Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Empirical Finance 11 (2004) 353–377
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

Ranking mutual funds using unconventional utility theory and stochastic dominance

H.D. Vinod*

Economics Department, Fordham University, 441 East Fordham Road, Bronx, NY 10458-9993, USA

Accepted 16 June 2003

Abstract

We make the choice between mutual funds realistic and rigorous by incorporating (i) unconventional non-expected utility theories implying four desirable properties (4DPs) of utility functions, and (ii) the fourth-order stochastic dominance (4SD). Kimball's (Econometrica 58 (1990) 53) diminishing prudence is shown to need 4SD. A new weighting incorporates 4DPs into various criteria including the Sharpe ratio. We use 1281 mutual funds rated with one to five stars by Morningstar, a trapezoidal approximation for evaluating 4SD and propose a new cumulative 4SD measure with superior properties for statistical inference. An appendix describes a new maximum entropy (ME) bootstrap for dependent time series.

JEL classification: G11; C14; D81
Keywords: Morningstar; Prospect theory; Portfolio selection; Sharpe ratio; Lower partial moments; Prudence; Maximum entropy; Dependent bootstrap

1. Introduction

With over 12,000 mutual funds now available, investors need tools to help select the best funds in which to invest. A popular mutual fund rating system is provided by Morningstar, which rates mutual funds with one to five stars, with best funds receiving five stars. This paper deals with the following question: How can we supplement or improve Morningstar's solution to the underlying problem by bringing together diverse and newer insights from Economics, Statistics and Finance?

* Corresponding author. Tel.: +1-718-817-4065; fax: +1-718-817-3518.
E-mail address: Vinod@fordham.edu (H.D. Vinod).

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2003.06.002

The relevant insights from Economics are to be found in theories of decision-making under uncertainty, expected utility theory (EUT), and risk aversion. Let Tb3 denote the risk-free return from the 3-month US Treasury Bills. Let x denote the *excess* return (defined as return minus Tb3) from a portfolio of mutual funds, which yields utility $u(x)$. Economists have long sought to make $u(x)$ as flexible and realistic as possible. Let the derivatives of $u(x)$ be denoted by primes. Let D_3 denote a well-known class of utility functions satisfying $u' \geq 0$, $u'' \leq 0$ and $u''' \geq 0$, which satisfies diminishing marginal utility and non-increasing absolute risk aversion (NIARA). Kimball (1990) notes that the Arrow–Pratt risk aversion merely measures how much one dislikes uncertainly, and defines "prudence" by $(-u''')/u''$ as a measure of 'sensitivity of the optimal choice of a decision variable to risk'. Kimball states that "Prudence measures the propensity to prepare and forearm oneself in the face of uncertainty". High-income individuals tend to have reduced prudence evidenced by reduced precautionary savings due to reduced income uncertainty.

Now, a realistic $u(x)$ should exhibit Kimball's non-increasing prudence (KNIP), which is proved (see Lemma 1 below) to require $u'''' \leq 0$. This paper develops methods of choosing mutual funds for agents whose utility functions belong to the class D_4, which extends the D_3 class by further requiring that $u'''' \leq 0$, that is KNIP.

The vast statistical literature comparing probability density function (PDFs) $f(x)$ or cumulative distribution functions (CDFs) $F(x)$ is related to stochastic dominance (SD). Denote SD of order i by iSD with $i=1,\ldots,4$. We shall see that evaluation of the iSD criterion for j-th portfolio (iSDj) involves integrals of $F(x)$ and applies to all $u(x)$ functions in the D_i class. Sir R.A. Fisher's "permutation distribution" (revived by the bootstrap literature) is used by Vinod (2001) in his exact nonparametric 'matched pairs' maximum distance test statistic max(3SDj). We show that the change in investor's terminal net worth can be approximated by (Σ1SDj) and is the most relevant quantity for the investor's problem. Evaluation of higher order dominance involves matrix premultiplications, which are linear transformations. Our statistical testing will use (ΣiSDj) which are shown to be asymptotically normal and efficient.

A strand of recent economics literature focuses on what is paradoxical in conventional EUT. Non-EUT literature is surveyed by Starmer (2000) explaining the Allais paradox with simple graphics. He also plots an inverted S-shaped probability weighting function on a unit square leading to "decision weights." He sites experimental evidence which suggests the real world has loss aversion, skewness preference and violations of independence and monotonicity assumed in conventional EUT. Kahneman and Tversky's (1979) prospect theory recently won a Noble prize for Kahneman. It has an editing phase wherein "dominance heuristic" eliminates stochastically dominated prospects from further consideration—a common practice in choosing mutual funds. The second feature of prospect theory involves distinct attitudes towards gains and losses (e.g. loss aversion) with a reference point at the status quo wealth. This paper lets the status quo wealth increase at the risk-free rate (Tb3) and uses Prelec's (1998) probability weighting function, $\exp(-(-\ln p)^{\alpha})$, shown later in Fig. 1, to provide for distinct treatment of gains and losses from the status quo. We assume that past data on a mutual fund reveals $f(x)$ as the ex ante probabilities of achieving an ordered set of excess returns $x_{(t)}$. Next, we transform these probabilities into decision weights w_t for each ordered $x_{(t)}$. Now we modify our Σ4SDj

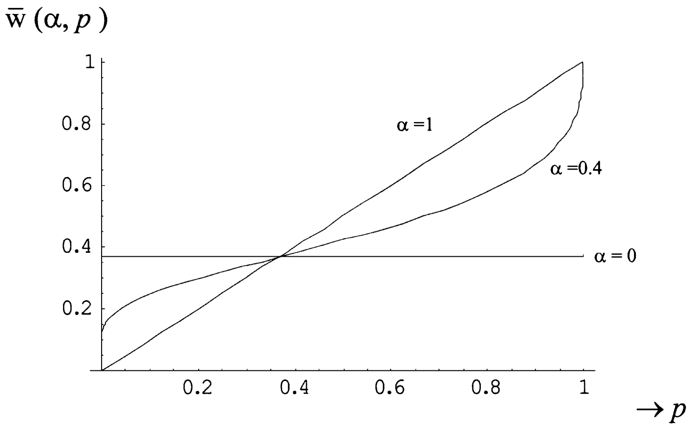

Fig. 1. Plot of $\bar{w}(\alpha, p) = \exp(-(-\ln p)^{\alpha})$ of Eq. (3) for $\alpha = 0.001$ (nearly flat line), $\alpha = 0.4$ (curved line), and $\alpha = 0.999$ (nearly the 45° line, nearly EUT compliant).

statistic quite simply by using the product $[w_t x_{(t)}]$ instead of $x_{(t)}$. This is how we incorporate the recent advances to conventional utility theory as well as non-EUT. To the best of my knowledge, non-EUT methods have not been applied to ranking of mutual funds. This paper fills that gap.

In Finance, this is the portfolio selection problem. Markowitz's mean-variance optimal solution holds, provided $u(x)$ is quadratic and $f(x)$ is normal. Denote the Sharpe ratio as SharpeR $= \bar{x}/s$, where $\bar{x}$ is the mean and s is the standard deviation and is based on Markowitz's solution. SharpeR is widely used and cited in the literature and pedagogy of finance and is often used for ranking stocks and mutual funds. It is known that Sharpe ratio can lead to misleading ranking of portfolios when its numerator is negative. This paper suggests a new solution to this old problem by using a simple 'add factor' to ensure correct rankings. A numerical example with 1281 mutual funds described later shows that negative SharpeR are common in practice. Hence, our solution is potentially important.

More general $f(x)$ and $u(x)$ are used to extend Markowitz's results. Bawa (1975) cites references and his theorem 7 proves that third-order stochastic dominance (3SD) implies dominance under mean-lower partial variance rule. Vinod (2001) considers downside standard deviation (DSD) or the square root of lower partial variance. Vinod (2000) considers inference methods for weighted standard deviations. We suggest a new weighting of downside Sharpe ratio defined as DnSha=$[\bar{x}_w/\text{DSD}]$, where the weights in the numerator are our w_t from the non-EUT considerations.

Section 2 begins with a Taylor series which links moments of $f(x)$ with expected utility from EUT. It briefly reviews relevant non-EUT results and develops implied weights for the portfolio selection problem. Section 3 reviews the relevant stochastic dominance literature and discusses KNIP and Anderson (1996) nonparametric measures for the first four orders of stochastic dominance. Section 4 discusses numerical results using ranking of 1281 international mutual funds and Section 5 has our summary and conclusions. Since our ME bootstrap has applications beyond the present context, it is discussed separately in Appendix A.

2. Moments of *f*(*x*), EUT and non-EUT weights

In this section, we first clarify the link between the expected value Eu(x) and moments of $f(x)$ and then discuss non-EUT insights. Once uncertainty is introduced, x becomes a random variable described by $f(x)$ and maximization of utility becomes ambiguous. Denote the moments of $f(x)$ as follows: $\bar{x}$ = mean, s^2 = variance, and m_k = k-th sample central moment. Now EUT replaces the ambiguity in utility maximization by focusing on its expected value, Eu(x). An elegant feature of EUT is the ingenious use of Taylor series to suggest plausible implications for the first four moments, not just the mean.

Expand $u(x)$ around $\bar{x}$ by the Taylor series and compute the expectation as: $\mathrm{Eu}(x) = \mathrm{Eu}(\bar{x}) + \mathrm{E}(x-\bar{x})u' + \mathrm{E}(x-\bar{x})^2 u''/2! + \mathrm{E}(x-\bar{x})^3 u'''/3! + \ldots,$ Thus,

$$\mathrm{Eu}(x) = u(\bar{x}) + s^2 u''/2! + m_3 u'''/3! + m_4 u''''/4! + \cdots, \tag{1}$$

where the partials are evaluated at the sample mean and $|x-\bar{x}| < 1$ is needed for convergence. Under ordinal utility theory, only the order or the rank of x values matter and one can divide all values by a large number chosen to satisfy $|x-\bar{x}| < 1$. Unfortunately, we cannot similarly change units to satisfy $|x-\bar{x}| < 1$ when x denotes excess returns and the inequality must be explicitly verified. Thus, Eq. (1) has established a link between moments of $f(x)$ and maximization of expected utility. We claim that the EUT implication for m_4 is new in this paper.

Remark 1. Note that to maximize Eu(x), it is not enough that the mean $\bar{x}$ is high. In fact, to maximize the left-hand side of Eq. (1) under EUT we must maximize each positive term on the right-hand side and simultaneously minimize each negative term on the right-hand side. What are the signs? The law of diminishing marginal utility (MU) implies $u'>0$ and $u''<0$. Hence, $s^2\ u''/2!$ is a negative term of Eq. (1) since $s^2 \geq 0$, by definition. This correctly suggests that variance should be as small as possible. We shall explain in Section 3 that a necessary condition for non-increasing absolute risk aversion (NIARA) is $u''' \geq 0$. Hence, the third term of Eq. (1) is a positive term when $m_3>0$, that is, when $f(x)$ has positive skewness. It makes intuitive sense that any EUT-maximizing agent would prefer positive skewness, since it means larger probability of profits than losses. Regarding the fourth term of Eq. (1) note that $m_4 \geq 0$, by definition, and Lemma 1 shows that Kimball's non-increasing prudence (KNIP) requires $u'''' \leq 0$. As wealth or income increases, caution or prudence decreases and a risky choice of long shot bets tends to increase, making the fourth term negative. Again, it makes intuitive sense that the agent wanting to maximize Eu(x) would prefer to minimize m_4, which is a new implication of EUT. Ali (1974) shows that preference for low kurtosis (m_4) means a preference for less thick tails, which amounts to a preference for avoiding the risk associated with low extreme observations in the left tail, even if they are matched by high extreme observations in the right tail. We stop at fourth derivatives. There is no economic reason to link Eu(x) with fifth or higher derivatives of $u(x)$ or m_j for $j \geq 5$. Moreover, m_j for $j \geq 5$ have poor sampling properties.

Let us now turn to unconventional utility theory by describing non-EUT-inspired experiments. These experiments study actual behavior of human economic agents. Imagine x changes from US\$10 to \$20 and we compare this change to the change from US\$100,010 to \$100,020. The absolute change is the same, but the latter has lesser impact.

In probability comparisons, a chance of 1/3 is half of 2/3. Similarly, a chance 1 in 3 million is also half of a 2 in 3 million chance. Despite the constant ratio, the latter has much less impact on the human decision maker, as the ratios get smaller (denominators get larger). Agents are quite willing to buy a lottery with a small chance of winning and yet buy fire insurance against a similarly small chance of their own house being burnt down. Agents exhibit "loss aversion" implying $u'(x) < u'(-x)$ for $x \geq 0$. Prelec (1998) gives detailed description of failures of EUT and lists four desirable properties (4DPs) of utility functions described in Remark 3 below.

Lorenz curves were initially developed in the context of income distributions to facilitate comparisons by bringing them to a common scale plotted on [0,1] intervals along both axes, i.e., the unit square. The Lorenz transformation should be credited with much progress in studies of income distributions and in the non-EUT literature. See Davidson and Duclos (2000) for an example of the former with useful references. Indeed, the 4DPs are stated in terms of CDFs mapped on one axis similar to Lorenz. Due to recent lowering of computational costs, we use a nonparametric "empirical" CDF, denoted by p_t which goes from 0 to 1 in increments of $(1/T)$, where T denotes the number of observations. Prospect theory (Kahneman and Tversky, 1979) suggests useful "decision weights" associated with $p_t \in [0,1]$, the CDF of x. Note that our financial excess returns $x \in (-\infty, \infty)$, are simply quantiles related to the p_t by the standard "probability integral transformation" described below in Eq. (2) where we omit the subscript t to avoid notational clutter.

The probability integral transformation is: $p = F(x)$. Given a number $p \in [0,1]$ a "quantile of x of order p," or $(100p)$-th percentile of the distribution is defined as:

$$x(p) = \inf\{x \mid F(x) \geq p\}. \tag{2}$$

It is simply the inverse of the CDF (or empirical CDF) and satisfies the probability relations: $\Pr(x \leq x(p)) \geq p$ and $\Pr(x \geq x(p)) = 1 - p$. It is well known that p is distributed as a uniform random variable $U(0,1)$, irrespective of the original distribution of returns $f(x)$. Hence, the estimated quantiles have better statistical sampling properties than the higher moments in Eq. (1). Levy's (1992) survey indicates further advantages of quantiles as the inverse function of the CDF. In our context, we first arrange the excess returns x_t for $t = 1,2,\ldots,T$ in an increasing order of magnitude and denote the order statistics by $x_{(t)}$. The corresponding cumulative probabilities are: $p_t = 1/T, 2/T, \ldots, 1$.

Remark 2. Prospect theory suggested by Kahneman and Tversky (1979) studies the utility associated with $p_t = F(x_{(t)})$ a $U(0,1)$ variable, instead of x itself. Since $x = F^{-1}(p)$, the utility function $u(x)$ becomes $u[F^{-1}(p)]$. Prospect theory and rank-dependent models suggest a nonlinear transformation of p into "decision weights" $w(p)$ defined on the unit square. Assuming that past mutual fund data yield $F(x)$, the CDF of ex ante probabilities p_t of $x_{(t)}$, we can transform p_t's into decision weights w_t. Starmer (2000) explains the technical reasons why the use of cumulative p is useful in avoiding violations of monotonicity and transitivity assumptions of the EUT.

Remark 3. Recall the Non-EUT-inspired experiments mentioned above. Given a return x, its perceived importance to the economic agent depends on the relative position of $x_{(t)}$ in its CDF, that is, it depends on p_t. Let us omit the subscript t and let $\bar{w}(\alpha,p)$ denote the perceived importance to human agents of p with an additional parameter α described below. Prelec (1998) provides a summary of the current knowledge about the perceived importance in an elegant graph. The graph, illustrated in our Fig. 1, is seen to be defined over the unit square mentioned above. If EUT were valid, the perceived importance $\bar{w}$ (α,p) will equal p implying a 45° line marked $\alpha = 1$ in Fig. 1. Now, Non-EUT-inspired experiments suggest the shape of the relation between p on the horizontal axis and its perceived importance $\bar{w}$ (α,p) on the vertical axis satisfies following four desirable properties (4DPs).

(i) "Regressive" means $\bar{w}(\alpha, p)$ intersects the 45° line (see the line marked $\alpha = 1$ in Fig. 1) from above. Thus, the weight $\bar{w}(\alpha, p)$ or perceived importance looms high when p is small. For example, it means risk aversion for small-p losses in the form of eagerness to buy fire insurance on the left side.
(ii) "Asymmetry" means that the fixed point $p = \bar{w}(\alpha, p)$ occurs at $p = 1/3$, not at $p = 0.5$. It is analytically convenient to use the fixed point at p=(1/e) = 0.37, where the lines marked $\alpha = 0$, 0.4 and 1 intersect in Fig. 1, instead of $p = 1/3$. After all, 0.37 is close enough to the theoretical value of $p = 1/3 = 0.33$ and far from 0.5. We interpret asymmetry as a desire for positive skewness of $f(x)$.
(iii) "Inverted S-shaped" means that it is concave in gains near the origin and then becomes convex in losses, shown as the curved line in Fig. 1. Kahneman and Tversky (1979) interpret the shape restrictions of prospect theory in terms of "diminishing sensitivity" $u''(x) \leq 0$ for $x \geq 0$, and $u''(x) \geq 0$ for $x \leq 0$, and "loss aversion" implying $u'(x) < u'(-x)$ for $x \geq 0$.
(iv) "Reflective" utility means assigning as much utility weight to a given loss-probability as to a given gain-probability.

A limitation here is that we are ignoring the joint distribution of portfolios being compared. Prelec (1998) defines $\bar{w}(p,\alpha)$, a simple, powerful and flexible function of α which satisfies the 4DPs listed above. Its slope equals α at its inflexion point. We have

$$\bar{w}(p, \alpha) = \exp(-(-\ln p)^{\alpha}), \text{ where } 0 < \alpha < 1. \tag{3}$$

Fig. 1 plots Eq. (3) with p measured in increments of 0.005 ($T = 201$). If $\alpha \to 0$ we have a step function, which is flat everywhere except at the end-points, implying maximum departure from EUT. Note that for graphing convenience, the curve in Fig. 1 marked as $\alpha = 0$ is actually evaluated at $\alpha = 0.001$, and is flat in the middle. If $\alpha \to 1$ we have p=(p,α), or perfect compliance with EUT (the 45° line of Fig. 1). Smaller the α, greater is the departure from EUT. For example, the $\alpha = 0$ line departs more from the EUT than the $\alpha = 0.4$ curved line in Fig. 1. Hence, we interpret α as our EUT-compliance parameter.

Next, we transform $\bar{w}(p,\alpha)$ into decision weights $w_t(\alpha, p_t)$ applied to a particular observation as follows. Since we seek weights on order statistics $x_{(t)}$, we "first difference" the weights in Eq. (3). For any given α, the appropriate weights for $x_{(t)}$ are:

$$\begin{aligned} w_t(\alpha, p_t) &= \exp(-(-\ln p_t)^{\alpha}) \text{ if } t = 1, \text{and} \\ &= \exp(-(-\ln p_t)^{\alpha}) - \exp(-(-\ln p_{t-1})^{\alpha}) \text{ for } t > 1. \end{aligned} \tag{4}$$

The weight w_t for the first observation is different from the w_t for other observations. We need to use Eq. (4) only once for each choice of T to get the weights w_t. What is the interpretation of these weights? Note that when $T=5$, all five weights for an EUT-compliant person having $\alpha=1$ equal 0.2 ($=1/5$). However, if $\alpha=0.1$ weights are: (0.350, 0.0207, 0.0215, 0.0303, 0.577), if $\alpha=0.5$ weighs become: (0.281, 0.103, 0.105, 0.134, 0.376), and if $\alpha=0.9$ they are: (0.216, 0.181, 0.182, 0.193, 0.228). The smallest value ($=x_{(1)}$, possibly a loss) of x_t is weighted more than the middle values. In addition, the largest value ($=x_{(T)}$, possibly a big gain) is also weighted by a number larger than $1/T$.

Given $\alpha\in[0,1]$, the decision weights w_t of Eq. (4) are readily computed by a computer algorithm (A Gauss proc is available upon request). Conversely, given sufficient data from the past behavior of an individual, we can estimate the implicit α appropriate for that individual. An individual with a small α adheres less closely to the norms of the expected utility theory, and is less EUT-compliant. Lattimore et al. (1992) suggest a weight function with two parameters.

$$w(\alpha,\beta,p_t) = \alpha(p_t)^{\beta}/\left[\alpha(p_t)^{\beta}\sum_{k=1,k\neq 1,}^{T}\alpha(p_t)^{\beta}\right] \tag{5}$$

The EUT is a special case when $\alpha=\beta=1$. When $\beta<1$, the curve (Eq. (5)) is shaped as inverted-S and when $\alpha<1$, the weights do not add up to unity and are called prospect pessimism.

We claim that the weights in Eqs. (4) or (5) are important tools in solving the portfolio selection problem in a more flexible and realistic setting of non-EUT compliant agents. Professionals often use certain summary statistics to weed out inefficient portfolios from a large list, before making detailed comparisons. The simplest summary statistic is the mean and we define its weighted versions relevant for any given α as:

$$\bar{x}_w = \sum_{t=1}^{T} w_t x_{(t)} \Big/ \sum_{t=1}^{T} w_t,$$

where weights from Eq. (4) depend on α.

Remark 4. The ex ante probabilities for agents failing to satisfy EUT (i.e., when $\alpha<1$) can be approximated by simply replacing the PDF $f(x_t)$ by $f[w_t x_{(t)}]$. A simple multiplication of the ordered $x_{(t)}$ and w_t defined in Eq. (4) is all that is needed to satisfy 4DPs. Thus, instead of the average return $\bar{x}$, the weighted average $\bar{x}_{\mathrm{w}}$ is claimed to be a more realistic criterion for comparing portfolios. In the literature dealing with robust statistical methods, trimming and winsorization are used to, respectively, delete or down-weight extreme observations. The basic idea is that the extreme circumstances that lead to extreme outcomes are unlikely to repeat themselves and hence distract from the main properties of $f(x)$. By contrast, in portfolio analysis, the extreme losses and profits are very important and get the highest $w_t(\alpha)$ even for $\alpha=0.9$ near unity. This is an interesting example where econometricians need caution in adapting statistical ideas.

Similar to weighted mean, the weighted variance from a sample of size T is:

$$s_{\mathrm{w}}^2 = \sum_{t=1}^{T} w_t(\alpha)(x_{(t)} - \bar{x})^2 \Bigg/ \left\{ T \sum_{t=1}^{T} w_t(\alpha) - 1 \right\}, \tag{6}$$

where the w_t weights are from Eq. (4), as above. Generally, the sum of the weights should be unity, in which case the denominator is the usual $(T-1)$. The weighted standard deviation from the square root of Eq. (6) is denoted as WSD_α and is a new summary statistic to measure the risk experienced by non-EUT compliant agent. The refined weights from Eq. (5) with the additional parameter β would lead to $\mathrm{WSD}_{\alpha,\beta}$. The additional parameter needs to be specified for each agent and adds an extra dimension in the ranking of mutual funds. Since we have 1281 funds to compare in our Section 4, the additional dimension makes results of using Eq. (5) too cumbersome to report and left for future work.

Now consider a particular weighting scheme which gives equal weight on an abridged range and zero weight elsewhere. In statistics, partial moments of x are moments defined over a subset of the original range $[x^*, x^{\mathrm{U}}]$. If $f(x)$ is unit normal or $N(0,1)$ with $x \in (-\infty, \infty)$, lower partial moments are defined over the lower part of the range $(-\infty, 0]$, or the downside. The downside standard deviation is DSD=$[-x\phi(x) + \Phi(x)]^{0.5}$, where $\phi(x)$ is the PDF and $\Phi(x)$ is the CDF of $N(0,1)$. Hence, this formula gives the theoretical DSD for $x \in (-\infty, \infty)$ upon using the $N(0,1)$ tables. It is obvious that DSD is a better measure of risk of loss than the usual standard deviation, which treats losses and gains in a symmetric manner.

Next, we define the weighted versions of higher central moments as $m_{j\mathrm{w}}$ for $j \geq 3$:

$$m_{j\mathrm{w}} = \sum_{t=1}^{T} w_t(\alpha)(x_{(t)} - \bar{x})^j \Bigg/ \left\{ T \sum_{t=1}^{T} w_t(\alpha) - 1 \right\}. \tag{7}$$

In particular, the weighted third central moment, which can be negative, is needed to define Pearson's measure of skewness. In Eq. (7), the weights depend on α, the EUT-compliance parameter of Eq. (3). The estimated $m_{3\mathrm{w}}$ (skewness) has a sampling distribution with high variance due to the third power of deviations from the mean needed in Eq. (7) with $j=3$. Now, the (possibly negative) skewness measure is defined by scaling with appropriate standard deviation as:

$$S_{\mathrm{k}} = m_{3\mathrm{w}} / s_{\mathrm{d}w}^3. \tag{8}$$

This measure is applicable even if $\alpha \neq 1$. All traditional moments are a special case of our weighted moments when $\alpha = 1$. Our main point is that weighted moments are typically more meaningful for portfolio theory, since all agents do not always comply fully with the axioms of EUT.

We have already defined the Sharpe Ratio as: SharpeR$=\bar{x}/s$. Vinod and Morey (2000) discuss the double bootstrap confidence intervals for the Sharpe ratio. Morey and Vinod (2001) propose a version of the Sharpe ratio, which allows for estimation risk. A new weighted version of the Sharpe ratio using the downside standard deviation (DSD) can incorporate investor preference for positive skewness. Since there is no reason to retain unweighted $\bar{x}$ in the numerator, let us define downside Sharpe ratio as: DnSha=$[\bar{x}_w/\mathrm{DSD}]$, where $\bar{x}_w$ needs the EUT α parameter. Similarly, we can define a weighted SharpeR as WtSha=$[\bar{x}_w/\mathrm{WSD}_\alpha]$, where we use the weighted standard deviation defined earlier. For brevity, we evaluate only the DnSha for the 1281 mutual funds in our numerical work described later in Section 4 and omit the WtSha.

2.1. A Sharpe ratio anomaly

If SharpeR are all positive, it is perfectly appropriate to rank funds by the Sharpe ratio, seeking to maximize the SharpeR. An old unsolved problem is that when some ratios have negative numerators SharpeR gives incorrect ranking of money-losing funds. Consider our example 1 (EX1) with two money-losing funds A and B. They have a common mean of (-1) and two distinct standard deviations 1 and 10. Now $\mathrm{SharpeR}_{(A)} = \bar{x}/s = -1$ is more negative than $\mathrm{SharpeR}_{(B)} = -1/10 = -0.1$ implying that $\mathrm{SharpeR}_{(A)} < \mathrm{SharpeR}_{(B)}$. This suggests the incorrect ranking that we should prefer fund B, even though B is doubly undesirable with negative return $(= -1)$ and higher risk measured by standard deviation $(= 10)$. We claim that using absolute values $|\bar{x}|$ in the numerator of a revised Sharpe ratio cannot solve this problem. To see this, consider our second example (EX2) where two portfolios C and D have $\bar{x}^{(C)} = -1$, $\bar{x}^{(D)} = -3$, and $s^{(C)} = s^{(D)} = 1$. Here, the absolute values will obviously give the wrong ranking: $\mathrm{SharpeR}_{(C)} = 1 < \mathrm{SharpeR}_{(D)} = 3$. We need a new solution, which does not change the ranking for moneymaking funds, and fixes the problem only for money losing funds. We suggest adding a large enough add factor (AF>0) and defining a new adjusted Sharpe ratio as AdjSh=$(\bar{x} + \mathrm{AF})/s$. Now with AF = 3, EX1 has $\mathrm{AdjSh}^{(A)} = 2 > \mathrm{AdjSh}^{(B)} = 2/10 = 0.2$, which yields the correct ranking. For EX2, note that $(\mathrm{AF} + \bar{x}^{(C)}) = 2 > (\mathrm{AF} + \bar{x}^{(D)}) = 0$. Again this yields the correct ranking.

Remark 5. The definition SharpeR $= \bar{x}/s$ has two problems. First, it is not defined if $s = 0$, and one may have to replace s by a 'small' positive number. The second problem described above is that higher values of SharpeR are not always more desirable when the comparison involves money-losing funds. The sign of SharpeR $= \bar{x}/s$ depends only on the sign of $\bar{x}$, since $s \geq 0$, by definition. Note that adding a common positive constant add factor (AF>0) to all $\bar{x}$ maintains the correct ranking of money-making funds. It is easy to verify that adjusted ratio AdjSh=$(\mathrm{AF} + \bar{x})/s$ does not have anomalous rankings provided AF exceeds the absolute value of the most negative $\bar{x}$. We define our adjusted downside SharpeR as AdjDnSh=$(\mathrm{AF} + \bar{x}_w)/\mathrm{DSD}$. Of course, the add factors should be used only if needed in the presence of negative means, and only for ranking purposes.

A popular personal finance package called Quicken has three categories of risk: small, medium and high using a combination of time horizon and s the standard deviation. Since one can have high risk-tolerance and yet a short time horizon, Quicken seems to mix up the two concepts. If firms like Quicken use DSD or WSD_α from Eq. (6) instead of s, it will improve matters. Our WSD_α can conveniently parameterize the risk measure by $\alpha \in [0,1]$, depending on the extent of compliance with EUT axioms. Thus, we claim that incorporating more realistic and meaningful indicators of agents' attitude toward risk is now feasible, and it will include important insights of economic theory.

Morningstar's star rating system from Morningstar manuals is described in Vinod and Morey (2002). Section 4 compares various ranking methods with Morningstar's *overall* star rating, which is the most popular and most advertised.

3. Stochastic dominance and risk measures

The purpose of this section is to review the insights from stochastic dominance literature, which emphasizes risk aversion, and an extension of Anderson's (1996) nonparametric measure to suit our portfolio selection problem. Although we will eventually satisfy the 4DPs by using transformed $w_t x_{(t)}$ instead of x_t, let us avoid clutter and use x in the initial parts of this section. Huang and Litzenberger (1988) and Campbell et al. (1997) are good references to this and related empirical literature.

Arrow (1971) formulated following desirable properties of utility functions. (i) MU is positive or nonsatiety with respect to wealth. (ii) Decreasing MU leads to aversion to variance, or risk aversion. (iii) Risky assets are not "inferior goods" or the utility function satisfies non-increasing (or roughly speaking, diminishing) *absolute* risk aversion (NIARA). The Arrow–Pratt absolute risk aversion function is defined as $(-u'')/u'$. The requirement for NIARA is that the derivative of any suitable risk aversion function with respect to x be non-positive. A necessary condition for this is that $u''' \geq (u'')^2/u'$.

A class of utility functions satisfying non-negative MU, $u' \geq 0$, is called D_1. Similarly, functions satisfying $u' \geq 0$ and $u'' \leq 0$ belong to D_2, which satisfies the "law of diminishing MU." Agents whose $u(x) \in D_2$ are usually called risk averse. Recall that D_3 denotes a class of utility functions satisfying $u' \geq 0$, $u'' \leq 0$ and $u''' \geq 0$, which can satisfy diminishing marginal utility and non-increasing absolute risk aversion (NIARA). A parametric cubic $u(x)$ is used in the literature to show that NIARA is more complicated and stringent requirement than checking the signs of derivatives. Unfortunately, the sufficient conditions are difficult to state, let alone verify, when we wish to keep $u(x)$ unspecified and nonparametric.

To fix ideas, consider the parametric power utility function: $u(x) = x^{1-\gamma}/(1-\gamma)$, with $\gamma > 0$. For this $u(x)$, $u' = x^{-\gamma}$, $u'' = -\gamma x^{-\gamma-1}$, where $u'' < 0$ and $u''' = \gamma(\gamma+1)x^{-\gamma-2}$. Clearly, $u''' \geq 0$ will hold for $\gamma > 0$. Since the Arrow–Pratt measure $(-u'')/u'$ defined above equals γ/x here, as x increases, γ/x decreases, implying NIARA for the power utility. The *relative* risk aversion function is defined as $(-u''x/u')$. Since $(-u''x)/u'$ equals a constant γ, power utility belongs to the constant relative risk aversion (CRRA) family. Experimental evidence in Gordon et al. (1972) suggests that a typical investor has *increasing* (not constant) relative risk aversion. Hence, the power utility and other forms of CRRA family are unacceptable. Unfortunately, the classical formulation needs a single analytic $u(x)$ satisfying NIARA for all x. We inject flexibility by splitting $u(x)$ into T component parts near each order statistic $x_{(t)}$. For example, even if relative risk aversion function is constant for $u(x)$ near one or more $x_{(t)}$ values, the constant itself need not be the same at the T order statistics.

Kimball (1990) notes that risk aversion is not quite suitable for the (investor's) choice problem. His 'prudence'$=(-u''')/u''$, measures the 'propensity to prepare or forearm oneself in the face of uncertainty'. As x increases, this propensity is expected to decrease. The 'prudence' for 'power utility' equals $(\gamma+1)/x$, which does decrease as x increases. A realistic $u(x)$ should exhibit Kimball's non-increasing prudence (KNIP).

Lemma 1. *The requirement $u'''' \leq 0$ is necessary for non-increasing "prudence."*

Proof. Non-increasing "prudence" requires $(\mathrm{d}/\mathrm{d}x)\{(-u''')/u''\} \leq 0$, that is $\{(-u'''')u'' - (-u''')(u''')/(u'')^2 \leq 0$. That is, prudence requires that: $u''''u'' \geq (u''')^2 \geq 0$. Since $u'' \leq 0$, the necessary condition for the left side $u''''u'' \geq 0$ is: $u'''' \leq 0$ as claimed by the lemma. □

The classes D_1 to D_4 are sequentially more restrictive and designed to satisfy regularity conditions for the first four orders of stochastic dominance to be discussed next. The class D_4 satisfying $u' \geq 0$, $u'' \leq 0$, $u''' \geq 0$, and $u'''' \leq 0$ is the most restrictive. Lemma 1 has shown that its special requirement $u'''' \leq 0$ is necessary for Kimball's (1990) non-increasing prudence (KNIP) to hold true.

3.1. First-order stochastic dominance (1SD)

Consider $x \in [x^*, x^U]$ a closed interval, and two mutual funds or uncertain prospects A and B, defined by their PDFs $f_a(x)$ and $f_b(x)$ with corresponding CDFs $F_a(x)$ and $F_b(x)$. Our $f_a(x)$ and $f_b(x)$ are the distributions of two competing funds (portfolios) A and B. An important question in financial economics is to compare them despite the uncertainty associated with both. Denote the difference as: $F_{ab}(x) = [F_a(x) - F_b(x)]$, defined over the combined support of $f_a(x)$ and $f_b(x)$. Fund A dominates B if

$$F_{ab}(x) \leq 0, \text{ or } F_a(x) \leq F_b(x) \text{ for all } x \in [x^*, x^U]. \tag{9}$$

One prefers fund A over fund B if $f_a(x)$ is mostly to the right-hand side of $f_b(x)$. However, evaluated at a fixed x the CDF for the distribution on the right should be *lower.* This is called first-order stochastic dominance and we write it as $A \geq_1 B$, which means that the expected utility of fund A is no smaller than the expected utility of fund B for every non-decreasing $u(x) \in D_1$ of nonsatiable individuals. Huang and Litzenberger (1988, ch. 2) have explanatory graphs and show that Eq. (9) is equivalent to $r_A = r_B + \tilde{e}$, where rates of return are denoted by subscripts A and B, and $\tilde{e}$ is a positive random variable.

3.2. Second-order stochastic dominance (2SD)

The fund A dominates fund B in 2SD, that is, $A \geq_2 B$ for the class D_2 of utility functions, if we have

$$\int_{x^*}^{x} F_{ab}(y)\mathrm{d}y \leq 0 \text{ for } x \in [x^*, x^U]. \tag{10}$$

This involves an integral of the difference between two CDFs (F_{ab}), which are already integrals of PDFs. All risk-averse nonsatiable individuals will prefer A over B.

3.3. Third-order stochastic dominance (3SD)

The fund A dominates fund B in 3SD, that is, $A \geq {}_3B$ for the class D_3 of utility functions, if in addition to Eq. (10) we have

$$\int_{x^*}^{x} \int_{x^*}^{y} F_{ab}(z)\mathrm{d}z\mathrm{d}y \leq 0 \text{ for all } x \in [x^*, x^{U}]. \tag{11}$$

3SD means that all individuals who prefer positive skewness, and who are risk averse and nonsatiable, will prefer A over B.

3.4. Fourth-order stochastic dominance (4SD)

The fund A dominates fund B in 4SD, that is, $A \geq {}_4B$ for the class D_4 utility functions if in addition to Eq. (11) we have

$$\int_{x^*}^{x} \int_{x^*}^{y} \int_{x^*}^{z} F_{ab}(w)\mathrm{d}w\mathrm{d}z\mathrm{d}y \leq 0 \text{ for all } x \in [x^*, x^{U}]. \tag{12}$$

4SD means that all individuals who have diminishing prudence, prefer positive skewness, are risk averse and nonsatiable will prefer A over B.

We stop at fourth-order since higher order conditions cannot be justified on economic grounds. Using moment generating functions, Thistle (1993) develops infinite degree stochastic dominance (∞SD) for the special case when the support of f(x), the PDF of x, is [x>0]. Since zero or negative excess returns are quite common, we find that (∞SD) is not viable or meaningful in our context.

3.5. Extension of stochastic dominance algorithms to incorporate 4DPs

This subsection discusses Anderson's (1996) numerical algorithm based on modified trapezoidal rule to convert the integrations in Eqs. (10)–(12) into matrix multiplications. Before that, we need to redefine our x to incorporate the four desirable properties of the non-EUT methods. Here, we need a numerical algorithm using the weights (Eq. (4)), with the added complication arising from having to consider two distributions for two funds. Assume that we want to compare the fund A with the fund B from past data on returns, with T_a data points for A, and T_b data points for B. Denote the *ordered* returns by $x_{(j)}$ defined over the combined set with $j=1,2,\ldots,k$, where $k=T_a+T_b$. Clearly, $x_{(1)}$ is the smallest return in the combined set and $x_{(k)}$ is the largest. In effect, we have k mutually exclusive and exhaustive class intervals with possibly variable widths d_j.

One can compute $x_{(j)}$ as cumulated d_j measured from the smallest value $x_{(1)}$. Thus, geometrically, the empirical CDF $F_a(x)$ has step increments of $1/T_a$ occurring only at specific values of $x_{(j)}$ that coincide with the values from fund A. Now our $k \times 1$ vector $\boldsymbol{p}^{Au}$ with all elements equal to $1/T_a$ represents unweighted individual probabilities for the PDF f_a. Similarly, the empirical CDF $F_b(x)$ and the vector $\boldsymbol{p}^{Bu}$ (all values $=1/T_b$) are well defined. The CDF represents the area under the PDF. A trapezoidal algorithm was used in

Vinod (1985) for integrations over PDFs. Our next task is to use Anderson's numerical trapezoidal algorithm for the integrals in Eqs. (9)–(12).

Let $\mathbf{I}_f$ denote a $k \times k$ matrix of ones along and below the main diagonal and zeros above the main diagonal. Anderson (1996) showed that the matrix multiplication $(\mathbf{I}_f \mathbf{p}^{\mathrm{A}})$ computes the $k \times 1$ vector of cumulative probabilities and provides an approximation to the empirical CDF, F_{a}. We suggest extending Anderson's algorithm to incorporate the four desirable properties (4DPs) of $u(x)$ from the prospect theory as follows. We simply work with weighted $\mathbf{p}^{\mathrm{A}}=[w_t \circ \boldsymbol{p}^{\mathrm{Au}}]$ where $\circ$ denotes an element-wise (Hadamard) product of the two vectors for fund A. Similarly we denote the weighted $\mathbf{p}^{\mathrm{B}}=[w_t \circ \boldsymbol{p}^{\mathrm{Bu}}]$ for fund B. As explained in Remark 4, the weights depend on α and k. We use a gauss language procedure based on Eq. (4) mentioned before to get the weights.

From Eq. (9), we need to compute an integral of $F_{\mathrm{ab}}(x)=F_{\mathrm{a}}(x)-F_{\mathrm{b}}(x)$ to check 1SD. Now $\mathbf{I}_f(\mathbf{p}^{\mathrm{A}}-\mathbf{p}^{\mathrm{B}})$ is a vector with k elements denoted by $1\mathrm{SD}_j$ for $j=1,2,\ldots,k$, where $k=T_{\mathrm{a}}+T_{\mathrm{b}}$. Thus, $1\mathrm{SD}_j$ represents a rich set of k evaluations to check whether first-order stochastic dominance is present at the k quantiles.

Due to random variation, $f_{\mathrm{a}}(x)$ may exceed $f_{\mathrm{b}}(x)$ in most of the relevant range but may fail for some points of the range. Do we then still prefer fund A to B? The answer can be based on $\max(1\mathrm{SD}_j)$, which is the Kolmogorov–Smirnov (K–S) maximum distance between two CDFs statistic. Even for the first order, the K–S statistic has a nonstandard sampling distribution and needs Gibbons' tables. Since the K–S tables are not available for higher orders, Larsen and Resnick (1993, 1996) propose a novel bootstrap.

When money can be invested in the two competing funds or portfolios over time, what matters is a comparison of the final net worth of the investor at the end of a time horizon. Since the K–S statistic does not approximate the final net worth, we suggest using the earlier result that Eq. (9) is equivalent to $r_{\mathrm{A}}=r_{\mathrm{B}}+\tilde{e}$. The difference in returns as it changes the net worth is approximated by the summation $\Sigma 1\mathrm{SD}j$ over the range of j. If $\Sigma 1\mathrm{SD}j$ is *negative*, we conclude that fund A dominates B in cumulative net worth of the investor. Since $x_{(j)}$ values are ordered, the corresponding utility function $u(x_{(t)})$ is well defined at these k values and *locally* belongs to the class D_1 of utility functions satisfying non-negative $u' \geq 0$, defined before. In fact, Anderson (1996, p. 1186) notes that $1\mathrm{SD}j$ for each j (viewed as a quantile) is asymptotically normal. We claim that the sum $\Sigma 1\mathrm{SDj}$ is a more efficient statistic, as its sampling variance is proportional to $1/k$. The central limit theorem for dependent data from McLeish (1974) proves asymptotic normality of the sum $\Sigma 1\mathrm{SD}j$ without requiring that each item in the sum be independent.

Next, we consider the utility class D_2 and compute the integral (Eq. (10)), where we compute the area with $x^j=x_{(j)}$ as the variable of integration. This requires integration of CDFs by the modified trapezoidal rule to accommodate unequal interval widths. Denote

$$C(x^j)=\int_0^{x^j} F(z)\mathrm{d}z \approx 0.5\left\{F(x^j)d_j+\sum_{i=1}^{j-1}(d_i+d_{i+1})F(x^i)\right\}. \tag{13}$$

Let d_j ($j=1,2,\ldots,k$). Although k will be large, we illustrate two useful matrices for $k=3$:

$$\mathbf{I}_f = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{I}_F = 0.5 \begin{bmatrix} d_1 & 0 & 0 \\ d_1 + d_2 & d_2 & 0 \\ d_1 + d_2 & d_2 + d_3 & d_3 \end{bmatrix}$$

Now the matrix multiplication $\mathbf{I}_F\mathbf{I}_f(\mathbf{p}^{\mathrm{A}} - \mathbf{p}^{\mathrm{B}})$ computes Eq. (13) for 2SD with elements 2SDj. Again, if the cumulative sum Σ2SDj is negative, we conclude that portfolio A dominates B in the utility class D_2. Since pre-multiplication by $\mathbf{I}_F$ is a linear transformation it retains asymptotic normality of each 2SDj. As before, we prefer the summation Σ2SDj statistic, since it can be proved to have lower variance i.e., is more efficient.

For the third order, we need to further integrate the $C(x^j)$ of Eq. (13) as:

$$\int_0^{x^j} C(z)\mathrm{d}z \approx 0.5\left\{ C(x^j)d_j + \sum_{i=1}^{j-1}(d_i + d_{i+1})C(x^i) \right\}. \tag{14}$$

It is easy to verify that matrix multiplication $\mathbf{I}_F\mathbf{I}_F\mathbf{I}_f(\mathbf{p}^{\mathrm{A}} - \mathbf{p}^{\mathrm{B}})$ yields elements denoted by 3SDj from the integral in Eq. (14). Again, if the cumulative sum Σ3SDj is negative, we conclude that fund A dominates B in the utility class D_3 locally satisfying $u' \geq 0$, $u'' \leq 0$ and $u''' \geq 0$. Similarly, matrix multiplication $\mathbf{I}_F\mathbf{I}_F\mathbf{I}_F\mathbf{I}_f(\mathbf{p}^{\mathrm{A}} - \mathbf{p}^{\mathrm{B}})$ yields elements denoted by 4SDj from the analogous integral for the fourth order (4SD). As before, pre-multiplication by a matrix is a linear transformation of a normal random variable. Thus, Σ3SDj and Σ4SDj are consistent, asymptotically normal and efficient test statistics.

Remark 6. The stochastic dominance literature appeals to the mathematical economists as it can involve intricate propositions. If the mutual fund A dominates B in 4SD, then the superiority of A over B does not depend on the functional form of the utility function, as long as $u(x) \in D_4$. This is a powerful statement, since we cannot generally estimate the exact form of $u(x)$. We see no reason to settle for 1SD or 2SD, since $u''' \geq 0$ and $u'''' \leq 0$ are obviously useful conditions. Recall that our four desirable properties (4DPs) listed in Remark 3 are important for realism. Hence, we suggest using weights (Eq. (4)) to modify the order statistics $x_{(t)}$ by $w_t x_{(t)}$, where weights w_t will depend on the EUT compliance parameter α. Then, our evaluations of 4SD do satisfy 4DPs. Moreover, we avoid parametric assumptions about $f(x)$ by working with empirical CDFs and trapezoidal algorithms for their integrations. If the cumulative sum $\Sigma 4\mathrm{SD}_j$ is *negative*, we conclude that the fund dominates a reference fund (S&P500 index) in a locally flexible $f(x)$ and in a theoretically realistic utility class. We use ‘tests of the difference between two large sample means’ to check if the differences between the ranked funds are statistically significant. Denoting the mean by $\bar{x}$, variance by v and sample size by n with subscripts 1 and 2 for the two samples, the test statistic from elementary textbooks is $z=(\bar{x}_1 - \bar{x}_2)/\mathrm{SE}$, where SE is the standard error, $\mathrm{SE} = \sqrt{[(v_1/n_1)+(v_2/n_2)]}$. Assuming $n>30$ for both samples, z is asymptotically $N(0,1)$. We reject the null hypothesis that the difference of two population means is zero if $|z|>1.96$ at the usual 95% confidence level.

Thus, we can recommend ranking of mutual funds based on $\Sigma 3SD_j$ and $\Sigma 4SD_j$ as a practical gist incorporating several important lessons from both standard and sophisticated recent theories in Economics, Finance and Statistics.

4. Numerical examples

In this section, we include examples to show the feasibility of our proposals. To illustrate our model, we use data from the January 2001 Morningstar Principia CD ROM Data Disk, which covers January 1991 to December 2000 period (Morningstar On-Disk Principia, 1991–2001). At the time this study was started, this was the latest set. From the data disk, we select data on all funds in Morningstar's International Fund Category that received a Morningstar *overall* star rating with stars (= Mstars) ranging from 1 to 5 described in Section 3. There are 1281 total funds of which 508 are young funds (5> years of data ≥ 3), 619 middle-aged funds (10> years of data ≥ 5), and 154 old funds (≥ 10 years of data). Since international funds tend to be relatively new, we have only 154 old funds. Of the 1281 funds, 129 funds have five stars, 306 funds have four stars, 454 funds have three stars.

Stochastic dominance concepts check if fund A dominates fund B (see Section 3). We always designate S&P500 (ticker symbol SPX) as our fund B. One by one, each of the 1281 funds becomes our fund A. We collect monthly returns from the Morningstar CD-ROM for all available months during January 1991 to December 2000. For SPX (from Yahoo) and these 1281 funds, we find their respective *excess* returns by subtracting the Tb3 and mix the two (A and B). We measure d_j as distances from the minimum of the mixed set. While the matrix $\mathbf{I}_f$ is simple, the matrix $\mathbf{I}_F$ and vectors $\mathbf{p}^A$ and $\mathbf{p}^B$ require considerable work. We need d_j to define $\mathbf{I}_F$. Next, we use matrix multiplications 1281 times to get $\Sigma i\mathrm{SD}j$ for $i = 1$ to 4.

The Sharpe ratios (SharpeR) in Morningstar's data set used here have 412 negative values, 1 fund (No. 84) is missing, 10 are zero and 858 are positive. Note that money-losing funds are not a rarity. We also omit Fund No. 84 in SharpeR context. We compute the downside Sharpe ratio (DnSha = $\bar{x}_w$/DSD). For $\alpha = 0.1$, DnSha for 359 funds are negative, none is zero and others are positive. For $\alpha = 0.5$, 456 funds have negative DnSha, none is zero and others are positive. For $\alpha = 1$, DnSha for 309 funds are negative, one is near zero and others are positive.

Now we turn to the two problems with Sharpe ratio mentioned in Remark 5. Since none of the 1281 mutual funds has a zero standard deviation, no adjustment for zero denominators is needed. However we do need the 'add factor' adjustment to solve the second problem. We have $\bar{x}_w$ the weighted downside average returns for $\alpha = 0.1$, 0.5 and 1 for each of the 1281 funds. The most negative value among these is -21.45. Hence, we choose our add factor to be: AF = 25, a nearby larger round number. For the denominator, we use downside standard deviation (DSD) leading to DnSha = $\bar{x}_w$/DSD. After add factor of 25, we define AdjDnSh=(AF + $\bar{x}_w$)/DSD. All these versions of Sharpe ratios are computed for each of the 1281 funds and the three choices of α.

It is not possible to report the results for 1281 funds in any detail. We summarize the main results for various methods of ranking 1281 funds by reporting the simple correlation coefficients among various pairs of variables denoted as Cor(.,.). Capital asset pricing model (CAPM) involves a familiar regression, whose intercept Alpha and slope Beta are commonly used by Wall Street brokers for ranking stocks and mutual funds. The Alpha and Beta of CAPM should not be confused with the α of Eq. (3) and β in Eq. (5). Note that variables (Mstars, SharpeR, Alpha, Beta) based on Morningstar's. calculations do not change with our EUT α. Hence, correlations of these four variables remain fixed and change only when they are correlated with DnSha, AdjDnSh type variables which do change with α. The correlations are split into three panels of Table 1 for $\alpha = 0.1$, 0.5 and 1.

It is intuitive that Cor(Mstars, SharpeR) = 0.848 is large, suggesting similar rankings of funds. Note that Cor(Mstars, Alpha) = 0.823 is also high. Now Cor(Mstars, Beta) = − 0.377 should be negative, since high Beta means a relatively risky fund in terms of non-diversifiable risk. However, its low magnitude may indicate that the CAPM Beta and Mstars are dissimilar ranking tools. On Wall Street, it is customary to use several tools to incorporate further individual preferences including tax efficiency, price-earning ratios, cash, inventory, gross margins, etc. The rankings proposed here can be used to add new tools supported by sophisticated theories in statistics, economics and finance.

For each of the 1281 funds, we have computed trapezoidal approximations to integrals and sums Σ1SDj , Σ2SDj, Σ3SDj, and Σ4SDj described earlier over the data period having about 120 monthly returns for $\alpha = 0.1, 0.5, 1$. Our main point is that our proposal is practical and potentially useful. It is possible to devise a set of alternatives to the Mstars based on stochastic dominance, and that they will, in general, lead to different choices of mutual funds by agents. It is interesting that Cor(Σ1SDj,Σ2SDj) = 0.35 and Cor(Σ1SDj, Σ4SDj) = 0.017 are low in the bottom panel for the perfectly EUT compliant agent (i.e. for $\alpha = 1$). However, they all become near 0.6 for $\alpha = 0.1$ or 0.5 in the other two panels. The Cor(Σ3SDj,Σ4SDj) also always exceed 0.6. Similar correlations among cumulative sums are generally lower for $\alpha = 1$ than for $\alpha = 0.1$.

The Cor(DnSha, AdjDnSh) between the downside Sharpe ratio and its adjusted version using the add factor (= 25) is only 0.22 when $\alpha = 1$ in the bottom right corner. However, it increases to 0.6 for $\alpha = 0.5$ and 0.1. This suggests that 'add factor' can make a potentially important difference to the rankings of mutual funds, especially for non-EUT compliant agents. The absolute value |Cor(AdjDnSh,ΣiSDj)| always exceeds |Cor(DnSha,ΣiSDj)| and |Cor(SharpeR,ΣiSDj)| for all $i = 1$ to 4 with one exception. The only exception in the bottom left corner when $\alpha = 1$ for Σ1SDj suggests poor behavior of Σ1SDj, as well as, superiority of AdjDnSh over DnSha and SharpeR. Thus, our focus on downside standard deviation and inclusion of the add factor are bringing Sharpe Ratio rankings closer to stochastic dominance rankings of orders 2 to 4.

Since large values of both the CAPM Beta and ΣiSDj suggest poor performance of the fund, it is gratifying that Cor(Beta,ΣiSDj) $\geq$ 0.65 for $i = 2,3,4$. Again, poor behavior when $i = 1$ suggests that Σ1SDj may not be a useful criterion. The |Cor(Beta,DnSha)| = 0.21 is smaller than |Cor(Beta,SharpeR)| = 0.27 and also smaller than |Cor(Beta,AdjDnSh)| = 0.64 for $\alpha = 1$ suggesting that the 'add factor' is bringing Sharpe ratio rankings closer to those based on Beta. However, |Cor(Beta,AdjDnSh)| and |Cor(Beta,DnSha)| get closer to each other as the agent becomes less EUT-compliant (α decreases).

Table 1
Correlation matrix for $\alpha = 0.1$, 0.5, 1 in three panels

	1SD*j*	2SD*j*	3SD*j*	4SD*j*	Mstars	SharpeR	Alpha	Beta	DnSha
α = 0.1									
1SD*j*	1								
2SD*j*	0.966	1							
3SD*j*	0.834	0.945	1						
4SD*j*	0.607	0.771	0.93	1					
Mstars	− 0.31	− 0.414	− 0.486	− 0.463	1				
SharpeR	− 0.336	− 0.436	− 0.501	− 0.474	0.848	1			
Alpha	− 0.284	− 0.391	− 0.472	− 0.465	0.823	0.886	1		
Beta	0.559	0.68	0.762	0.732	− 0.377	− 0.273	− 0.374	1	
DnSha	− 0.428	− 0.511	− 0.549	− 0.49	0.56	0.536	0.432	− 0.534	1
AdjDnSh	− 0.473	− 0.563	− 0.608	− 0.554	0.456	0.415	0.276	− 0.547	0.594
α = 0.5									
1SD*j*	1								
2SD*j*	0.959	1							
3SD*j*	0.833	0.952	1						
4SD*j*	0.582	0.746	0.891	1					
Mstars	− 0.438	− 0.498	− 0.521	− 0.433	1				
SharpeR	− 0.439	− 0.494	− 0.517	− 0.434	0.848	1			
Alpha	− 0.363	− 0.428	− 0.47	− 0.407	0.823	0.886	1		
Beta	0.62	0.737	0.782	0.679	− 0.377	− 0.273	− 0.374	1	
DnSha	− 0.571	− 0.624	− 0.616	− 0.489	0.699	0.717	0.625	− 0.556	1
AdjDnSh	− 0.692	− 0.745	− 0.715	− 0.565	0.389	0.36	0.187	− 0.64	0.635
α = 1									
1SD*j*	1								
2SD*j*	0.351	1							
3SD*j*	0.159	0.923	1						
4SD*j*	0.017	0.314	0.588	1					
Mstars	− 0.635	− 0.649	− 0.499	− 0.117	1				
SharpeR	− 0.592	− 0.557	− 0.429	− 0.109	0.848	1			
Alpha	− 0.52	− 0.423	− 0.299	− 0.052	0.823	0.886	1		
Beta	− 0.084	0.655	0.65	0.195	− 0.377	− 0.273	− 0.374	1	
DnSha	− 0.622	− 0.457	− 0.325	− 0.073	0.846	0.987	0.918	− 0.207	1
AdjDnSh	− 0.006	− 0.808	− 0.732	− 0.222	0.364	0.343	0.176	− 0.64	0.223

Symbols Σ1SDj is represented as 1SD*j*, and similarly till 4SD*j*. All these should have the prefix Σ placed before them. The CAPM Alpha and Beta are as computed by Morningstar.

Since the number of stars (Mstars) are integers 1 to 5, this variable is not particularly suited for correlation analysis. However, we note that Cor(Mstars,Alpha) = 0.82, Cor(Mstars,DnSha) = 0.85 ($\alpha = 1$), and Cor(Mstars,SharpeR) = 0.85. This suggests that for EUT compliant agents Mstars is related to the indicated variables. However, Cor(Mstars,DnSha) reduces to 0.7 when $\alpha = 0.5$, and further reduces to 0.6 when $\alpha = 0.1$ for non-EUT compliant choices. A long list of similar conclusions can be drawn by studying Table 1 of correlations among various criteria for ranking mutual funds.

Another way to summarize the results for 1281 funds is to report the best and worst set of funds. Tables 2 and 3 report the best five and worst five funds sorted by the Σ4SD*j*

Table 2

Best five funds sorted by $\Sigma 4\mathrm{SD}j$ for $\alpha = 0.1$, 0.5, 1

α	Fundno	$\Sigma 1\mathrm{SD}j$	$\Sigma 2\mathrm{SD}j$	$\Sigma 3\mathrm{SD}j$	$\Sigma 4\mathrm{SD}j$	Mstars	SharpeR	Alpha	Beta	DnSha	AdjDnSh
0.1	734	−0.609	−9.695	−84.8	−536.9	5	0.97	16.06	0.76	3.781	37.593
0.1	735	−0.609	−9.692	−84.76	−536.7	5	0.97	16.12	0.76	3.796	37.512
0.1	547	−0.596	−9.485	−83.57	−533.5	5	0.67	10.78	0.82	1.162	43.585
0.1	1046	−0.593	−9.407	−83.36	−532.6	5	0.73	19.49	0.7	4.348	37.148
0.1	1268	−0.591	−9.392	−82.91	−531.6	4	0.54	7.14	0.64	4.671	44.6
		Signif. pairs	(1:3,4,5)	(2:3,4,5)	(1,2:4, & 1,2:5)	(1,2:4)					
0.5	722	−0.272	−4.255	−33.97	−195.4	4	0.45	0.62	0.42	1.594	65.762
0.5	931	−0.27	−4.346	−34.25	−195.1	3	0.25	−1.25	0.41	0.787	68.172
0.5	161	−0.287	−4.296	−33.39	−188.9	3	−0.04	−2.8	0.37	1.807	55.513
0.5	1241	−0.304	−4.629	−34.5	−188.5	4	0.91	2.67	0.26	2.668	89.443
0.5	929	−0.262	−4.254	−33.27	−188.1	3	0.1	−2.22	0.41	0.555	68.453
1	931	0.06	−0.688	−5.608	−27.04	3	0.25	−1.25	0.41	0.433	65.015
1	161	0.052	−0.553	−5.028	−25.87	3	−0.04	−2.8	0.37	0.012	60.75
1	722	0.056	−0.504	−4.834	−25.11	4	0.45	0.62	0.42	0.757	58.123
1	929	0.07	−0.636	−5.192	−24.6	3	0.1	−2.22	0.41	0.219	65.12
1	439	0.049	−0.567	−4.886	−24.37	3	0.25	−1.17	0.46	0.451	55.096

$\Sigma 4\mathrm{SD}j$ is negative for best funds. Fundno is fund number. Statistically significant pairs are indicated for $\alpha = 0.1$ along the row marked "signif. pairs". The notation uses the colon and commas to separate the ranks. For example, (1:3,4,5) means that rank 1 fund is significantly different from funds ranked 3, 4, and 5. The 95% bootstrap-t confidence intervals for $\Sigma i\mathrm{SD}j$ when $i = 1,..,4$ are: (− 0.8603, − 0.6128), (− 19.67, − 9.442), (− 240.2, − 83.14), (− 1745 − 533.8) based on Appendix A.

Table 3
Worst five funds sorted by Σ4SDj for $\alpha = 0.1, 0.5, 1$

α	Fundno	Σ1SDj	Σ2SDj	Σ3SDj	Σ4SDj	Mstars	SharpeR	Alpha	Beta	DnSha	AdjDnSh
0.1	318	0.655	26.46	559.8	8617	2	−0.57	−20.12	1.46	−1.53	14.027
0.1	43	0.653	26.54	563.2	8779	2	−0.48	−17.47	1.45	−0.9	15.376
0.1	44	0.655	26.62	565.7	8844	2	−0.49	−17.86	1.45	−0.903	15.362
0.1	11	0.657	27.08	591.1	9375	1	−0.24	−10.98	1.57	−0.617	18.294
0.1	957	0.665	40.01	1267	28,590	1	−0.61	−26.66	2.51	−3.063	7.406
0.5	44	0.318	11.54	219.1	3220						
0.5	1249	0.318	11.82	226.9	3353	1	−0.92	−27.01	1.46	−0.419	11.43
0.5	11	0.306	12.45	250.8	3782	1	−0.24	−10.98	1.57	−0.677	12.814
0.5	177	−0.111	−0.993	85.01	9417	4	0.28	1.69	0.67	0.57	29.67
0.5	957	0.329	19.89	569.9	12,110	1	−0.61	−26.66	2.51	−1.496	6.758
1	647	0.146	4.499	64.67	764	3	0.24	21.3	1.6	0.936	9.687
1	345	0.129	4.49	70.85	1039	3	0.3	32.06	1.68	1.104	9.109
1	957	−0.033	4.912	109.6	1747	1	−0.61	−26.66	2.51	−0.269	7.196
1	230	−0.012	0.667	43.14	1846	5	0.52	55.71	0.4	1.437	10.362
1	177	0.033	1.483	115.1	10,360	4	0.28	1.69	0.67	0.598	26.355

criterion for $\alpha = 0.1, 0.5, 1$, respectively. Our third order dominance criterion is obviously distinct from Morningstar's or various Sharpe Ratios or CAPM Alpha or Beta. The information provided by our criteria depends on the EUT parameter α when weights are involved. It is interesting that only for $\alpha = 0.1$ Morningstar assigns five stars (best fund) to four out of five best funds rated by our Σ4SDj criterion.

For testing the rankings, we compare all pairs of funds among the top 14 and find that 91 are statistically significant [absolute value of the difference of means statistic (Remark 6) exceeds 1.96]. Table 2 reports the significant pairs among top five funds.

Several additional ways of sorting the results are possible. Let us focus on the problem of the small investor who accepts Morningstar's stars and wants to restrict attention to only funds having five stars. Among 129 such funds one can sort by our Σ4SDj criterion and select about 10 or 15 funds whose Σ4SDj numbers are near the best and close to each other and further sort by AdjDnSh. The results are omitted for brevity but available upon request. Clearly, the ranking by Σ4SDj has a low discriminating power, and hence the efficient set of funds is larger when sampling errors are incorporated. We suggest determining the admissible set by considering only the worst-case scenario confidence limits, say WC(Σ4SDj), for each fund. Since negative values of Σ4SDj represent dominance WC limit is the upper limit. For the SharpeR where high ratio is desirable, the WC limit is the lower limit. Appendix A develops a new maximum entropy bootstrap for computing confidence intervals for any statistic.

5. Summary and conclusions

The literature dealing with the expected utility theory (EUT), non-EUT, stochastic dominance, and related methods from economics, statistics and finance is vast. We seek to distill useful lessons for the practical problem of choice between mutual funds. Within conventional theory, Kimball (1990) suggests why a flexible $u(x)$ should exhibit non-increasing prudence (KNIP). Our Lemma 1 proves that this requires $u'''' \leq 0$, satisfied by class D_4 of $u(x)$ relevant for the fourth-order stochastic dominance (4SD). We have included short reviews of related concepts, and six remarks are identified to help focus on important issues and novelties in this paper. The recent unconventional theory (non-EUT) suggests that a typical investor's $u(x)$ should satisfy further four desirable properties (4DPs) listed in Remark 3. An important contribution of this paper is the development of new algorithms involving weights to modify 1SD to 4SD criteria to also satisfy the 4DPs.

Given data x_t on excess returns for T different time periods, we suggest arranging them in increasing order as $x_{(t)}$ for empirical cumulative distribution functions. We also suggest using Prelec's (1998) weight function on cumulative probabilities and "first difference" his weights to obtain relevant weights for $x_{(t)}$. The interpretation of α as the EUT-compliance parameter and the "first-differencing" formula (Eq. (4)) for w_t weights are new. Brokerage firms need a tool to classify clients with reference to their attitude to risk, such as extent of loss aversion, skewness preference, etc. Finance professionals can potentially use the α parameter for this purpose. Since past investment decisions may reveal an agent's implicit α, this opens exciting avenues for future research. For example, one can estimate international, racial, or gender differences in EUT-compliance.

We show that weighted versions of traditional indexes and criteria can be designed to satisfy 4DPs. For example, we mention the limitations of standard deviation (s) as a measure of risk (volatility) and propose downside standard deviation (DSD) and weighted standard deviation (WSD) to allow for 4DPs. Next, we use them as denominators of the well-known Sharpe ratio. Remark 5 explains why Sharpe ratios lead to anomalous rankings of money-losing funds and how new add factors solve the anomaly.

To illustrate the feasibility of our methods, we use Morningstar's overall star ratings and related data for 1281 International mutual funds. Our data cover a recent period having about 120 monthly returns. We use the 90-day Treasury bill rate (Tb3) as risk-free rate and compute all returns as 'excess returns' over the Tb3 rate. For brevity, we choose only three α values of 0.1, 0.5 and 1, where the last value $\alpha = 1$ is interpreted to mean a perfectly EUT-compliant investor. We use trapezoidal approximations to integrals appearing in the first- to fourth-order stochastic dominance formulas.

We suggest that cumulative sums Σ3SD*j* are better indicators in our context (as cumulative return to the investor) than the maximum of these values, max(3SD*j*), typically used in statistical literature. We first compare the 1281 funds with the SPX in binary comparisons and then use the Σ3SD*j* and Σ4SD*j* for ranking the funds. We report results of a correlation analysis of various fund-ranking methods including Mstars, SharpeR, DnSha, AdjDnSh, Σ1SD*j*, Σ2SD*j*, Σ3SD*j*, Σ4SD*j*, CAPM intercept Alpha and slope Beta. We also report the best five and worst five funds sorted by our Σ4SD*j* criteria, side by side with other ranking methods.

Further theoretical and numerical extensions of these methods are possible and needed. A practitioner would be interested in comparisons of out-of-sample performance in both bull and bear markets. This paper demonstrates that we can formally and practically remedy some of the failures of the expected utility theory by incorporating a new weighting scheme into the theory of stochastic dominance. We also establish a need for considering the fourth order stochastic dominance to satisfy Kimball's non-increasing 'prudence.' Appendix A provides a new bootstrap, which extends the range of data beyond the observed range and remains applicable to dependent time series.

Acknowledgements

Two anonymous referees made valuable comments and suggestions. Their insistence on inclusion of either a simulation or a bootstrap for inference has resulted in Appendix A. I thank Colin Rose for help with the graphic in Appendix A.

Appendix A. Maximum entropy (ME) bootstrap for time series of returns

The traditional bootstrap suffers from the following three properties, which are especially unsuitable when x_t represents financial returns in excess of a benchmark (S&P500). The new ME bootstrap is designed to avoid all three properties simultaneously.

(P1) The traditional bootstrap sample repeats some x_t values while deleting as many others. There is no reason to believe that excess returns a few basis points away from observed x_t are impossible, and do not belong in the resample.

(P2) The traditional bootstrap sample lies in the closed interval [min(x_t), max(x_t)]. The vast Finance literature dealing with value at risk (VaR) proves that investors do worry about potentially large losses from returns lower than min(x_t) in the data.

(P3) The traditional bootstrap sample shuffles x_t such that any dependence information in the sequence x_{t+1} follows x_t, which in turn follows x_{t-1}, and so forth, is lost. If we try to restore the original order to a shuffled resample, we end up with essentially the original set $\{x_t\}$, except that some dropped x_t values are replaced by the repeats of adjacent values. Hence, it is impossible to generate a large number ($J=999$) of sensibly distinct resamples, if we wish to restore the time-series properties to the traditional bootstrap.

The aim of Appendix A is to avoid these properties without introducing unnecessary arbitrariness. We follow the Appendix in Vinod (1985), who uses the maximum entropy arguments to extend the range of income data. The entropy has been widely used in both social and natural sciences as a measure of uninformativeness in a system. Let the random variable x_t take T possible values and let $f(x)$ denote its density. The entropy is defined by the mathematical expectation of Shannon information:

$$H = E(-\log f(x)) = \text{Entropy}. \tag{15}$$

We impose mass and mean-preserving constraint and select the $f(x)$ which maximizes H. Such an $f(x)$ is called the 'ME distribution' and consists of a combination of T pieces joined together. First, we sort the data in an increasing order of magnitude, denote the order statistics by $x_{(t)}$ and compute new 'intermediate points' defined as the average of successive order statistics:

$$z_t = 0.5(x_{(t)} + x_{(t+1)}) \qquad t = 1, \ldots . T-1 \tag{16}$$

The intermediate points help define our intervals defined by pairs of successive z's. For $t=2$ to $t=T-1$ we have, $I_2=(z_1, z_2), \ldots, I_{T-1}=(z_{T-2}, z_{T-1})$. The key novelty is to also include open-ended intervals $I_1(-\infty, z_1)$ and $I_T(z_{T-1}, \infty)$ at the two end points for $t=1$ and $t=T$. Thus, we have exactly T intervals, each of which contains exactly one $x_{(t)}$.

The mass-preserving constraint (on moments of order zero) says that a fraction $1/T$ of the mass of the probability distribution must lie in each interval. This is achieved by the ME bootstrap by giving an equal chance to each interval $I_1(-\infty, z_1), I_2(z_1, z_2), \ldots, I_T(z_{T-1}, \infty)$ of being included in the resample. The traditional bootstrap selects uniform pseudo-random numbers between [0,1], rounds them into a sequence of random integers in $[1,T]$ to define the shuffled selection from the set $\{x_t\}$. It gives each observation an equal chance $(1/T)$ of being included in the resample. An advantage of the ME bootstrap is that it uses the true pseudo random numbers, not rounded integers.

The mean-preserving constraint (on order 1 moments of $f(x)$) is $\Sigma x_t = \Sigma x_{(t)} = \Sigma m_t$, where m_t denote the mean of $f(x)$ within intervals I_t. This property is satisfied by the traditional bootstraps, since its $f(x)$ is a delta function with the mass $(1/T)$ at x_t in I_t. The mean-preserving requirement is a bit complicated for the ME bootstrap, since it requires that the mean m_t in the interval I_t is equal to a weighted sum of the order statistic $x_{(t)}$ with weights from the set {0.75,0.50,0.25} explained below.

Only if the $f(x)$ maximizes our 'ignorance' defined by the entropy (Eq. (15)), we can call it the ME distribution. The problem of finding such $f(x)$ was solved long ago in the statistical literature dealing with so-called characterization problems (Kagan et al., 1973). A reference for economists in a different context is Theil and Laitinen (1980). The solution states that: (i) $f(x)=(1/\theta)\exp([x-z_1]/\theta)$ with $\theta = m_1 = 0.75x_{(1)} + 0.25x_{(2)}$ in I_1. (ii) $f(x)=(1/\theta)\exp([z_{T-1} - x]/\theta)$ with $\theta = m_T = 0.25x_{(T-1)} + 0.75x_{(T)}$ in I_T. (iii) $f(x) = 1/(z_k - z_{k-1})$ is a uniform density where the mean of the uniform $(z_{k-1} + z_k)/2$ equals $m_k = 0.25x_{(k-1)} + 0.50x_{(k)} + 0.25x_{(k+1)}$ in the intervals $I_k(z_{k-1}, z_k)$ for $k=2, \ldots, T-1$. The weights satisfy the mean-preserving constraint $\Sigma x_t = \Sigma m_t$. Thus, the ME distribution avoids arbitrary choices and is completely known from the data.

For further clarity, let us consider a simple example of $T=5$ data points where $x_{(t)}=\{4,8,12,20,36\}$. From consecutive averages, we have $z_{(t)}=\{6,10,16,28\}$. At the left extreme, the ME distribution in the interval I_1 $(-\infty, 6)$ is $(1/\theta)\exp([x-6]/\theta)]$ with the sample mean $\theta = 0.75*4 + 0.25*8 = 5$. The next interval I_2 is (6,10) and the ME distribution in it is uniform with the sample mean $0.25*4 + 0.5*8 + 0.25*12 = 8$. Similarly for the intervals (10,16) and (16,28) the means of the uniform are 13 and 22, respectively. The last open ended interval is $I_T(28, \infty)$ distributed as $(1/\theta)\exp([28-x]/\theta)$ with $\theta = 0.25*20 + 0.75*36 = 32$. Now the mean preserving constraint $\Sigma x_t = 80 = \Sigma m_k = 5 + 8 + 13 + 22 + 32$, based on the weights 0.75, 0.50 and 0.25 is indeed satisfied. Thus the ME bootstrap distribution consists of T attached parts over the entire Real line $(-\infty,\infty)$ and avoids the property P2 of the traditional bootstrap.

It is fortunate that the ME density involves the exponential, since its CDF is $[\exp(-\theta x)-1]$ and inverse CDF at any probability p_r is $[\theta \ln(1-p_r)]$. Thus, pseudo-random numbers from the ME density are easy to obtain on a computer. The steps are as follows.

(1) Define a $T\times 2$ sorting matrix called $\mathbf{S}_1$ and place the observed time series x_t in the first column and the set of integers $a=\{1,2,\ldots,T\}$ in the second column of S_1.
(2) Sort the matrix $\mathbf{S}_1$ with respect to the numbers in its first column. This sort yields the order statistics $x_{(t)}$ in the first column and a vector $\boldsymbol{a}^{\mathrm{s}}$ of sorted a in the second. From $x_{(t)}$ construct z_t, m_t and I_t defined earlier.
(3) Choose a seed, create T uniform pseudo-random numbers p_{s} in the [0,1] interval, and identify the range $R_t=(t/T,(t+1)/T]$ for $t=0$ to $T-1$ wherein each p_{s} falls.
(4) Match each R_t with I_t. If $p_{\mathrm{s}}\in R_0$ or R_{T-1}, use the inverse CDF of the exponential. Otherwise, use linear interpolation and obtain a set of T values $\{x_{jt}\}$ as the j-th resample. Make sure that each mean of the uniform equals the correct mean m_t.

    Unlike properties P1 and P2 of the traditional bootstrap, our set $\{x_{jt}\}\in(-\infty,\infty)$ does not rule out values outside $[\min(x_t),\max(x_t)]$. Since we do not round p_{s} to nearby integers, it is extremely rare that any member of $\{x_{jt}\}$ equals the observed x_t. Finally, we turn to avoiding the property P3 by additional steps, which further transform $\{x_{jt}\}$ to recapture time series properties of the original series x_t.
(5) Define another $T\times 2$ sorting matrix $\mathbf{S}_2$. Reorder the T members of the $\{x_{jt}\}$ set for the j-th resample obtained in step 4 in an increasing order of magnitude and place them in column 1. Also place the sorted set $\boldsymbol{a}^{\mathrm{s}}$ of step 2 in column 2 of $\mathbf{S}_2$.
(6) Sort the $\mathbf{S}_2$ matrix with respect to the second column to restore the order $\{1,2,\ldots,T\}$ there. Denote the jointly sorted column 1 of the elements by $\{x_{\mathrm{s}jt}\}$. These represent our ME bootstrap resample, where the additional subscript s reminds us of the sorting step, which has restored the time dimension to correspond with the original data.
(7) Repeat steps 1 to 6 a large number of times for $j=1, 2,\ldots, J$ ($J=999$, say).

We conclude that we have removed all three undesirable properties of the traditional bootstrap. The bootstrap literature is vast and there are piecemeal attempts to remove one of P1 to P3. For example, the so-called 'moving blocks' bootstrap deals with P3 by assuming 'm-dependent' time series. Since no other bootstrap removes P1, P2 and P3 simultaneously, any comparison with them will not be fair to the ME bootstrap. Although interesting, a comprehensive comparison will further lengthen this paper.

We illustrate by using the top ranked mutual fund (No. 734 from 1281 funds) studied over the boom period of Jan. 1991 to Dec. 2000. I used GAUSS version 4.1 on a Pentium III, 850Mhz processor, 'Windows Me' OS machine. Replacing initial seed = 321 by seed + 1 at each iteration, the entire bootstrap took about 7 min. The seven steps above yield $J=999$ resamples and hence J values of various statistics discussed in the text. The 25th and 975th order statistics for any statistic yield the so-called naïve 95% confidence interval. The bottom of Table 2 reports the so-called bootstrap-t intervals, Mittelhammer et al. (2000, p. 727), which are known to be more reliable than the naïve intervals.

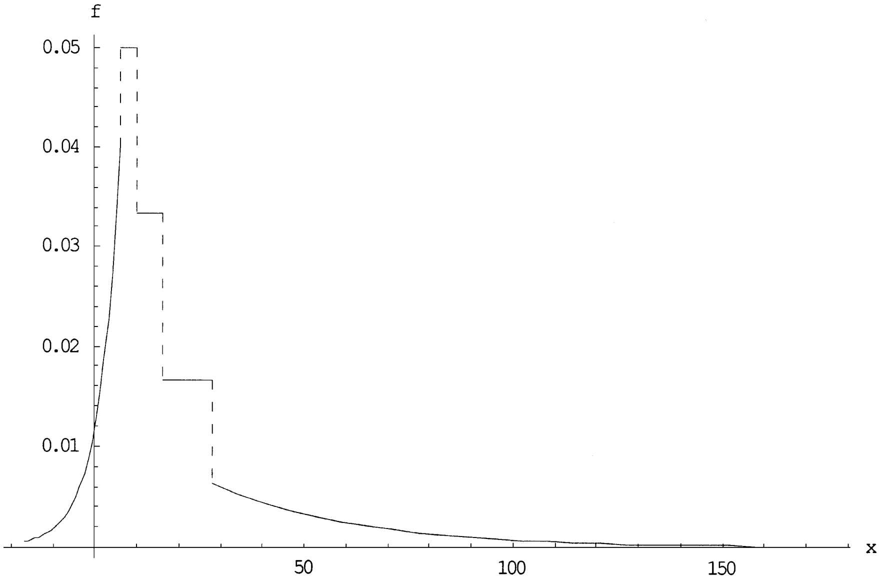

Fig. 2. Plot of the maximum entropy density when $x_{(t)}$={4 8, 12, 20, 36}.

The 95% bootstrap-*t* confidence interval is (− 1745, − 533.8) for ΣΔ4SD*j*. The worst-case here is at the upper limit, WC(Σ4SD*j*) = − 533.8. Since it remains negative, the fund 734 will dominate the S&P index fund with approximately 95% chance. Since our intervals incorporate the exponential tails (see Fig. 2), it is not surprising that our bootstrap interval is somewhat wide, although arguably more realistic. The wide interval reveals that any simple ranking of funds by 4SD has a low discriminating power. However, given adequate computing resources, we can construct bootstrap-*t* confidence limits for all 1281 mutual funds and rank them by WC(Σ4SD*j*), say. We can reject funds with positive WC(Σ4SD*j*) as worse than the S&P index fund. If the admissible set is too large, it can be further curtailed by limiting ourselves to the top 100 funds, say. Thus, allowing for sampling error need not make the admissible set unwieldy.

A detailed analysis by using the methods shown here on typical portfolios in the year 2000 would have provided a similarly wide confidence interval, warning the investors of potential losses before the market collapse of the year 2001. The new ME bootstrap methods described in this appendix are applicable elsewhere. Vinod (2003) considers an application to non-stationary time series common in econometrics and additional applications are planned. These methods deserve further critical study.

References

Ali, M.M., 1974. Stochastic ordering and kurtosis measure. Journal of the American Statistical Association 69, 543–545.

Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64 (5), 1183–1193.

Arrow, K., 1971. Essays in the Theory of Risk Bearing. Markham Publishing, Chicago.
Bawa, V.S., 1975. Optimal rules for ordering uncertain prospects. Journal of Financial Economics 2, 95–121.
Campbell, J.-Y., Lo, A.W., MacKinley, A.C., 1997. The Econometrics of Financial Markets. Princeton Univ. Press, Princeton.
Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68 (6), 1435–1464.
Gordon, M.G., Paradis, M.G., Rorke, C., 1972. Experimental evidence on alternative portfolio decision rules. American Economic Review LXII, 107–118.
Huang, C., Litzenberger, R.H., 1988. Foundations for Financial Economics. Prentice-Hall, New York.
Kagan, A.M., Linnik, Yu.V., Rao, C.R., 1973. Characterization Problems in Mathematical Statistics. Wiley, New York.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.
Kimball, M.S., 1990. Precautionary savings in the small and in the large. Econometrica 58 (1), 53–74.
Larsen, G.A., Resnick, B.G., 1993. Bootstrapping a distance test for stochastic dominance analysis. Review of Quantitative Finance and Accounting 3, 61–69.
Larsen, G.A., Resnick, B.G., 1996. Refining the bootstrap method of stochastic dominance analysis: the case of the January effect. Review of Quantitative Finance and Accounting 7, 65–79.
Lattimore, P.K., Baker, P., Witte, A.D., 1992. The influence of probability on risky choice: a parametric examination. Journal of Economic Behavior Organization 17, 377–400.
Levy, H., 1992. Stochastic dominance and expected utility: survey and analysis. Management Science 38 (4), 555–593.
McLeish, D.L., 1974. Dependent data central limit theorems and invariance principles. Annals of Probability 2 (4), 620–628.
Mittlehammer, R.C., Judge, G.G., Miller, D.J., 2000. Econometric Foundations. Cambridge Univ. Press, New York.
Morey, M.R., Vinod, H.D., 2001. A double Sharpe ratio. In: Lee, C.-F. (Ed.), Advances in Investment Analysis and Portfolio Management. JAI-Elsevier, New York, pp. 57–65.
Morningstar On-Disk, Principia, Principia-Plus, and Principia-Pro Manuals (Various Editions from 1991–2000). Morningstar Publications. Chicago.
Prelec, D., 1998. The probability weighting function. Econometrica 66 (3), 497–527.
Starmer, C., 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. Journal of Economic Literature 38, 332–382.
Theil, H., Laitinen, K., 1980. Singular moment matrices in applied econometrics. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis, vol. V. North-Holland, New York, pp. 629–649.
Thistle, P.D., 1993. Negative moments, risk aversion and stochastic dominance. Journal of Financial and Quantitative Analysis 28 (2), 301–311.
Vinod, H.D., 1985. Measurement of economic distance between blacks and whites. Journal of Business and Economic Statistics 3, 78–88.
Vinod, H.D., 2000. Foundations of multivariate inference using modern computers. Linear Algebra and Its Applications 321, 365–385.
Vinod, H.D., 2001. Non-expected utility, stochastic dominance and new summary measures for risk. Journal of Quantitative Economics 17 (1), 1–24.
Vinod, H.D., 2003. Constructive ensembles for time series in econometrics and finance. Proceedings of the Symposium on the Interface: Computing Science and Statistics, Salt Lake City. Published as a CD.
Vinod, H.D., Morey, M.R., 2000. Confidence intervals and hypothesis testing for the Sharpe and Treynor performance measures: a bootstrap approach. In: Abu-Mostafa, Y.S., LeBaron, B., Lo, A.W., Weigend, A. (Eds.), Computational Finance 1999. MIT Press, Cambridge, pp. 25–39.
Vinod, H.D., Morey, M.R., 2002. Estimation risk in mutual fund ratings: the case of Morningstar. Journal of Investing 11 (4), 67–75.