where $w = (w_1, \ldots, w_n)^\top$ is the vector of asset weights, $\mu = (\mu_1, \ldots, \mu_n)^\top$ is the vector of asset means, Ω is the $n \times n$ covariance matrix of asset returns comprising the portfolio, σ_*^2 is a user specified maximal variance and $1^\top$ is the n dimensional unit vector $(1, \ldots, 1)$.

The last two constraints rule out short selling and in essence, state that the feasible set of weights is in the $n - 1$ dimensional standard simplex: $w \in \Delta^{n-1}$. Thus, we can rewrite the above program as:[3]

$$\max_{w \in \Delta^{n-1}} (w^\top \mu - C w^\top \Omega w) \tag{3}$$

or equivalently,

$$\min_{w \in \Delta^{n-1}} (C w^\top \Omega w - w^\top \mu) \tag{4}$$

where C may be interpreted to be a risk aversion parameter. A very low value of C indicates lower aversion to risk thereby prompting the investor to concentrate portfolio weights on assets with the highest expected returns, while high values of C emphasize lower variance at the expense of high expected returns.

2.1. Mean risk analysis

While mean–variance analysis was the first attempt to model the tradeoff between risk and return in a systematic way, researchers now have generalized the above setup so as to accommodate more accurate notions of portfolio risk.

If $\rho(\cdot)$ is a measure of portfolio risk, the minimization program becomes:

$$\min_{w \in \Delta^{n-1}} (C \rho(w) - w^\top \mu) \tag{5}$$

Different notions of portfolio risk will result in different allocational decisions made by the investor.

2.1.1. Risk measures
In order to compare and contrast the role played by risk measures, we analyze portfolio performance based on several prevalent measures of risk.

2.1.1.1. Foster–Hart risk. Foster and Hart (2009) propose a new risk measure that quantifies the level of wealth an agent must possess in order to render future investments in risky assets bankruptcy proof.

---

[3] The standard $n - 1$ simplex is defined as $\{ w \in [0,1]^n : w_i \geqslant 0 \text{ and } \sum_{i=1}^n w_i = 1 \}$.

Risky investments are referred to as "gambles"—*bounded* random variables with positive expectations and a positive probability of losses:

$$\mathbb{E}(g) > 0 \text{ and } \mathbb{P}(g < 0) > 0$$

The main result of Foster and Hart (2009) is that for each such gamble g, there is a scalar $R(g)$ which is the unique positive solution to the following equation:

$$\mathbb{E}\left(\log\left[1 + \frac{g}{R(g)}\right]\right) = 0 \tag{6}$$

This scalar $R(g)$—the Foster–Hart risk measure (*FH risk* in short)—is interpretable as the *minimal reserve* needed to undertake the risky investment g which renders the agent bankruptcy-proof.

In addition, FH risk exhibits other properties deemed desirable in a risk measure. Some of them include (positive) homogeneity, subadditivity, consistency with first and second order stochastic domination, continuity etc. Also, the FH risk is always weakly greater in magnitude than the maximal loss possible under that gamble. For a detailed discussion, see Foster and Hart (2009). We note that FH risk does not qualify as a "coherent" risk measure in the sense of Artzner et al. (1998) since it does not obey the axiom of cash invariance. Due to the same reason, it does not qualify as a convex risk measure either.[4] Riedel and Hellmann (2015a) extend FH risk to dynamic settings and show that it is time-inconsistent as well.

2.1.1.2. General Foster–Hart risk. Riedel and Hellmann (2015b) show that the defining equation for FH risk has no solution for many common continuous distributions. They, however, extend FH risk to such continuous random variables consistently by identifying it with the *maximal loss L* of the gamble. Since in the Foster–Hart setup, gambles are bounded, such maximal losses are finite.

$$R(g) = \begin{cases} r^* \text{ such that } \mathbb{E}\left(\log\left[1 + \frac{g}{r^*}\right]\right) = 0 & \text{if } \mathbb{E}\left(\log\left[1 + \frac{g}{L}\right]\right) < 0 \\ L & \text{if } \mathbb{E}\left(\log\left[1 + \frac{g}{L}\right]\right) \geqslant 0 \end{cases} \tag{7}$$

2.1.1.3. The empirical Foster–Hart risk heuristic. In order for the Foster–Hart risk to be defined, the gamble under consideration must be a random variable whose expectation is positive ($X : \mathbb{E}(X) > 0$). However, during severe economic downturns, portfolios may exhibit mean *negative* returns, thereby making the Foster–Hart risk undefined.

Hence we define the notion of the *empirical Foster–Hart risk heuristic* in which, just as above, we identify the Foster–Hart risk with the maximum loss L for scenarios in which the gamble's mean is negative.

While extending FH risk to gambles with negative means is an important problem, the said extension is nontrivial. Indeed, according to the Foster–Hart framework, the FH risk of such gambles is infinite since in the long run, the only way to avoid bankruptcy when faced with gambles with negative means is to reject them, as playing them will result in bankruptcy almost surely. This may also be observed by applying the axiomatic derivation of FH risk by preference order approach (Hart, 2011), where gambles with negative means are always rejected by risk averse agents.

---

[4] Convex risk measures combine the axioms of positive homogeneity and subadditivity into a convexity axiom. Hence all coherent measures are convex but not vice versa. FH risk obeys all other axioms of coherence—consistency with first order stochastic domination, positive homogeneity and subadditivity—but the applicability of the axiom of cash invariance in the bankruptcy-proof Foster–Hart setup is questionable (Foster and Hart, 2009, p. 801).

Therefore, in order to render FH risk practically useful in the context of portfolio optimization, a heuristic solution is required. While substituting the maximal loss for such gambles is admittedly arbitrary, it does preserve the highly important left tail sensitivity of the Foster–Hart risk, which we elaborate upon in Section 6.3 and is key to understanding the superiority of FH risk for portfolio optimization. Moreover, for our sample, such cases occupy a small proportion of scenarios (about 4%) and hence we believe that our heuristic does not distort the optimization program much.

We do note that imposition of such a heuristic makes the empirical FH risk discontinuous. However, the FH risk is continuous only when the said gambles' maximal losses have the same limit L (i.e., $L(g_n) \to L(g)$) (Foster and Hart, 2009, p. 795). In highly volatile markets, the FH risk itself may be discontinuous as there is wide variation in daily maximal losses with returns responding dramatically to signals such as news events, interest rate announcements etc. As above, finding a suitable empirical heuristic that respects continuity is nontrivial.

As a final comment we add that ours is not the only feasible heuristic. Indeed, any value $x : L < x < \infty$ can be replaced in lieu of L while respecting FH risk's left tail sensitivity. Our choice of the maximal loss as the empirical heuristic is guided by its simplicity, consistency with Riedel and Hellmann (2015b); and by being the smallest value consistent with FH risk's left tail sensitivity property (see Section 6.3 for details).

2.1.1.4. Value at Risk. The Value at Risk (more commonly, *VaR*) is simply a user specified quantile of the return distribution and is the most widely used risk measure among practitioners today—primarily due to its ease in computation and simple interpretation.

2.1.1.5. Average Value at Risk. As the name suggests, Average Value at Risk (*AVaR*, also referred to as *Expected Shortfall (ES)*) averages the different VaR values beyond the user specified confidence level.

$$\text{AVaR}_\alpha(g) := \frac{1}{\alpha}\int_0^\alpha \text{VaR}_p(g)dp \tag{8}$$

AVaR is more informative about the losses beyond the stipulated confidence level and, unlike VaR and FH, satisfies the axioms of coherence.

2.1.2. Transaction costs

Although mean risk analysis is an improvement over mean variance analysis, it needs to be modified further before it can prove useful to practitioners. This is so because notwithstanding the static nature of the optimization program described above (note the absence of time subscripts t), reallocation of portfolio weights at each time period is a cost that partially offsets cumulative portfolio performance. In order to deal with this problem systematically, we discuss the issue of transaction costs.

Pogue (1970) was one of the first studies to incorporate transaction costs into the mean variance framework. In order to accommodate transaction costs into the mean risk framework, our approach essentially amounts to including a penalty function in the objective:

$$\min_{w \in \Delta^{n-1}} \left(C\rho(w) - w^\top \mu + \lambda \cdot TC(w)\right) \tag{9}$$

where $TC(\cdot)$ denote the transaction cost function and λ denotes the transaction cost aversion parameter.

Neglecting the transaction cost term ignores financial market frictions and can lead to consistent overestimation of portfolio performance. Their introduction however, reflects the investor's concern about maximizing returns but keeping the cost of doing so at a minimum. Consequently, for higher values of transaction cost

aversion, lower levels of rebalancing are more suitable and conversely.

We propose a general quadratic transaction cost function and note that they have been commonly considered in literature (Fabozzi et al., 2006, p. 76):

$$\lambda \cdot TC(w_t) = \lambda \left(\alpha + \beta \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}| + \gamma \sum_{i=1}^{n} (w_{i,t} - w_{i,t-1})^2 \right) \quad (10)$$

Hence the minimization program at time t now becomes:

$$\min_{w_t \in \Delta^{n-1}} \left(C\rho(w_t) + \lambda TC(w_t) - w_t^\top \mu \right) \quad (11)$$

or, in expanded form:

$$\min_{w_t \in \Delta^{n-1}} \left(C\rho(w_t) + \lambda \left(\alpha + \beta \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}| + \gamma \sum_{i=1}^{n} (w_{i,t} - w_{i,t-1})^2 \right) - w_t^\top \mu \right) \quad (12)$$

2.2. Statistical modeling of asset returns

While the use of a risk measure improves allocational efficiency, it comes at a computational cost. In order to compute risk of a portfolio, one must know the *entire* distribution according to which portfolio returns are governed. This in turn necessitates that the entire joint multivariate distribution of individual assets be known.[5]

Classically, portfolio returns were assumed to be Normally distributed but research done since the '60s suggests that while the Normal distribution may be a satisfactory assumption for lower frequency return data (say monthly or quarterly data), it provides terrible fits in general for higher frequency return data (say, weekly or daily) (Cont, 2001). Researchers now agree that return distributions of various financial instruments are *skewed and leptokurtotic*—asymmetric, and more peaked around the mean, with fat, Pareto-type tails.[6]

However, the lack of existence of moments presents difficulties in dealing with general, Stable distribution based approaches. *Tempered* Stable distributions (see Barndorff-Nielsen and Shephard (2001) and Barndorff-Nielsen and Levendorskii (2001)) present a possible solution and have been used successfully for modeling returns (Kim et al., 2011). Recent findings from Kim et al. (2012) motivated us to choose a specific distribution—the "*multivariate Normal* Tempered Stable" distribution—to model the asymmetric and interdependent nature of stock returns.

In order to illustrate the role played by the multivariate NTS distribution in modeling asset returns, we compare and contrast it with the more standard assumption of asset returns being distributed as multivariate Normal and multivariate Student's T distributions respectively.

2.2.1. The normal tempered stable distribution

The Normal Tempered Stable (NTS) distribution is built upon the Tempered Stable subordinator with parameters $\alpha \in (0, 2)$ and $\theta > 0$ with characteristic function:

$$\phi_T(u) = \exp \left(-\frac{2\theta^{1-\frac{\alpha}{2}}}{\alpha} ((\theta - iu)^{\frac{\alpha}{2}} - \theta^{\frac{\alpha}{2}}) \right) \quad (13)$$

The n-dimensional NTS distributed random vector $X = (X_1, \ldots, X_n)^\top$ is defined to be:

$$X = \mu + \beta(T - 1) + \sqrt{T}(\gamma \circ \xi) \quad (14)$$

where $\mu, \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}^n_+, \xi \sim \mathcal{N}(0, \Sigma), (\gamma \circ \xi) = (\gamma_1 \xi_1, \ldots, \gamma_n \xi_n)^\top$; and T is the Tempered Stable subordinator with parameters α and θ. The subordinator is independent of ξ. The n-dimensional NTS random vector so constructed, is denoted as $X \sim NTS_n(\alpha, \theta, \beta, \gamma, \mu, \Sigma)$.

Standardization of the NTS random vector involves setting $\mu = (0, \cdots, 0)^\top, \gamma_i = \sqrt{1 - \beta_i^2 \left(\frac{2-\alpha}{2\theta} \right)}$ with $|\beta_i| < \sqrt{\frac{2\theta}{2-\alpha}}$, where β_i and γ_i are the i-th elements of β and γ, respectively, for $i = 1, 2, \cdots, n$. This yields a standard NTS random vector with 0 mean and unit variance and is denoted as $X \sim stdNTS_n(\alpha, \theta, \beta, \mathbb{I}_n)$, where $\mathbb{I}_n$ is the $n \times n$ identity matrix.[7]

2.2.2. Modeling temporal dependencies

Using a suitable leptokurtotic distribution can help us capture some stylized facts of asset returns. However, we must also account for autocorrelation and volatility clustering (Cont, 2001). Hence we introduce a stochastic process for the evolution of returns that is consistent with such facts.

A common assumption is the use of ARMA(1,1)–GARCH(1,1) stochastic process:[8]

$$r_{t+1} = \mu_{t+1} + \sigma_{t+1} \epsilon_{t+1} \quad (15)$$

where r_{t+1} denotes asset *log* return at time $t + 1, \mu_{t+1}$ is the conditional mean, while σ_{t+1} is the conditional standard deviation. ϵ_{t+1} is the "*innovation*" and is a standard random variable with 0 mean and unit standard deviation.

The conditional mean and the conditional standard deviation themselves are modeled as autoregressive processes, giving rise to the nomenclature of ARMA (Autoregressive Moving Average) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity).

The ARMA(1,1) process for the conditional mean is:

$$\mu_{t+1} = c + ar_t + b\sigma_t \epsilon_t \quad (16)$$

where $c, b \in \mathbb{R}$ and $|a| \leqslant 1$

The GARCH(1,1) process for the conditional standard deviation is:

$$\sigma_{t+1}^2 = d + f(\sigma_t \epsilon_t)^2 + g\sigma_t^2 \quad (17)$$

where $d > 0$ and $f, g \geqslant 0$. We assume the innovations to be standard NTS.

As remarked before, the FH risk has been extended to a general, dynamic framework in Riedel and Hellmann (2015a) and our imposition of an ARMA–GARCH stochastic process for modeling the dynamic return process as described above is consistent with their formulation.

3. Data

Our data comprise daily log returns from stocks included in the Dow Jones Industrial Average (DJIA) index for a period encompassing January 3rd, 2000 to December 31st, 2014. The Dow Jones Industrial Average has historically consisted of 30 large, publicly owned firms based in the United States. The list of DJIA constituents as it stood in December 2015 is included as Table 7 in the appendices to this paper.

---

[5] Choosing standard deviation as a risk measure only requires the knowledge of portfolio standard deviation, which in turn merely requires computation of the asset returns' covariance matrix, as opposed to their entire joint distribution.

[6] There is a substantially large body of literature (see in particular Mandelbrot (1963a,b, 1967), Fama, 1963, 1965 etc.) that provides extensive evidence for this assertion. For a comprehensive literature review, see Haas and Pigorsch (2009).

[7] We note that the sum of standard NTS random variables with common α and θ parameters is again an NTS random variable (Kim et al., 2012)—a fact that we use in portfolio analysis with NTS innovations in later sections.

[8] See in particular, Cont (2001), Berkowitz and O'Brien (2002), Kim et al. (2011) and Pagan (1996) among others.

We use 29 of the 30 listed companies for statistical analysis. The one company that we drop from this list is Visa Inc. listed on the New York Stock Exchange. The reason we do not use daily returns for this firm is because it was made public only in March 2008 and was included as a DJIA component as late as in September 2013.

We exclude American non-business days from the data set, which gives us 3773 return observations for each DJIA component.

4. Estimation methodology

We model log returns of DJIA constituents as an ARMA(1,1)–GARCH(1,1) stochastic process:

$$r_{t+1}^i = \mu_{t+1}^i + \sigma_{t+1}^i \epsilon_{t+1}^i$$
$$\mu_{t+1}^i = c_i + a_i r_t^i + b_i \sigma_t^i \epsilon_t^i$$
$$(\sigma_{t+1}^i)^2 = d_i + f_i (\sigma_t^i)^2 + g_i (\sigma_t^i \epsilon_t^i)^2 \tag{18}$$

The indices $i \in \{1, 2, \ldots, 29\}$ represent individual DJIA constituents. $\epsilon_t = (\epsilon_t^1, \epsilon_t^2, \ldots, \epsilon_t^{29})^\top$ is the multivariate innovation. We choose the 29-dimensional NTS distribution to model the joint distribution of innovations since it captures the skewed, interdependent and leptokurtotic nature of real life stock returns.

As remarked before, in order to contrast the NTS hypothesis's contribution to statistical modeling, we alternately model innovations as multivariate Normal and multivariate Student's T respectively as well. The degrees of freedom parameter for T innovations is estimated from fitting a univariate ARMA–GARCH-T (AGT) model to the DJIA index.

We employ the rolling window estimation technique with a 1250 days (about five years) forward moving time window for daily estimation of the parameters of the stochastic processes.[9] Our first time window starts from January 3, 2000 and ends on December 22, 2004; while the last time window starts from January 14, 2010 and ends on December 31, 2014. Hence we carry out 2524 daily estimations of model parameters.

Computation of the covariance matrix Σ for the multivariate NTS, Student's T and Normal innovations is done on the basis of the most recent 250 days, or almost one year of daily returns. We explicitly model the dependence between various constituents' innovations by means of their covariance matrix since movements in the stock price of one may have effects on that of the others.

Estimating the ARMA–GARCH-Normal (AGNormal) and the AGT model is done by the standard Maximum Likelihood Estimation (MLE) technique. More care needs to be taken, however, for estimating the parameters of the ARMA–GARCH–NTS (AGNTS) model. Our estimation methodology is adapted from Kim et al. (2011)—among the first studies to incorporate tempered stable innovations in an ARMA–GARCH framework—as well as from Kim et al. (2012).

The NTS distribution has two tail parameters α and θ and one skewness parameter β. We assume common tail parameters α and θ for individual constituents and estimate them from the DJIA index. This leaves the skewness parameters $\{\beta_i\}_{i=1}^{29}$ to be calibrated by MLE for each asset in the DJIA portfolio. Finally, we follow Kim et al. (2012) in joining NTS marginals by means of their covariance matrix into multivariate NTS.

5. Empirical study

The main goal of our empirical study is to establish that among the three candidate stochastic processes for modeling the asset returns of the DJIA portfolio, AGNTS captures their stylized statistical behavior best.

In order to demonstrate this, we divide our empirical study into two main parts:

1. Tests comparing the performance of the statistical model: AGNTS versus AGT versus AGNormal.
2. Backtesting risk estimates computed from the above models.

5.1. Testing statistical models

We compare the performance of the three candidate statistical models—AGNormal, AGT and AGNTS—by computing their residuals and testing to verify if they conform to the said hypotheses. We do this for both in-sample and out-of-sample estimates as described below.

5.1.1. In sample test: fitting the DJIA index

We take the complete DJIA index sample from January 3, 2000 to December 31, 2014 and fit the three candidate models to it. We compute residuals for each statistical model and compare their empirical density and empirical quantiles to those hypothesized by the statistical models—standard normal, standard T and standard NTS respectively.

Hence, we set up a univariate ARMA(1,1)–GARCH(1,1) stochastic process, fit the univariate DJIA index to it and compute the residuals of the univariate regression to check how different their empirical distribution and hypothesized distributions are.

For the AGNormal model, the DJIA index residuals must conform to the standard normal distribution. However, when we compare the graphs of the empirical density versus that of the standard normal in Fig. 1, we observe that the empirical density shows significant excess kurtosis and fat tails than the standard normal. The quantile–quantile (QQ) plot also displays substantially poor fits in the tails—all of which echo previous findings compiled in Cont (2001).

For the AGT model's hypothesis, the DJIA index residuals must conform to a standard T random variable with v degrees of freedom (which upon estimation via MLE, comes out to be 7.20). Hence we compare empirical residual densities and quantiles with those of a standard T with the same degrees of freedom. As the accompanying Fig. 2 shows, while the fits are better than those provided by standard normal, the empirical distribution is more kurtotic and exhibits fatter tails than even the T distribution.

When comparing empirical densities and quantiles with those of standard NTS in Fig. 3, we see that the fits are substantially better and that the NTS hypothesis for residuals captures both kurtosis and fat tails much better than its competitors.

Hence it can be claimed that the DJIA index—a stylized, one dimensional market-based representation of the American economy—is best represented as an AGNTS stochastic process.[10]

5.1.2. The Kolmogorov Smirnov (KS) test

As remarked before, for estimation of all three candidate stochastic processes, we employ the rolling window estimation technique with window length of approximately five years. Based on these estimations, we test daily residuals for all 29 constituents of the DJIA index.

In order to test how well they conform to the hypotheses of being distributed according to standard normal, standard T and standard NTS respectively, we employ the Kolmogorov–Smirnov test—a nonparametric test that compares the shapes of empirical distributions to those of a hypothesized distribution.

---

[9] The optimal length of time windows is still a matter of debate. Very long time windows cause stack overflow issues in MATLAB, while very short time windows are not able to capture the stylized facts of skewness and fat-tailedness.

[10] We qualify our claim by noting that it is based on the in-sample properties observed from 2000 to 2014; and that the competitors were the AGNormal and the AGT processes.
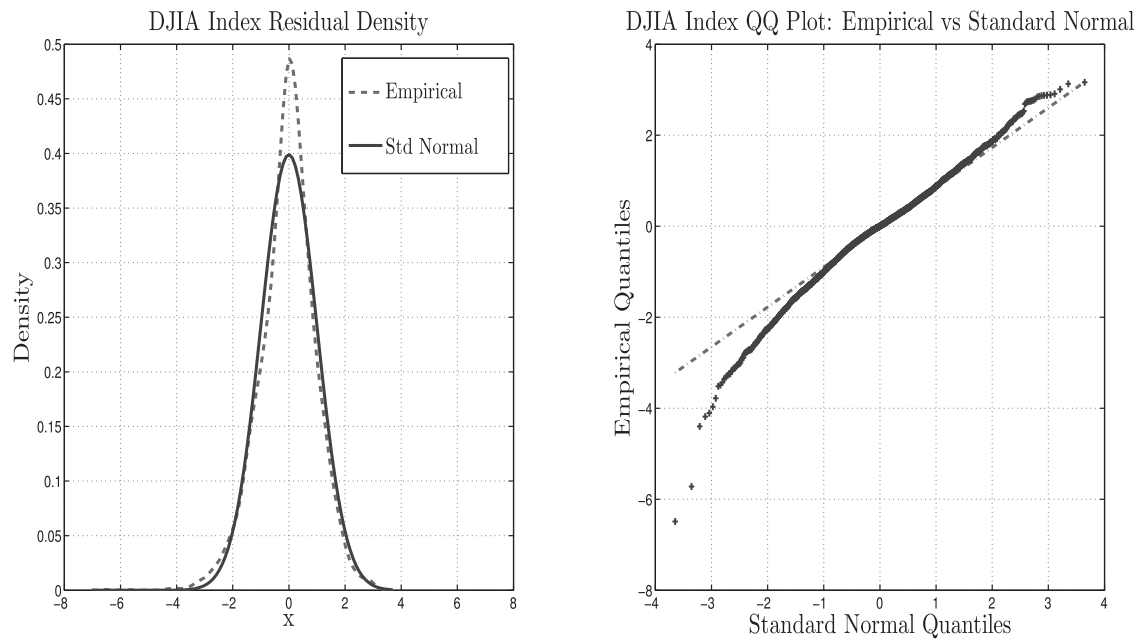
Fig. 1. Comparing the empirical distribution of the DJIA index residual to that of standard normal. The left hand graph compares the empirical density of the residuals to that of standard normal. In line with previous literature, the empirical density is much more leptokurtotic than the standard normal. The right hand graph plots empirical quantiles versus standard normal quantiles. Note that while the standard normal quantiles fit the empirical quantiles satisfactorily around the center, the fits becomes progressively worse in the tails.
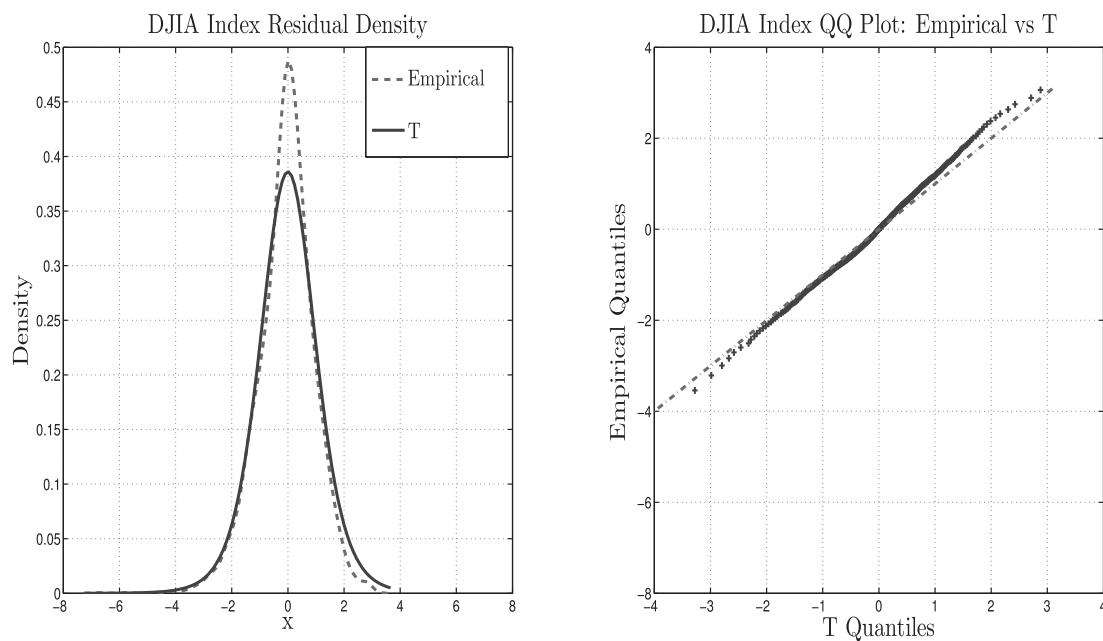
Fig. 2. Comparing the empirical distribution of the DJIA index residual to that of standard *T*. The estimated degrees of freedom for the index is 7.20. The empirical density versus the standard *T* density graph on the left side reveals better fits than standard normal. However, empirical density is still substantially more leptokurtotic than the standard $T_{7.20}$ density. While the fatter tails of the *T* distribution provide a better fit in the QQ plot, the fits are poorer in the tails.

For each of the 29 DJIA constituents, we apply the KS test daily—a total of 2524 such estimations for each asset. We report the number of days the null hypotheses are rejected at significance levels 10%, 5%, 1% and 0.5% respectively. The fewer the number of rejections, the better is the performance of the statistical model in question.[11]

As careful observation of Table 1 will confirm, the NTS hypothesis strongly outperforms its competitors. For the significance levels of 10% and 5%, for 20 of the 29 stocks, we observe the order AGNTS⪰AGT⪰AGNormal. For more conservative levels of 1 and 0.5%, the margin of outperformance increases sharply: AGNTS⪰AGT⪰AGNormal is observed for 27 out of the 29 stocks.

The stock of the chemical industry firm DuPont (ticker: DD) is the only exception to this observation. At all levels of significance, DuPont is best described by the AGNormal model than by the

---

[11] The null hypothesis being that the samples are drawn from the standard normal, standard Student's *T* and standard NTS distributions respectively.
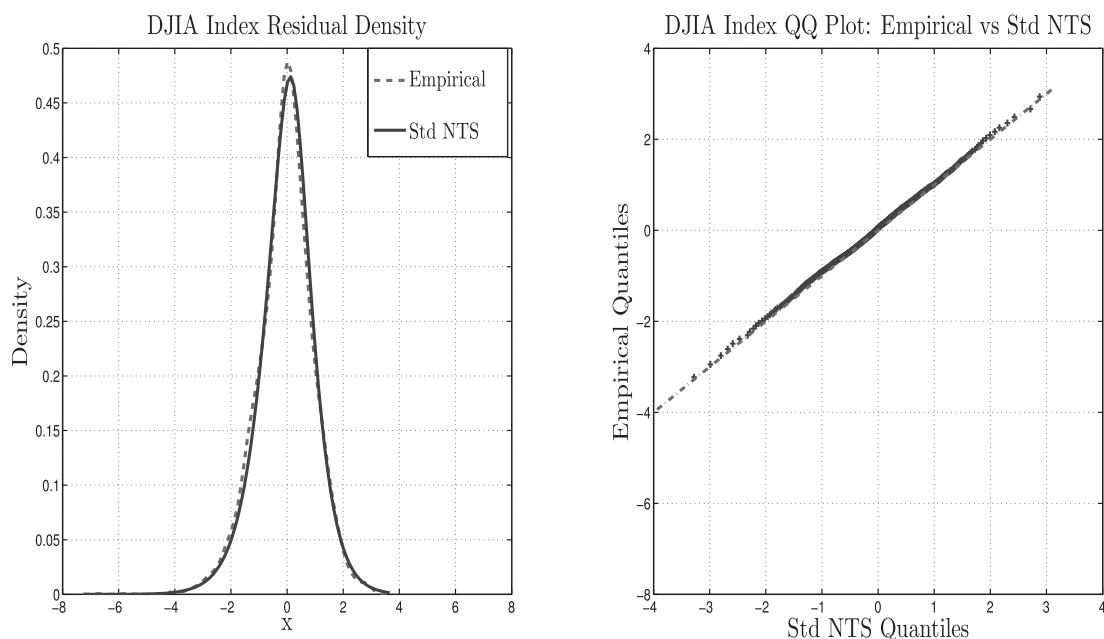
Fig. 3. Comparing the empirical distribution of the DJIA index residual to that of standard NTS. The standard NTS is able to provide good fits to DJIA index in general. The NTS density is sufficiently leptokurtotic to match the empirical density; and the QQ plot is an almost straight line—fitting the observations both in the center *and* in the tails.

Table 1
Number of rejections of distributional assumptions for each DJIA constituent on the basis of the KS test (out of a total of 2524 daily estimations). The null hypothesis is that innovations are the standard normal, standard Student's *T* and standard NTS distributions respectively. NTS strongly outperforms *T* and normal distributional assumptions at all significance levels. For significance levels of 10 and 5%, the order NTS≻*T*≻Normal is observed for 20 of the 29 stocks. For the more conservative significance levels of 1% and 0.5% respectively, the same order is observed for 27 out of the 29 stocks. The only exception to this outperformance is stock for DuPont (ticker: DD) for which AGNormal is the best performer, followed by AGT and then AGNTS.

	Significance level 10%			Significance level 5%			Significance level 1%			Significance level 0.5%		
	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS
AXP	2360	1064	373	2195	799	189	1591	65	18	1236	41	0
BA	1865	12	77	1255	10	3	275	2	1	117	2	1
CAT	2521	1240	358	2460	1231	150	1826	1121	13	1521	935	4
CSCO	2509	1479	866	2323	1172	624	1914	670	168	1775	564	58
CVX	1233	2065	1582	1175	1618	1413	1175	1010	577	1175	648	334
DD	492	1064	1039	348	895	713	74	154	164	57	34	71
DIS	2517	199	87	2440	63	5	1112	3	1	757	3	1
GE	2256	217	313	2038	116	154	1394	102	74	1263	88	70
GS	1592	20	11	1401	12	6	492	9	4	272	9	4
HD	2497	445	53	1879	186	3	1035	4	1	573	2	1
IBM	2355	587	428	2161	370	147	1149	69	1	988	29	1
INTC	1894	217	622	1218	68	219	181	10	8	152	9	4
JNJ	2450	983	322	2276	630	182	1896	484	119	1465	475	109
JPM	2215	568	283	1732	136	138	1003	10	8	705	10	6
KO	2524	1552	604	2524	1436	363	2265	1026	12	1988	863	8
MCD	1805	969	1113	1576	756	848	940	266	223	381	60	140
MMM	2520	1462	977	2389	1403	760	2220	1284	449	2184	1182	392
MRK	2524	1434	668	2377	1351	463	2067	1259	239	1974	1219	167
MSFT	2524	1600	752	2524	1506	663	2090	1154	389	1966	1053	278
NKE	2524	2514	516	2524	2445	494	2523	2291	205	2519	2196	105
PFE	2500	605	113	2285	278	3	1206	115	1	1028	95	1
PG	2524	2180	332	2524	1668	206	2524	1134	177	2497	1121	177
T	1124	223	483	261	50	175	113	8	1	59	8	1
TRV	2524	955	471	2519	622	293	2176	216	68	1783	181	25
UNH	2519	1553	387	2519	1479	385	2338	1362	381	2228	1351	350
UTX	2522	1247	489	2522	1235	292	2415	999	175	1968	845	116
VZ	942	30	277	212	6	39	4	4	1	1	4	1
WMT	1829	65	1	1778	6	1	1265	3	1	977	3	1
XOM	1053	567	1528	451	462	1116	386	394	595	383	359	333

others. Among the other stocks, ExxonMobil (XOM) and McDonald's (MCD) are also outliers in that for the former, AGNormal and AGT outperform AGNTS at all levels except 0.5%; and for the latter, AGT outperforms AGNTS for all levels except 1%.

However, for all other DJIA stocks, the AGNTS model produces better fits in general for all significance levels. The degree of outperformance increases as significance levels become more and more conservative.

Hence KS tests indicate that among the three candidate stochastic processes, AGNTS produces the best fits for almost all DJIA constituents.

5.2. Backtesting

As a further test for checking the robustness of risk estimates produced by the three statistical models, we employ backtesting by means of the Christoffersen Likelihood Ratio Test (CLR Test) introduced in Christoffersen (1998). The CLR test has three parts: the CLR unconditional coverage test ('uc'),[12] the CLR independence test ('ind'); and the CLR joint test of coverage and independence ('cc').

In order to subject risk estimates from different statistical models to systematic backtesting, we employ the following strategy:

1. We compute VaR estimates for all individual assets in the DJIA portfolio and subject them to CLR backtests.
2. We construct a benchmark, equally weighted DJIA portfolio and conduct CLR backtesting exercises on the DJIA portfolio VaR.
3. To preclude the possibility that there are errors systematically nullifying each other over time, we subdivide the entire period into five two year, non overlapping periods and backtest the equally weighted DJIA portfolio for each such period.

5.2.1. Backtesting individual assets in the DJIA portfolio

We report p values of the unconditional coverage (uc), independence (ind) and conditional coverage (cc) tests according to the AGNormal, AGT and AGNTS in Table 2 based on 2524 daily 99% VaR estimations for each asset in the DJIA index. As the careful observation of Table 2 reveals, the AGNTS model produces the best results comparatively as measured by the magnitude of p values.

Indeed, for both the unconditional and the conditional coverage test, we observe strong outperformance by AGNTS. For 19 of the 29 stocks, we observe the following order in terms of statistical model performance: AGNTS≻AGT≻AGNormal. Out of the remaining 10 components, 8 of them have AGT as the winner, while 2 stocks—Nike and Procter & Gamble—show best backtesting results under the AGNormal hypothesis.

For the independence test, all three statistical models show strong performance and report healthy p values. Indeed, just on the basis of the independence test, we cannot judge the candidates' performance since all of them perform fairly well.

Finally, looking at the last two rows of Table 2 shows that AGNTS dominates AGT and AGNormal for most of the stocks in terms of not failing the null hypothesis for significance levels 5% and 1%. For example, for the 'uc' test at the 5% level, we fail to reject the AGNTS hypothesis for 26 out of 29 stocks whereas the corresponding numbers for AGT and AGNormal are 20 and 7, out of 29. Similarly, for the 'cc' test at 5%, AGNTS wins, the corresponding numbers being 28, 25 and 7 out of 29 respectively.

5.2.2. Backtesting portfolio risk

We report p values for CLR backtests for portfolio VaR in Table 3. We report results for three thresholds of portfolio VaR estimates: 99%, 95% and 90%.

90% portfolio VaR shows AGNormal to be the winner. For both the unconditional and the conditional coverage tests, it strongly outperforms its rival processes. For the independence tests, all models produce high p values.

For both 99% and 95% portfolio VaR estimates, the AGNTS model strongly outperforms its rivals AGT and AGNormal. For all three tests: unconditional, independence and conditional coverage,

AGNTS produces uniformly higher p values than its competitors. The AGT model occupies the second spot, while the AGNormal model produces in general, the worst results. We however acknowledge that for the independence test, similar to the findings in Table 2, all three models show satisfactorily high p values and hence perform reasonably well. The margin of victory of AGNTS is due to its accurate backtesting properties in the unconditional and conditional coverage tests.

5.2.3. Backtesting multi period portfolio risks

As a final exercise in backtesting, we partition our entire period from 2004–2014 into five, equally spaced, two year periods; and subject portfolio VaR values to CLR backtests. Our first period starts from December 2004 to December 2006, second from December 2006 to December 2008 and so on.

We report p values for multi period CLR tests in Table 4. For the first time period from 2004–2006, the AGNormal reports uniformly higher p values than its competitors, while the AGNTS model takes second place. For the period 2006–2008, all models fail the CLR backtests and produce very low p values.

For the next three two year periods however, AGNTS strongly dominates other contenders and exhibits the best backtesting results in general. This is true for all three CLR tests—'uc', 'ind' and 'cc'. Finally, we note that similar to the backtests reported in Tables 2 and 3, it is hard to distinguish the three tests on the basis of the 'ind' test. It is the 'uc' and the 'cc' tests that determine which model performs better.

Hence on the basis of extensive statistical testing—in sample, out-of-sample and by a battery of different backtests—we have shown the superiority of statistical modeling of DJIA asset returns by the AGNTS stochastic process. Hence while studying the portfolio optimization program corresponding to different risk measures, we choose to model asset returns by the AGNTS stochastic process.

6. Results

We wish to solve the following optimization program:

$$\min_{w_t \in \Delta^{n-1}} \left(C\rho(w_t) + \lambda \left(\alpha + \beta \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}| + \gamma \sum_{i=1}^{n} (w_{i,t} - w_{i,t-1})^2 \right) - w_t^\top \mu \right) \tag{19}$$

where at each point in time, the portfolio return $r_p = w^\top r$ is an NTS random variable;[13] and evolves in time according to an AGNTS stochastic process. We assume that the risk aversion parameter $C = 1$, the transaction cost aversion factor $\lambda = 1$ and constants $\alpha = \beta = \gamma = 0.0050$ (50 basis points) and note that such assumptions on transaction cost constants in the equity markets are common in literature Fabozzi et al., 2006, p. 76. We also note that the assumption of the risk aversion constant $C = 1$ is presented as a benchmark case. Indeed, the general results about portfolio performance with different risk aversion constants are qualitatively similar to the one explored here. We report cumulative returns based portfolio performance results for the case with $C = 2$ and $C = 1/2$ in the appendices.

Before analyzing portfolio performance measures in detail, we briefly present descriptive statistics of daily realized returns generated under different definitions of portfolio risk in Table 5. As a cursory glance reveals, the daily realized return distributions under all definitions of risk are skewed and highly leptokurtotic. The most skewed return distribution belongs to the mean-FH program which displays extreme right skewness. The mean-standard deviation program is at the opposite extreme and exhibits highly negatively skewed behavior.

---

[12] This test is also called the "proportion of failures" test in Kupiec, 1995.

[13] $r = (r_1, \ldots, r_n)^\top$ being the individual asset return vector.

Table 2

p values of the CLR test for individual assets in DJIA at 99% VaR. The CLR null hypothesis is that the sequences of exceedances are Bernoulli distributed with its parameter equal to the significance level. For the tests uc and cc, AGNTS$\succeq$AGT$\succeq$AGNormal for 19 out of 29 stocks. Among the remaining 10, AGT wins 8 while AGNormal wins 2 (for Nike (NKE) and Procter & Gamble (PG)). For the CLR independence test, all statistical models report high p values and perform fairly well.

	99% VaR								
	uc			ind			cc		
	AGN	AGT	AGNTS	AGN	AGT	AGNTS	AGN	AGT	AGNTS
AXP	0.0084	0.3553	0.3673	0.5148	0.6509	0.6719	0.0251	0.5888	0.6388
BA	0.0002	0.5087	0.6956	0.3196	0.6715	0.6923	0.0005	0.7347	0.8565
CAT	0.0001	0.0004	0.0020	0.4109	0.4441	0.4788	0.0003	0.0015	0.0066
CSCO	0.0042	0.0302	0.0922	0.4966	0.5522	0.5908	0.0132	0.0802	0.2098
CVX	0.0042	0.0322	0.0539	0.2273	0.5432	0.5714	0.0080	0.0801	0.1330
DD	0.0084	0.3553	0.5087	0.5148	0.6509	0.6715	0.0251	0.5888	0.7347
DIS	0.0094	0.6826	0.3563	0.5008	0.6923	0.6509	0.0251	0.8565	0.5627
GE	0.0084	0.9083	0.6956	0.5148	0.7133	0.6923	0.0251	0.9286	0.8565
GS	0.1512	0.8666	0.6956	0.6106	0.7346	0.6902	0.3135	0.9309	0.8475
HD	0.0539	0.1644	0.9083	0.5714	0.8213	0.7133	0.1330	0.3708	0.9286
IBM	0.0000	0.0302	0.0539	0.3710	0.1680	0.1503	0.0001	0.0369	0.0555
INTC	0.0922	0.5087	0.8666	0.5908	0.6715	0.7346	0.2097	0.7347	0.9309
JNJ	0.00008	0.6956	0.5087	0.4109	0.6923	0.6715	0.0003	0.8565	0.7347
JPM	0.0042	0.3553	0.2370	0.4966	0.6509	0.6306	0.0132	0.5888	0.4427
KO	0.0163	0.8666	0.6461	0.5334	0.7346	0.7560	0.0459	0.9309	0.8575
MCD	0.0042	0.6957	0.5087	0.4966	0.6923	0.6715	0.0132	0.8565	0.7347
MMM	0.0000	0.0539	0.1512	0.0172	0.5714	0.6106	0.0000	0.1330	0.3135
MRK	0.0001	0.0084	0.0539	0.0448	0.2065	0.1503	0.0001	0.0140	0.0555
MSFT	0.0043	0.9083	0.9293	0.4966	0.7133	0.7243	0.0132	0.9286	0.9495
NKE	0.1513	0.0913	0.0922	0.6106	0.5908	0.5908	0.3135	0.2008	0.2097
PFE	0.0539	0.9083	0.5087	0.5714	0.7133	0.6715	0.1330	0.9286	0.7347
PG	0.5067	0.0301	0.0302	0.0772	0.5522	0.5832	0.1687	0.0801	0.0825
T	0.0000	0.0539	0.0922	0.2679	0.5714	0.5908	0.0000	0.1389	0.2097
TRV	0.0002	0.1512	0.0922	0.4274	0.6106	0.5908	0.0007	0.3135	0.2097
UNH	0.0539	0.5087	0.5257	0.5715	0.6715	0.6995	0.1330	0.7113	0.7347
UTX	0.0163	0.0922	0.0939	0.5334	0.5908	0.6208	0.0459	0.2097	0.2097
VZ	0.0009	0.0302	0.0539	0.4613	0.5511	0.5766	0.0032	0.0801	0.1330
WMT	0.0002	0.0302	0.0341	0.4274	0.5332	0.5522	0.0007	0.08001	0.0924
XOM	0.0000	0.0042	0.1512	0.2934	0.4966	0.6106	0.0000	0.01312	0.3135
$p > 0.05$	7/29	20/29	26/29	28/29	29/29	29/29	7/29	25/29	28/29
$p > 0.01$	9/29	25/29	28/29	29/29	29/29	29/29	17/29	28/29	28/29

Table 3

p-values of CLR test for DJIA portfolio VaR. The CLR null hypothesis is that the sequences of exceedances are Bernoulli distributed with its parameter equal to the significance level. For 99% and 95% VaR values, the AGNTS model comprehensively beats its rivals AGT and AGNormal. This is true for all three tests—the AGNTS model produces uniformly higher p values. For the aforementioned confidence levels, the AGT model captures the second spot. The performances for the independence test, though, are practically indistinguishable since all models report relatively high p-values. This ordering changes however, for the 90% confidence level in which AGNormal emerges as the winner.

uc			ind			cc		
AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS
99% VaR								
0.0004	0.0020	0.0163	0.4441	0.4788	0.5334	0.0015	0.0066	0.0459
95% VaR								
0.0000	0.0002	0.0063	0.9292	0.1906	0.2557	0.0003	0.0004	0.0125
90% VaR								
0.4634	0.0001	0.0005	0.3750	0.8780	0.5171	0.5156	0.0006	0.0018

Table 4

Multi-period p-values of CLR test for 99% portfolio VaR. The CLR null hypothesis is that the sequences of exceedances are Bernoulli distributed with its parameter equal to the significance level. For the first two-year period from December 2004 to December 2006, the AGNormal model emerges as the victor, while for the next two year period, all models fail and report p-values very close to 0. For the next three two year periods, the AGNTS model reports uniformly higher p-values for all three tests while the AGT model occupies the second place. In keeping with the results reported in Table 3, however, all models perform well in the independence tests and exhibit relatively robust p-values for all time period partitions.

	uc			ind			cc		
	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS	AGNormal	AGT	AGNTS
2004–2006	0.9566	0.1937	0.2217	0.7474	0.6052	0.6291	0.9481	0.3760	0.3951
2006–2008	0.0002	0.0001	0.0001	0.3288	0.2971	0.3121	0.0007	0.0002	0.0003
2008–2010	0.0935	0.6228	0.9566	0.5604	0.6988	0.7433	0.2068	0.8222	0.9481
2010–2012	0.0922	0.6125	0.9587	0.1488	0.6929	0.7474	0.0864	0.8326	0.9488
2012–2014	0.2006	0.2001	0.6362	0.6067	0.6012	0.7000	0.3862	0.4362	0.8302

Table 5
Descriptive statistics of realized returns of mean risk programs. As may be observed, the daily realized returns obtained under different mean risk formulations display significant skewness and excess kurtosis. The centralized moments presented here have been compiled from 2524 daily realizations of optimal portfolios under different risk regimes. While all daily optimal portfolio returns show similar values for standard deviations, they differ starkly in terms of skewness, with FH optimal portfolios significantly skewed to the right. All programs generate significant excess kurtosis as well, with standard deviation posting the highest value: 21.1971. (Excess kurtosis = Kurtosis - 3.) Also reported are the overall cumulative returns corresponding to the last day of analysis (December 31, 2014). FH optimal portfolios post the highest cumulative returns. The cumulative returns based ordering is: FH $\succeq$ AVaR $\succeq$ DJIA $\succeq$ EquallyWeighted $\succeq$ VaR $\succeq$ StandardDeviation

Mean risk	Std dev	Skewness	Kurtosis	Cum returns
DJIA	0.0118	−0.0846	14.0858	0.5009
Equally weighted	0.0123	−0.0290	14.8372	0.4652
Mean-std deviation	0.0098	−0.7712	24.1971	0.2836
Mean-VaR	0.0126	0.1886	19.7908	0.2859
Mean-AVaR	0.0129	0.1337	20.3799	0.5226
Mean-FH	0.0134	0.4908	17.6900	0.8333

All daily realized return distributions are leptokurtotic. Indeed comparison with the benchmark case of the normal distribution, whose kurtosis is 3, the *excess* kurtosis displayed by mean risk programs ranges from 11.09, for the case of the DJIA index, to an extremely high value of 21.20 for the mean-standard deviation program. Such significant deviations from normality can also be seen in the overall empirical densities of the daily mean risk realized returns.

Having discussed descriptive statistics of daily realized returns' distributions under different mean risk programs, we turn our attention to their portfolio performance. We employ two main out-of-sample criteria to evaluate portfolio performance: cumulative returns produced under different risk measures; and annualized reward-to-risk ratios.

6.1. Measuring portfolio performance: cumulative returns

On the basis of optimal weights chosen according to each optimization program, we construct four different series of cumulative returns, corresponding to standard deviation, 99% VaR, 99% AVaR and FH as measures of risk. As per our out-of-sample rolling window estimation, the first window starts on December 22, 2004 and the last window ends on December 31, 2014—a total of 2524 daily optimizations for each definition of risk.

We present the time series of cumulative returns based on optimal weights in Fig. 4. As can be seen from relative performances under different risk measures, FH strongly dominates cumulative returns produced by other measures of risk. Indeed, it is easy to observe that for the time period in question, FH$\succeq$AVaR$\succeq$DJIA$\succeq$Equally Weighted$\succeq$VaR$\succeq$Standard Deviation.

While it is true that the advent of the Great Recession and the subsequent crash in the American stock markets cause all optimization programs to suffer losses, Foster–Hart optimal portfolios are the quickest to bounce back and begin posting positive returns from as early as the third quarter of 2009. With the exception of Q1–Q3 2005 and Q4 2008–Q2 2009, where all optimal portfolios perform badly, the FH optimal portfolio produces strictly higher cumulative returns for all other time periods. The same trend is true if we vary the risk aversion constant from $C = 1$ to $C = 2$ and $C = 1/2$, as may be seen in Figs. 6 and 7 in the appendices, where FH more strongly outperforms when risk aversion is higher, especially post-crisis. These results are similar in spirit to those reported in Leiss et al. (2015) which presents evidence that the representative investor's risk aversion increases post crisis after unsustainable super-exponential growth expectations during the pre-crisis period.

Among other programs, AVaR performs the best, while Value at Risk and Standard Deviation finish third and last respectively. This is intuitive, since as the accuracy of risk measures increase (from standard deviation to FH risk)—so does the optimization program's ability to choose risk minimizing weights.

6.2. Measuring portfolio performance: reward-to-risk ratios

We augment our cumulative returns' based analysis with a comparative analysis of reward-to-risk ratios. In general, reward-to-risk ratios are defined as Rachev et al., 2008, p. 318:

$$RR(r_p) = \frac{v(r_p - r_b)}{\rho(r_p - r_b)} \tag{20}$$

where as before, $r_p = w^\top r$ is the portfolio return, r_b is the benchmark return, which in our case is the DJIA index returns; and $\rho(\cdot)$ is a risk measure. $v(\cdot)$ is some reward measure, which in our study, we assume to be the mean return: $v(\cdot) = \mu(\cdot)$.

Different definitions of risk produce different reward-to-risk ratios. According to the revised definition in Sharpe (1994), choosing $\rho(r - r_b) = \sigma(r - r_b)$ defines the celebrated Sharpe Ratio. Analogously, for other definitions of portfolio risk, we define Mean-VaR ratio, Mean-AVaR ratio and the Mean-Foster–Hart Ratio.[14]

$$\text{Sharpe Ratio} = \frac{\mu(r_p - r_{DJIA})}{\sigma(r_p - r_{DJIA})} \tag{21}$$

$$\text{Mean-VaR Ratio} = \frac{\mu(r_p - r_{DJIA})}{\text{VaR}(r_p - r_{DJIA})} \tag{22}$$

$$\text{Mean-AVaR Ratio} = \frac{\mu(r_p - r_{DJIA})}{\text{AVaR}(r_p - r_{DJIA})} \tag{23}$$

$$\text{Mean−Foster−Hart Ratio} = \frac{\mu(r_p - r_{DJIA})}{\text{FH}(r_p - r_{DJIA})} \tag{24}$$

We compute perfomance measures for different portfolios—equally weighted, standard deviation optimal, VaR optimal, AVaR optimal and FH optimal. For each of the above, we evaluate their performance based on *all* reward-to-risk ratios: Sharpe, Mean-VaR, Mean-AVaR and Mean-FH. As is customary in literature, we compute out-of-sample, ex-post, annualized reward-to-risk ratios for each risk measure in Table 6.[15]

We find strong evidence that FH optimal portfolios produce the best risk adjusted returns. As a glance at Table 6 reveals, this outperformance is robust to the definition of RR ratio. Indeed, each row corresponds to a different performance measure, ranging from Sharpe Ratio to Mean FH ratio. For each such definition, FH optimal portfolios produce the highest performance—from 0.3311 for Sharpe Ratio to 0.0235 for the Mean FH ratio. The order of performance for different optimal portfolios is robust to the definition of RR ratio as well. For all the above definitions, we observe the following order:

$$\text{FH} \succeq \text{AVaR} \succeq \text{Equally Weighted} \succeq \text{VaR} \succeq \text{StdDev} \tag{25}$$

Only FH optimal and AVaR optimal portfolios are able to beat the market, as reflected by the respective ratios' positive signs. The equally weighted portfolio performs almost as well as

---

[14] We note that there are alternative names for these ratios as well. For example, Caporin and Lisi (2011) and Zakamouline (2009) call our Mean-VaR ratio as "expected return over absolute VaR" and "reward to VaR" ratios respectively. Similarly what we define to be the "Mean-AVaR ratio is also referred to as "STARR", or Stable Tail-Adjusted Risk Return Ratio Rachev et al., 2008, p. 326. We also clarify that the use of the abbreviation "RR ratio" is in no way related to the "Rachev Ratio" in Biglova et al. (2004), which in fact, is a performance measure obtained by taking the ratio of AVaRs.
[15] We thank an anonymous referee for suggesting the construction of this table.
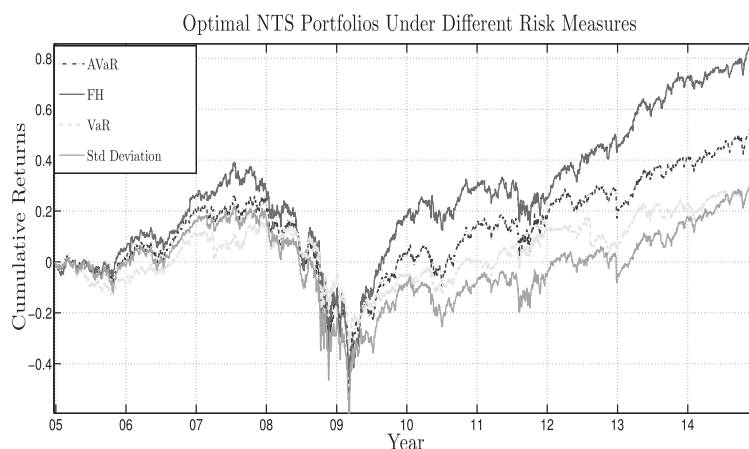
Fig. 4. Comparing portfolio performances under different definitions of risk under the assumption of a risk aversion constant $C = 1$. In terms of the time series of overall cumulative returns, one can clearly observe the almost uniformly higher returns delivered for the Foster–Hart optimal portfolios. Average Value at Risk at 99% confidence level occupies the second spot, followed by VaR at 99% confidence and standard deviation respectively.

Table 6

Annualized Reward-to-Risk Ratios (RR Ratios) for different risk measures. The benchmark for the RR ratios is the DJIA index. The definition of the ratios are $\frac{\mu(r_p^i - r_{DJIA})}{\rho^j(r_p^i - r_{DJIA})}$, where r_p^i are the portfolio returns implied by strategy i, (equally weighted, standard deviation, VaR, AVaR, FH optimization), and ρ^j denotes risk measures: standard deviation, VaR, AVaR and FH respectively. The RR ratios are computed from out-of-sample daily risk estimates which are then converted to annualized estimates for the period 2004–2014. For both VaR and AVaR, the confidence level is 99%. For all definitions of performance measures (RR ratios) the FH optimal portfolios dominate the others and deliver highest risk adjusted returns.

RR Ratio	Equally weighted	Std dev	VaR	AVaR	FH
Sharpe ratio	−0.0983	−0.2298	−0.2557	0.0253	0.3311
Mean-VaR	−0.0371	−0.0858	−0.0890	0.0089	0.1097
Mean-AVaR	−0.0185	−0.0467	−0.0481	0.0107	0.0777
Mean-FH	−0.0083	−0.0132	−0.0148	0.0001	0.0235

the market, posting risk adjusted returns which are slightly negative, while the VaR optimal and standard deviation optimal portfolios are comprehensively outperformed by the DJIA index.[16]

It is curious to observe that as the measure of portfolio risk becomes more sophisticated, the RR ratio—returns per unit of risk incurred—becomes more conservative. This may be observed by comparing the entries in Table 6 column-wise. For example, the Sharpe Ratio of the FH optimal portfolio is 0.3311, its mean-VaR ratio is 0.1097, the mean-AVaR ratio is 0.0777 while the FH ratio is 0.0235. This is so because for the same return, the RR ratio is smaller for more sophisticated risk measures since generally $StdDev \leqslant VaR_{0.99} \leqslant AVaR_{0.99} \leqslant FH$, i.e., the denominator of the RR ratio becomes progressively larger. The same is true for any other column as well—the performance measure decreases in absolute value for all optimal portfolios.

However, since accurate risk measures capture the notion of riskiness better, they are able to accurately discriminate against more risky stocks and help in allocating weights away from them.

Hence we establish the superiority of the Foster–Hart risk measure over both (out-of-sample) criteria of portfolio performance—cumulative returns and reward-to-risk ratios. We

therefore opine that the use of Foster–Hart risk is extremely beneficial from the point of view of institutional investors and fund managers who need to maximize risk adjusted returns for their clients. The use of an accurate risk measure such as Foster–Hart can help them assess risk accurately and hence help them allocate capital to less risky assets.

We also note that our results are relevant to the ongoing debate about whether different performance measures generate different rank orderings over the performances of optimal portfolios. In particular, Eling and Schuhmacher (2007) opine that hedge funds' performance ranking is very similar when evaluated by different performance measures, while Farinelli et al. (2008) and Zakamouline (2009) disagree with their findings and cite significant deviations from return distributions' normality as the main reason for their conclusion. Our study contributes to this discussion by providing evidence that non-normality[17] of return distributions *does* create differences in the performance rankings of different optimization strategies, though we note that our analysis is limited to the DJIA portfolio and one must be cautious when extending our conclusions directly to hedge funds' performance rankings.

6.3. Why is FH superior?

FH risk is constructed by emphasizing bankruptcy-proofing future investments. Why does it prove superior in optimal asset allocation?[18]

We suggest that this is so since the portfolio exits risky positions *earlier* under FH risk management. Indeed, we are not alone in suggesting this explanation. Leiss and Nax (2015) show evidence that FH risk provides additional information beyond that given by other risk measures and captures forward-looking characteristics of the market. Bali et al. (2012) find strong predictive power of FH risk for future downturns. In the same spirit, Bali et al. (2011) note that investing in assets when wealth levels are lower than FH risk leads to bad outcomes leading to decreasing wealth, and in the limit, to bankruptcy. All of these studies suggest that FH risk captures forward-looking expectations of option markets exceptionally well.

We offer a simple explanation for all the above evidence. The Foster–Hart risk measure is extremely sensitive to *left tail* events of the *log return* distribution. This may be understood by observing

---

[16] We note that "beating the market", when the market is represented as the DJIA index, is significantly harder than beating the risk free security rate r_f, often taken to be the US treasury bond. While the original formulation of "reward to variability" ratio Sharpe (1966) uses r_f as the benchmark to beat, our choice of the market index itself as the benchmark is significantly more difficult to outperform against. Hence the relatively smaller values of positive RR ratios for 99% AVaR and FH; and relatively higher negative values for the other mean risk programs.

---

[17] To be more precise, it is in fact the *non-ellipticity* of return distributions that is, perhaps, to blame. The same point is made in Eling and Schuhmacher (2007) as well.

[18] We thank two anonymous referees for suggesting a fuller discussion on this topic.

the log concavity of the defining Eq. 6, which makes FH risk very sensitive to fall in prices. Indeed, if we take two gambles g and f such that their log returns have identical distributions except that the maximum loss of f is $1000 more than that of g ($L_f = L_g + 1000$); by virtue of the property that FH risk is at least as high as maximum loss, $FH(f) \geqslant L_f = L_g + 1000$. Note that this *does not* depend on what the probability of its occurrence is—even if it were very small, say, 10^{-6}—the FH risk may still go up substantially.[19]

This exceptional sensitivity to tiny changes in the left tail is why FH risk is able to capture forward-looking downturn expectations embedded in the option markets. The same theoretical reason explains the results of the current study as well. The out of sample risk computation of the portfolio is done on the basis of Monte Carlo simulations, which, in turn are generated by an ARMA–GARCH process. As illustrated in this paper, the AGNTS model captures the underlying data generating process well. When markets are under distress, the out of sample Monte Carlo simulations produce returns with maximal losses much higher than those in tranquil times. Even if such large losses are distributed with very tiny probabilities, the FH risk assumes levels even higher than those losses, making the portfolio exit risky positions earlier.

7. Conclusions

In this study, we demonstrate that usage of a powerful and accurate measure of risk—the Foster–Hart risk—can be beneficial from the point of view of an investor seeking to maximize risk adjusted returns. We establish the superiority of Foster–Hart optimal portfolios by two different out-of-sample portfolio performance measures: the time series of cumulative returns; and reward-to-risk ratios for the Dow Jones Industrial Average constituents comprising 15 years of daily data: from January 3, 2000, to December 31, 2014. Our analysis indicates that Foster–Hart optimal portfolios outperform AVaR optimal, VaR optimal and Mean–Variance optimal portfolios under a variety of different assumptions on the risk aversion characteristics of an investor (see in particular Figs. 6 and 7).

In order to ensure that our optimal portfolios are not consistently overestimating their performance, we include *ex-ante* general transaction cost functions, including fixed, linear and quadratic penalty terms in the objective function itself. Employing such transaction costs ensures that we are careful to include market microstructure induced frictions; and respect investors' wishes to generate maximum returns while minimizing the cost of rebalancing their portfolio.

In order to compute optimal portfolios, we assume that stock returns of the DJIA constituents follow an ARMA(1,1)–GARCH (1,1) process with multivariate normal tempered stable innovations. In order to demonstrate that such a process faithfully captures the stylized, leptokurtotic behavior of asset returns, we run a battery of in-sample, out-of-sample and backtesting based tests and present extensive evidence of its superiority.

Our results have important applications for fund managers who seek to maximize their clients' portfolio performance while being mindful of the costs that frequent rebalancing of portfolios entail. Our union of fat-tailed statistical modeling with Foster–Hart risk modeling helps in portfolio allocation by diverting investors' wealth to minimally risky stocks. This result is striking also

because it exhibits that careful risk modeling is useful not just from a regulatory point of view, but also from that of a return maximizing fund manager.

Appendices

The following is the description of the appendices to the paper. The first section presents the 30 individual constituents of the Dow Jones Industrial Average (DJIA) index as it stood on December 31st 2014; the second section compares the cumulative returns delivered by Foster–Hart optimal portfolios to the DJIA market index and the benchmark equally weighted portfolio and shows that FH optimal portfolios consistently perform better than both. The third and the fourth sections perform portfolio optimization with investor risk aversion constant $C = 2$ and $1/2$ respectively, thereby considering the cases of risk aversion not limited to the benchmark case of $C = 1$.

As a reminder, the optimization program being solved is:

$$\min_{w \in \Delta^{n-1}} \left(C\rho(w) - w^\top \mu + \lambda \cdot TC(w) \right) \quad (26)$$

where the transaction cost function is:

$$\lambda \cdot TC(w) = \lambda \left(\alpha + \beta \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}| + \gamma \sum_{i=1}^{n} (w_{i,t} - w_{i,t-1})^2 \right) \quad (27)$$

Appendix A. The Dow Jones Industrial Average index

The constituents of the DJIA index, as it stood on December 31, 2014 is presented in Table 7.

Appendix B. Cumulative returns: FH versus DJIA index versus Equally Weighted DJIA Portfolio

Fig. 5 presents the relative cumulative returns' time series of FH optimal portfolios and compares them with those of the DJIA index and the benchmark equally weighted portfolio. The risk aversion constant $C = 1$, the transaction cost aversion parameter $\lambda = 1$ and the transaction cost constants α, β and γ are assumed to be 0.0050 (50 basis points).

Appendix C. Cumulative returns with risk aversion constant $C = 2$

Comparative cumulative returns' time series in the presence of highly risk averse investors with risk aversion $C = 2$ is presented in Fig. 6. The transaction cost constants $\alpha = \beta = \gamma = 0.0050$ and the transaction cost aversion $\lambda = 1$.

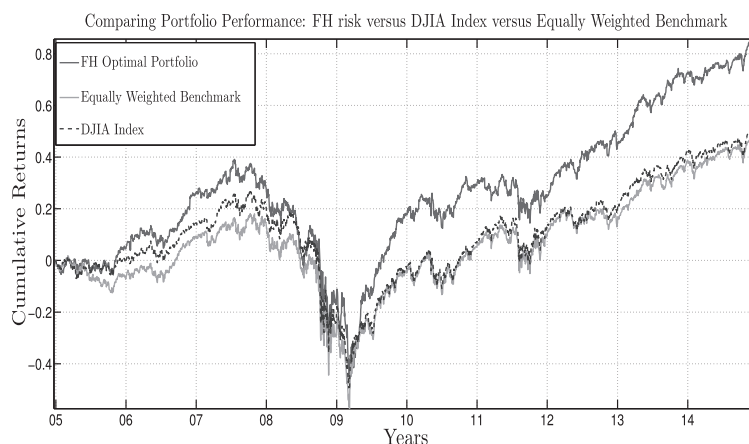As may be seen, the FH optimal portfolio outperforms its rivals comprehensively.

Table 7
List of the 30 DJIA constituents (as of December 31, 2014).

DJIA components		
3 M (MMM)	American Express (AXP)	AT&T (T)
Boeing (BA)	Caterpillar (CAT)	Chevron (CVX)
Cisco Systems (CSCO)	Coca Cola (KO)	DuPont (DD)
ExxonMobil (XOM)	General Electric (GE)	Goldman Sachs (GS)
The Home Depot (HD)	IBM (IBM)	Intel (INTC)
Johnson and Johnson (JNJ)	JP Morgan (JPM)	McDonald's (MCD)
Merck (MRK)	Microsoft (MSFT)	Nike (NKE)
Pfizer (PFE)	Procter & Gamble (PG)	Travelers (TRV)
UnitedHealth Group (UNH)	United Technologies (UTX)	Verizon (VZ)
Visa (V)	Wal-Mart (WMT)	Disney (DIS)

---

[19] Under our setup, "gambles" are identified with the DJIA portfolio's log return distributions (following the AGNTS process) whose support is the entire real line. While ostensibly, this is at odds with the FH definition of a gamble as a bounded random variable, our reliance on scenario generation by means of Monte Carlo simulations ensures that the simulated portfolio log returns $r_p^1, \ldots, r_p^N$ obey $L = -\min\{r_p^1, \ldots, r_p^N\} < \infty$ and hence a well defined FH risk satisfying $L \leqslant FH(r_p) < \infty$.

Fig. 5. Comparing the FH optimal portfolio with the DJIA index and the equally weighted benchmark. While the equally weighted benchmark closely tracks the index, the cumulative returns reported by the FH optimal portfolio strongly outperform both the index and the equally weighted benchmark.
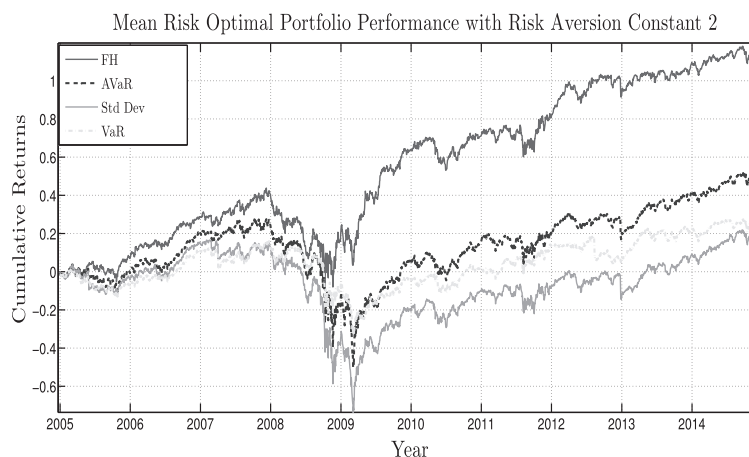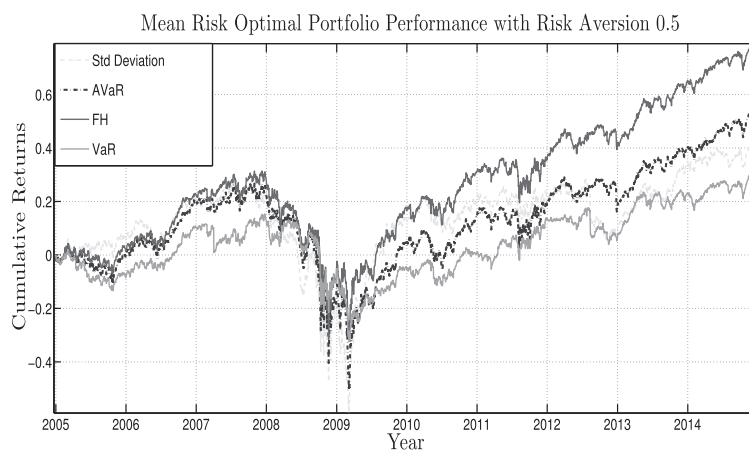

Fig. 6. Comparative performance of mean risk analysis under the assumption of the risk aversion constant $C = 2$. The overall cumulative returns' time series puts FH optimal portfolios in the first place again, thereby showing that its performance is not strongly dependent on the assumptions of the risk aversion parameter being unity. AVaR and VaR are computed for the 99% confidence level.

Appendix D. Cumulative returns with risk aversion constant $C = 0.5$

Fig. 7 shows comparative cumulative returns' time series in the presence of highly risk loving investors with risk aversion $C = 1/2$.

The transaction cost constants $\alpha = \beta = \gamma = 0.0050$ and the transaction cost aversion $\lambda = 1$.

Again, the FH optimal portfolio shows better performance than its competitors.

Fig. 7. Comparative performance of mean risk analysis under the assumption of the risk aversion constant $C = 0.5$. The overall cumulative returns' time series puts FH optimal portfolios in the first place again, thereby showing that its performance is not strongly dependent on the assumptions of the risk aversion parameter being unity. Note however, that in contrast to previous results, the mean-standard deviation portfolio beats the 99% VaR portfolio for the third spot.

References

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1998. Coherent measures of risk. Mathematical Finance 6 (3), 203–228.

Bali, T.G., Cakici, N., Chabi-Yo, F., 2011. A generalized measure of riskiness. Management Science 57 (8), 1406–1423.

Bali, T.G., Cakici, N., Chabi-Yo, F., 2012. Does aggregate riskiness predict future economic downturns? Fisher College of Business WP 2012-03-009.

Barndorff-Nielsen, O.E., Levendorskii, S.Z., 2001. Feller processes of normal inverse Gaussian type. Quantitative Finance 1 (3), 318–331.

Barndorff-Nielsen, O.E., Shephard, N., 2001. Normal modified stable processes. Economics Series Working Papers from University of Oxford, Department of Economics 72.

Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? Journal of Finance 57, 1093–1111.

Bianchi, M.L., Rachev, S.T., Kim, Y.S., Fabozzi, F.J., 2010. Tempered stable distributions and processes in finance: numerical analysis. Mathematical and Statistical Methods for Actuarial Sciences and Finance, 33–42.

Biglova, A., Ortobelli, S., Rachev, S.T., Stoyanov, S., 2004. Different approaches to risk estimation in portfolio theory. Journal of Portfolio Management 31 (1), 103–112.

Caporin, M., Lisi, F., 2011. Comparing and selecting performance measures using rank correlations. Economics: The Open-Access, Open-Assessment E-Journal 5, 1–34. http://dx.doi.org/10.5018/economics-ejournal.ja.2011-10.

Choi, J., Kim, Y.S., Mitov, I., 2015. Reward-risk momentum strategies using classical tempered stable distribution. Journal of Banking & Finance 58, 194–213.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1 (2), 223–236.

Eling, M., Schuhmacher, F., 2007. Does the choice of performance measure influence the evaluation of hedge funds? Journal of Banking & Finance 31, 2632–2647.

Fabozzi, F.J., Focardi, S.M., Kolm, P.N., 2006. Financial Modeling of the Equity Market: From CAPM to Cointegration. Wiley, Hoboken, New Jersey.

Fama, E.F., 1963. Mandelbrot and the stable Paretian hypothesis. Journal of Business 36, 420–429.

Fama, E.F., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M., Tibiletti, L., 2008. Beyond Sharpe ratio: optimal asset allocation using different performance ratios. Journal of Banking & Finance 32 (10), 2057–2063.

Foster, D.P., Hart, S., 2009. An operational measure of riskiness. Journal of Political Economy 117 (5), 785–814.

Haas, M., Pigorsch, C., 2009. Complex Systems in Finance and Econometrics. Springer, New York, Chap. Financial Economics, Fat-Tailed Distributions 308–339.

Hart, S., 2011. Comparing risks by acceptance and rejection. Journal of Political Economy 119 (4), 617–638.

Kim, Y.S., Giacometti, R., Rachev, S.T., Fabozzi, F.J., Mignacca, D., 2012. Measuring financial risk and portfolio optimization with a non-Gaussian multivariate model. Annals of Operations Research, 325–343.

Kim, Y.S., Rachev, S.T., Bianchi, M.L., Fabozzi, F.J., 2010. Tempered stable and tempered infinitely divisible GARCH models. Journal of Banking & Finance 34, 2096–2109.

Kim, Y.S., Rachev, S.T., Bianchi, M.L., Mitov, I., Fabozzi, F.J., 2011. Time series analysis for financial market meltdowns. Journal of Banking & Finance 35, 1879–1891.

Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 6, 6–24.

Leiss, M., Nax, H.H., 2015. Option-implied objective measures of market risk. Working Paper.

Leiss, M., Nax, H.H., Sornette, D., 2015. Super-exponential growth expectations and the global financial crisis. Journal of Economic Dynamics and Control 55, 1–13.

Mandelbrot, B., 1963a. New methods in statistical economics. Journal of Political Economy 71, 421–440.

Mandelbrot, B., 1963b. The variation of certain speculative prices. Journal of Business 36, 394–419.

Mandelbrot, B., 1967. The variation of some other speculative prices. Journal of Business 40, 393–413.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Markowitz, H.M., 1959. Portfolio Selection: Efficient Diversification of Investments. John Wiley and Sons, New York.

Pagan, A., 1996. The econometrics of financial markets. Journal of Empirical Finance 3, 15–102.

Pogue, G.A., 1970. An extension of the Markowitz portfolio selection model to include variable transactions' costs, short sales, leverage policies and taxes. Journal of Finance 25, 1005–1027.

Rachev, S.T., Stoyanov, S.V., Fabozzi, F.J., 2008. Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization. John Wiley and Sons, Hoboken, New Jersey.

Riedel, F., Hellmann, T., 2015a. A dynamic extension of the Foster–Hart measure of riskiness. Journal of Mathematical Economics 59, 66–70.

Riedel, F., Hellmann, T., 2015b. The Foster–Hart measure of riskiness for general gambles. Theoretical Economics 10, 1–9.

Sharpe, W.F., 1966. Mutual fund performance. Journal of Business 39 (S1), 119–138.

Sharpe, W.F., 1994. The Sharpe ratio. The Journal of Portfolio Management 21 (1), 49–58.

Zakamouline, V., 2009. The choice of performance measure does influence the evaluation of hedge funds. Working Paper.