

Smoothed Safety First and the Holding of Assets

M. Ryan Haley*

Department of Economics
University of Wisconsin – Oshkosh
800 Algoma Blvd.
Oshkosh, WI 54901, USA
+1 920 424 7150 (telephone)
+1 920 424 7413 (facsimile)
haley@uwosh.edu

Harry J. Paarsch
Department of Economics
University of Melbourne

Charles H. Whiteman
Henry B. Tippie College of Business
University of Iowa

11 February 2011

Abstract

We construct a simple nonparametric portfolio selection rule by minimizing an approximation to the empirical shortfall probability using a smooth sigmoid-shaped function, and then compare the performance of our rule with the Sharpe ratio and the Portfolio Performance Index. Monte Carlo simulations, an application, and bootstrap results suggest our method is competitive with existing methods, offering portfolios with a good balance of expected return, variance, skewness, and especially small shortfall probabilities. Our rule may be of particular interest to fund managers whose primary concern is minimizing the probability of realizing a fund return below some predetermined benchmark or target rate.

Keywords: portfolio choice, shortfall probability, safety first, smoothing

JEL Classification: G11

*Corresponding author.

1 Introduction and Motivation

Solving the portfolio-selection problem has been an important focus of research in finance during the past half century, with numerous strategies having been proposed. Some particularly popular and useful methods have emerged, most notably mean-variance rules and expected utility functions. A competing family of methods involves rules that minimize, in various ways, the probability of realizing a portfolio return below some predetermined benchmark (or target) level of return. Rules derived from this objective are often referred to as *shortfall methods*.

Perhaps the earliest and most celebrated of these shortfall methods is the Safety First (SF) principle of Roy [1952], revisited by Hanssman [1968] and many others. The semi-variance methods of Markowitz [1970] and the lower partial-moment methods of Bawa [1976, 1978] likewise build on the notion of minimizing the chance of realizing a return below some predetermined benchmark. Related to these approaches is the Sortino ratio (see, for example, Sortino and Price [1994]) which is a version of the Sharpe ratio (SR) (see, for example, Sharpe [1994]) that penalizes only downside risk. Others, such as Cheng [2001], have demonstrated the usefulness of downside-risk methods in real-estate portfolio management. Still others, such as Browne [1999], have proposed dynamic shortfall-based methods that focus on beating a benchmark by a predetermined percentage. Recently, Stutzer [2000] proposed the elegant Portfolio Performance Index (PPI) rule, which is based on maximizing the decay rate of the shortfall probability.¹

The now common notion of Value-at-Risk builds on the shortfall notion by quantifying how exposed a financial position is to a given percentage loss; see, for example, Jorion [2000]. Similar to Value-at-Risk is the novel (and somewhat under utilized) Conditional Value-at-Risk (see, for example, Rockafellar and Uryasev [2000], or Krokmal et al. [2002]), a portfolio selection rule that has a coherent measure of risk relative to expected tail size, convexity, and efficiency frontiers. The advantages of Conditional Value-at-Risk are partially offset by the nondifferen-

¹Haley and Whiteman [2008] developed a closely-related rule called “Generalized Safety First” (GSF). When net asset returns are independent and identically-distributed (IID), GSF is equivalent to the PPI of Stutzer [2000]; when the logarithm of gross asset returns are IID, GSF is equivalent to the Decay Rate Maximizing Portfolio of Stutzer [2003]. Net IID returns are used below, so in this context, PPI and GSF are interchangeable terms, but the analysis can be extended to the logarithm of gross returns if desired. The interested reader is referred to these papers for complete descriptions of these methods.

tiability of the objective function, which make high-dimension problems time consuming to solve. Lim et al. [2008] have, however, suggested ways to accelerate the optimization process, but most ways to do this are rather sophisticated and may, therefore, be inaccessible to practitioners. Our approach, as we shall demonstrate below, is simple, differentiable, and can be solved very quickly—even when the number of assets is quite large—using nothing more than a basic solver, such those found in spreadsheets, like Excel.

While using the shortfall probability as the basis for portfolio selection has many proponents across various areas of risk management, it is not without its critics. For example, Bodie [1995] has highlighted potential drawbacks with the zero-one loss function implied by shortfall-based decisions: shortfall-based rules focus on the probability of falling below a benchmark, ignoring the magnitude of such a fall; in addition, shortfall-based rules fail to weigh the magnitude by which the portfolio return may exceed the benchmark. However, in the case of normal returns, when the benchmark is set equal to the risk-free rate, shortfall methods (such as SF and the PPI) recommend the same portfolio as “nonshortfall” methods (such as the SR and the familiar negative exponential utility model). Moreover, even outside the normal family, minimizing SF is equivalent to maximizing the SR. Thus, the behavioral notion of shortfall minimization is actually more common and has been more influential than it may appear.²

Other researchers (such as Fishburn [1977], Benartzi and Thaler [1995], and Rabin [1998]) have proposed behavioral foundations similar to the notion of shortfall aversion—e.g., *loss aversion*.

One additional defense of shortfall-based methods is that they are rarely based on the *true* shortfall probability; instead, proxies for the shortfall probability are used, which often map into richer decision environments than those implied by direct calculation of the true shortfall probability. In fact, Stutzer [2000] has provided a good example of this phenomenon.

In this paper, we construct a simple nonparametric portfolio selection rule by minimizing an approximation to the empirical shortfall probability where the complementary Heaviside function, within the empirical shortfall probability, is approximated by a smooth sigmoid-shaped

²In fact, Markowitz [1999] has referred to Roy’s shortfall-based SF rule as a “tremendous contribution” and wrote (p. 5): “On the basis of Markowitz (1952), I am often called the father of modern portfolio theory (MPT), but Roy (1952) can claim an equal share of this honor.”

function. In the next section, we review shortfall-based portfolio selection, and lay the ground-work for the smoothing method (SM) presented in section three. Simulations results based on mixtures of bivariate normal densities are presented in section four, while results for an higher-dimensional application are presented in section five. In section six, we summarize our work and conclude with a discussion of topics for future research.

2 Portfolio Selection: Details of Shortfall Methods

Consider a universe of N assets having returns $(R_1, R_2, \dots, R_N)$. Introduce the portfolio weights $(w_1, w_2, \dots, w_N)$, denoting this vector by $\mathbf{w}$. Define the portfolio return as

$$R(\mathbf{w}) = \sum_{n=1}^N w_n R_n.$$

The notation $R(\mathbf{w})$ makes plain the dependence of the portfolio return on the vector $\mathbf{w}$. Central to this discussion is the shortfall probability

$$\Pr[R(\mathbf{w}) \leq d]$$

where d is the benchmark rate of return. The shortfall probability is simply the area in the lower tail of the probability density function of $R(\mathbf{w})$, to the left of d in figure 1.

All portfolio selection rules focus how best to choose $\mathbf{w}$. In this case, the corresponding population decision problem is

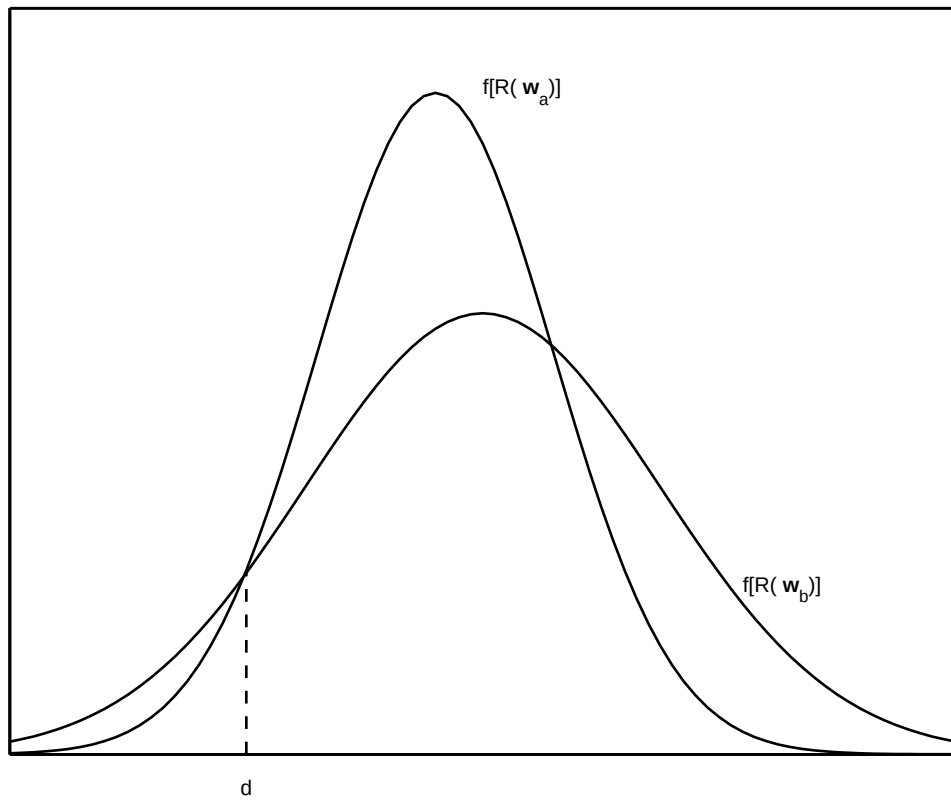
$$\min_{\mathbf{w}} \Pr[R(\mathbf{w}) \leq d] \equiv \min_{\mathbf{w}} \int_{-\infty}^d f[R(\mathbf{w})] dR(\mathbf{w}) \quad (1)$$

subject to the usual wealth exhaustion constraint.³ As in the extant literature, we assume that the joint distribution of returns is such that problem (1) has a well-defined solution.

While (1) is an intuitively appealing statement of the investor's decision problem, for a variety of reasons (such as being able to derive the probability density function $f(\cdot)$ from the joint

³The wealth exhaustion constraint, $\sum_{n=1}^N w_n = 1$, is implied, but suppressed for notational parsimony throughout the remainder of the paper. Also, short positions are permitted, but all the analysis done here can be adapted to the case of no shorting if desired.

Figure 1: Shortfall Probability*



*Different choices of $\mathbf{w}$ induce different portfolio return distributions, as illustrated here by the distinct weight vectors $\mathbf{w}_a$ and $\mathbf{w}_b$. Within the shortfall family, the goal is to find the $\mathbf{w}$ that produces the smallest probability (area) below the benchmark d .

density of returns), this problem is rarely solved directly in a portfolio selection environment. A common alternative is the corresponding finite-sample problem

$$\min_{\mathbf{w}} \frac{1}{T} \sum_{t=1}^T \mathcal{H}_{(R_t)}[R_t(\mathbf{w})] \quad (2)$$

where

$$R_t(\mathbf{w}) = \sum_{n=1}^N w_n R_{nt}.$$

Here, the complementary Heaviside function is defined as

$$\mathcal{H}_{(R_t)} = \begin{cases} 1 & \text{if } R_t(\mathbf{w}) < d, \\ \frac{1}{2} & \text{if } R_t(\mathbf{w}) = d, \\ 0 & \text{if } R_t(\mathbf{w}) > d. \end{cases}$$

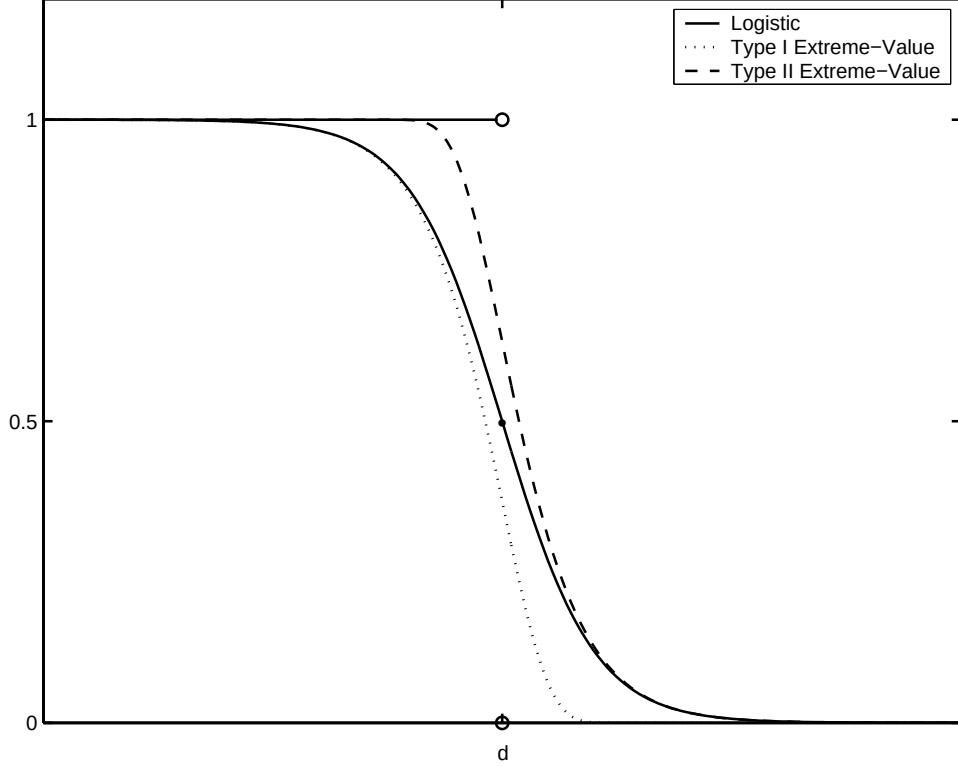
Note that the standard definition of the Heaviside function has it rising from left to right; the complementary version used here falls from left to right.

Under fairly general conditions concerning the data-generating process of returns, one can invoke some variant of the Glivenko–Cantelli theorem to argue that the solution to (2) will converge in probability to the solution to (1) as $T \rightarrow \infty$. However, solving (2) is not without complications. For example, $\mathcal{H}_{(R_t)}$ is nondifferentiable, which means that standard derivative-based optimization methods cannot be used to isolate the optimal portfolio.

A common way to circumvent this difficulty involves constructing an *upper bound* on the true shortfall probability; this strategy was used by Roy [1952] and Stutzer [2000, 2003] as well as Haley and Whiteman [2008].

Another strategy to overcome the nondifferentiability of the complementary Heaviside function (albeit a brute-force one) involves discretizing $\mathbf{w}$, forming a sample portfolio-return distribution for each discrete vector of weights, computing the shortfall probability for that sample portfolio, and then finding the one(s) having the smallest empirical shortfall probability. While this strategy is certainly feasible when N is small (such as in the allocation of tillable land to

Figure 2: Three Candidate Smoothing Functions



a small number of crops), it suffers from the curse of dimensionality when N is even moderately large, say ten. Also, uniqueness cannot be guaranteed under such a strategy. Thus, in general, discretizing $\mathbf{w}$ is an impractical strategy to solve (2).⁴ However, more sophisticated “global” algorithms, such as simulated annealing or genetic algorithms could be explored as possible ways to solve (2) effectively, but doing so is not the purpose of this paper. Our strategy is quite different: to solve a close approximation to (2) that is not NP-hard and which, therefore, circumvents the curse of dimensionality. Specifically, we proceed by replacing the complementary Heaviside function with a smooth (differentiable) approximation.

3 Smoothing Approach

Our smoothing method derives from the simple and intuitively appealing notion of approximating the sharp-cornered complementary Heaviside function that appears in (2) by a smooth

⁴In the presence of effective pruning strategies, this solution method is tractable for small N in the no-shortening case, but if shortening is permitted the problem becomes extremely computationally intensive, even for small N .

function. This intuition is illustrated in figure 2, which depicts the complementary Heaviside function and three candidate smoothing functions based on the logistic, type I extreme-value, and type II extreme-value cumulative distribution functions. By an appropriate choice of smoothing function, a researcher can recover a portfolio that closely approximates the portfolio that would have obtained by direct minimization of (2).

Defining a smoothing function is relatively straightforward because many functions have the standard Heaviside function as a limit. For example, given smoothing parameter h , both

$$\frac{1}{1 + \exp(-R_t/h)} \quad (3)$$

and

$$\exp[-\exp(-R_t/h)] \quad (4)$$

approach the Heaviside function as $h \rightarrow 0$. After experimenting with numerous candidates, we found that different smoothing function weigh moments of the return process differently, which in turn influences their finite sample behavior. The smoothing function we settled on derives from (4), but must be modified to fit $\mathcal{H}_{(R_t)}$. The reader will note that (3) is the logistic cumulative distribution function, while (4) is a transformation of the extreme-value (Gumbel) cumulative distribution function. Note, too, that, except for h , both (3) and (4) correspond to “parameter-less” probability laws, which simplifies the derivation and presentation of the smoothing approach.⁵

In econometrics, a standard practice is to use asymptotic results as approximations in finite-sampled applications. Here, we do the reverse: we stop short of the asymptotic limit (the complementary Heaviside function) to avoid nondifferentiability. Thus, a trade-off exists: the limiting result is exact, but computationally impractical, yet stopping short of that limit yields an imperfect representation of the shortfall objective in (2). As with SF and the PPI, our’s is an approximation to the true probability of shortfall, just not an upper-bound approximation.

⁵We experimented with other sigmoid and sigmoid-like functions (such as Extreme type I and II as well as the logistic, among others). We studied them using simulation methods and via Taylor-series expansions. The latter were particularly helpful in determining how different methods value even- and odd-order moments of the return distribution. After all these various background experiments, the most viable was the Extreme Type I. Of course, an infinite number of sigmoid or sigmoid-like functions exist from which we could select a smoothing function. Perhaps other good choices will be uncovered by future research.

How should one balance this trade-off? We apply a bandwidth selection rule used in the kernel-smoothed density estimation literature in statistics. While *ad hoc*, this rule serves, in our analysis, as a practical way to compromise between the exact (but nondifferentiable) limiting result and the need for smoothness. With a suitably smooth approximation, standard derivative-based optimization methods (e.g., quasi-Newton methods such as BFGS, DFP, or SR1) can then be used to isolate the optimal portfolio.

The smoothing method developed below uses (4) to approximate the complementary Heaviside function in (2). To begin, consider the cumulative distribution function of the type I extreme-value law:

$$1 - \exp\{-\exp[R(\mathbf{w})]\}.$$

To implement our method, three modifications are necessary: first, the relevant portfolio return is not $R(\mathbf{w})$, but rather the excess return $[R(\mathbf{w}) - d]$; i.e., portfolio returns net of the benchmark d . Thus, we consider

$$1 - \exp\{-\exp[R(\mathbf{w}) - d]\}.$$

Next, we want the approximating function to turn off or on quickly. To control this, we scale by the smoothing parameter h , so

$$1 - \exp\left\{-\exp\left[\frac{R(\mathbf{w}) - d}{h}\right]\right\}.$$

Finally, the approximation must fall from left to right. Adapting the above expression to mimic $\mathcal{H}_{(R_t)}$ and then embedding it into the portfolio-selection rule yields the following problem:

$$\min_{\mathbf{w}} \frac{1}{T} \sum_{t=1}^T \exp\left\{-\exp\left[\frac{R_t(\mathbf{w}) - d}{h(\mathbf{w})}\right]\right\}. \quad (5)$$

Observe that far to the left of $[R_t(\mathbf{w}) - d]$, the function $\exp[-\exp(\cdot)]$ is near one, while far to the right, the function is near zero. Note, too, how the smoothing parameter h is a function of the portfolio weights $\mathbf{w}$; we discuss this below.⁶

⁶Note that the smoothing function is only a device used to compute the optimal linear combination $\mathbf{w}$; it is *not* a distributional assumption concerning the portfolio returns. The shape of the portfolio return distribution follows from the properties of the joint distribution of asset returns.

3.1 Smoothing Parameter h

The purpose of h is to control the approximation to $\mathcal{H}_{(R_t)}$ in a way that balances two competing concerns: first, h must be small enough so the smoothing function mimics $\mathcal{H}_{(R_t)}$; but, second, h must be large enough to preserve the smoothness of the approximation to $\mathcal{H}_{(R_t)}$. Because h scales the portfolio returns, it correspondingly affects how abruptly the smoothing function falls from left to right. The smoothing functions depicted in figure 2 were plotted using an h of 0.1 so they could be easily visually distinguished from $\mathcal{H}_{(R_t)}$. In practice, h will be smaller than 0.1.

Making h too small, however, can mean that a computer will be unable to represent the numbers within the sum in (5). For example, in the data used by Stutzer [2000], a typical monthly return is around 0.01. For an h of 0.01 and a zero benchmark, a typical element in the SM would be $\exp[-\exp(0.01/0.01)]$ or 0.0660, while a smaller value of h , such as 0.002, would result in a typical element equal to $\exp[-\exp(0.01/0.002)]$ or 3.5074×10^{-65} . Any choice much smaller than 0.002 (e.g., 0.001) will result in machine underflow; i.e., a computer will represent the number as a zero, which can cause the optimization algorithm to fail.⁷ Thus, h must be set small enough so the complementary Heaviside function is well-approximated, but not so small that numerical difficulties arise. While the theory is informed by mathematics, the practice is art.

In some ways, choosing h is similar to choosing the bandwidth parameter in kernel-smoothed density estimation: if h is chosen too large, then the objective function (5) will not represent (2) well. Similarly, if h is chosen too small, then the smoothing will not “improve” the behavior of the approximation. Thus, the choice of h involves the practical, computational issues outlined above.

As an *ad hoc* rule, we embed $\mathbf{w}$ in the simple bandwidth selection formula

$$h(\mathbf{w}) = 1.06\sigma(\mathbf{w})T^{-0.2}$$

⁷Even if underflows do not occur, values on the order of 10^{-65} will contain substantial rounding error, and this can undermine the reliability of optimization algorithms.

from Silverman [1986]. This rule proved quite useful in balancing the two concerns in practice; that is, it delivered fast, reliable convergence and selected portfolios with very small shortfall probabilities and moments consistent with what the Taylor-series expansion of (5) would predict. Here, $\sigma(\mathbf{w})$ is the standard deviation of the portfolio return distribution induced by $\mathbf{w}$; i.e.,

$$\sigma(\mathbf{w}) = \sqrt{\frac{1}{T-1} \sum_{t=1}^T \left[R_t(\mathbf{w}) - \frac{1}{T} \sum_{t=1}^T R_t(\mathbf{w}) \right]^2}$$

where

$$R_t(\mathbf{w}) = \sum_{n=1}^N w_n R_{nt}.$$

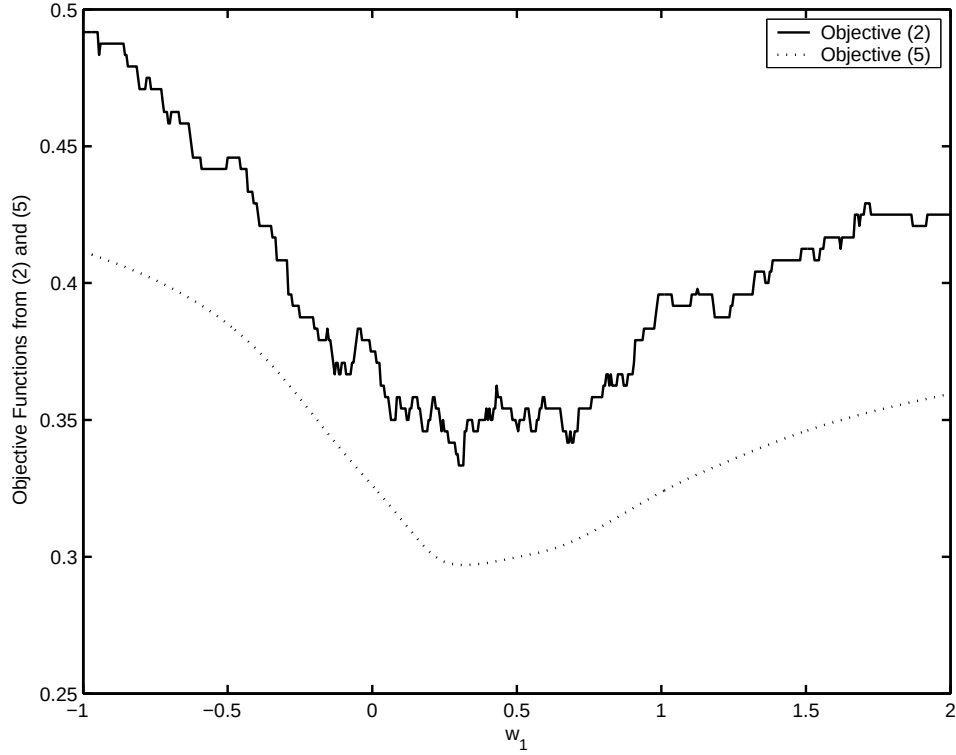
With the dependence of h on $\mathbf{w}$ now explicit, we can depict the shape of objective functions (2) and (5) by plotting them against $\mathbf{w}$ for the two-asset case; see figure 3. Note how the smoothing function tracks the shape of the actual shortfall objective function, but in a way that results in a well-defined minimization problem. The intuition governing the two-asset case applies to the N -asset case.

3.2 Expected Utility Representation

While (5) is a smooth approximation to (2), it also has an expected utility representation. Specifically, the objective function in (5) is the sample analog to the negative of the expected utility function $E\{-\exp[-\exp(\alpha r)]\}$, where r is the excess return (net of d) and $\alpha \equiv 1/h$. The interested reader can explore this connection in more detail using a Taylor-series expansion of $-\exp[-\exp(\alpha r)]$. While this expansion offers some insight into how SM values various moments of the portfolio returns, the comparison is incomplete because the shortfall probability is, of course, not tied to any particular moment. Note, too, that the SM does not require an exogenous utility parameter; i.e., for (5), α is, in a sense, endogenous to the computational strategy because of its dependence on h .⁸ Given this expected utility representation, the risk

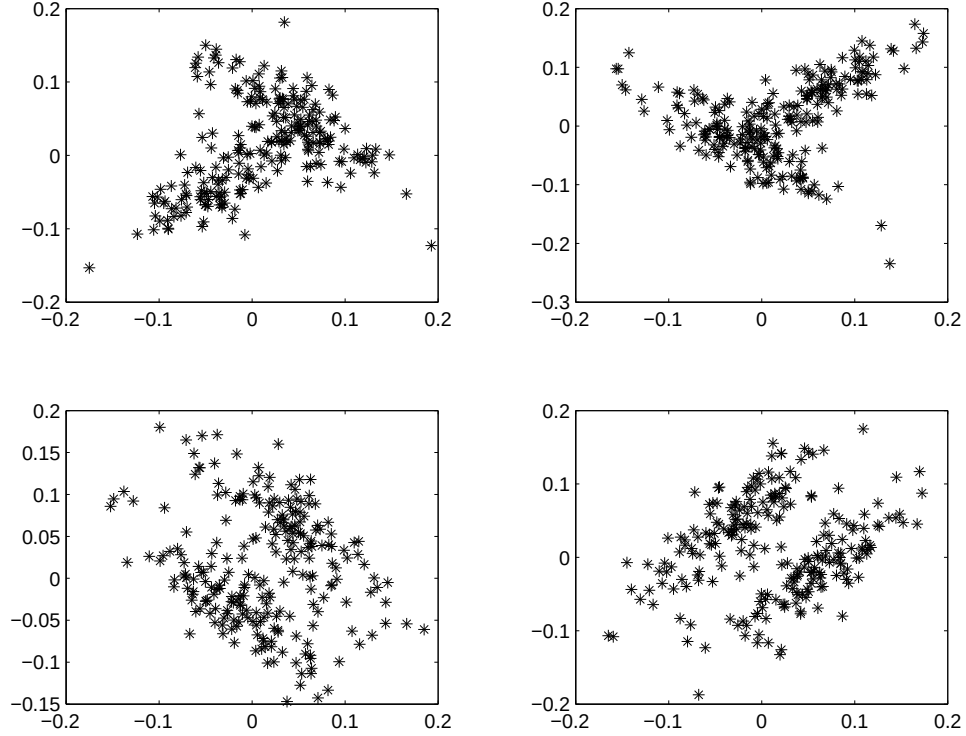
⁸Stutzer [2000, 2003] emphasized another endogenous utility approach. In the former paper, he drew a connection to the familiar negative exponential utility model, while in the latter paper he described an endogenous version of an expected utility function similar to standard power utility.

Figure 3: Objective Functions from (2) and (5) for the Two-Asset Case*



*This illustration uses 240 observations on the stocks AT and ORU, which comprise a subset of the dataset used by Stutzer [2000]; the monthly data span January 1977 to December 1996. The smoothing function isolates the same optimal weight on stock one as does the actual, discretized shortfall probability. The same relationship between these two objectives holds for the N -asset case. Values for h ranged (as $\mathbf{w}$ varied) from 0.0138 to 0.0409—sensible sizes for balancing the countervailing concerns over smoothness and approximation accuracy. The optimal weight was 0.310 with $d = 0.001$ and $h = 0.0138$. Note that portfolio selection by minimizing (1), while feasible when the N is small, is very computationally intensive (and possibly intractable) for real-world portfolio selection where N could reasonably be 100 or more.

Figure 4: Two-Dimensional Plots of the Four Representative Mixture Cases*



*Sample size of 240 randomly generated returns.

characteristics of the SM can be studied using the standard Arrow–Pratt measures of absolute and relative risk aversion. While this connection to expected utility is interesting, we prefer to focus on the simple and intuitively appealing shortfall-based interpretation of the SM. As Roy [1952, p. 433] so eloquently wrote:

In calling in a utility function to our aid, an appearance of generality is achieved at the cost of a loss of practical significance and applicability in our results. A man who seeks advice about his actions will not be grateful for the suggestion that he maximize expected utility.

4 Simulation Comparisons

A straightforward, but sometimes forgotten, feature of the SR is that it is equivalent to SF when the joint distribution of returns is multivariate normal.⁹ Stutzer [2000, 2003] demonstrated that the PPI reduces to the SR rule under normality. Haley and Whiteman [2008] summarized

⁹This is true more generally as well.

Table 1: Mixture of Normals Portfolio Performance Comparisons*

Statistic	SR	PPI	SM		SR	PPI	SM
	$d = -0.05$				$d = -0.02$		
Case 1							
Mean	0.0129	0.0127	0.0127		0.0131	0.0129	0.0139
St.Dev.	0.0539	0.0539	0.0539		0.0541	0.0539	0.0585
Skewness	0.8637	0.8866	0.8875		0.8332	0.8757	0.4664
$\Pr \leq d$	0.0476	0.0426	0.0421		0.3398	0.3488	0.3119
Case 2							
Mean	0.0127	0.0128	0.0128		0.0129	0.0130	0.0128
St.Dev.	0.0538	0.0542	0.0562		0.0540	0.0545	0.0547
Skewness	-0.8871	-0.8457	-0.6322		-0.8728	-0.8341	-0.7970
$\Pr \leq d$	0.1572	0.1571	0.1537		0.2678	0.2677	0.2657
Case 3							
Mean	0.0127	0.0126	0.0126		0.0128	0.0127	0.0133
St.Dev.	0.0407	0.0408	0.0407		0.0408	0.0408	0.0446
Skewness	0.0032	0.0077	0.0074		0.0032	0.0062	-0.0019
$\Pr \leq d$	0.0292	0.0284	0.0276		0.3057	0.3076	0.2777
Case 4							
Mean	0.0126	0.0126	0.0126		0.0127	0.0127	0.0127
St.Dev.	0.0521	0.0522	0.0523		0.0522	0.0522	0.0525
Skewness	0.0048	0.0181	0.0095		0.0050	0.0121	-0.0040
$\Pr \leq d$	0.1148	0.1155	0.1131		0.2649	0.2655	0.2615

*A sample size of 240 was chosen to mimic 20 years of monthly data; we used 500 Monte Carlo replications. The cases correspond to the four types of mixtures depicted in the corresponding quadrants of figure 4. All values are averages across the 500 samples. The mixing parameter for all results reported in this table was 0.5, but similar results obtained with different values as well.

these relationships and emphasized one additional inter-relationship: under normality, all of these methods are not only equivalent to one another, but they also minimize the *actual* short-fall probability in (1) and (2). Furthermore, because SM is very close to (2), it should (and does) mimic all of these existing methods under normality. Thus, if the joint distribution of asset returns appears to follow the multivariate normal, then all of these methods will directly minimize the shortfall probability. Hence, studying the normal case is unnecessary. Instead, using data generated from a mixture of normals, we explored four finite-sample cases involving modest departures from normality.

A parametric mixture model is a common and flexible way to characterize the multivariate distribution of asset returns, admitting features such as skewness and thick tails. We created

finite-sample datasets using two bivariate normal distributions and a mixing parameter, and then applied the SR, PPI, and SM. While we experimented with different mixing parameters, different levels for d , and so forth, it is impossible to report all of these findings. We report the results from the four main cases, as depicted in figure 4. The results, expressed as means of the 500 replications for two different values for d , are presented in table 1. Specifically, for each of the 500 replications, all the methods were computed for each index.¹⁰ The pseudo-code of the process is as follows:

```

for case = 1:4
    for d = -0.05, -0.02
        for repetition = 1:500
            create dataset
            solve SR, PPI, SM
            record moments for SR, PPI, SM
            clear dataset
        end
    end
end
end

```

The results are consistent with our expectations. In *all* cases, the SM delivered the smallest average shortfall probability while simultaneously delivering reasonable values for mean, standard deviation, and skewness. In some instances, the SM had noticeably smaller shortfall probabilities, in other cases its advantages were quite small. Regarding the standard moments, the SM generally delivered means similar to or perhaps slightly larger than its competitors, although its standard deviations tended to be slightly larger as well; with some regularity, the SM beat the SR and PPI in skewness. Over all, the SM appears to be a competitive method of portfolio selection, one that retains a focus on the shortfall probability even outside of the normal case—a feature that may be attractive to some investors.

¹⁰We used the MATLAB optimization routine `fminunc`, using a quasi-Newton method with a line search subroutine and numerical derivatives. If the reader has a strong preference for analytic first and second derivatives, then they can be derived, but they are cumbersome to manage and return similar results when used in the optimization process. Of course, `fmincon` could be used as well.

Table 2: Descriptive Statistics for Monthly Returns*

Ticker	Minimum	Maximum	Mean	St.Dev	Skewness
ACK	-0.294	0.330	0.014	0.087	0.087
ALK	-0.323	0.519	0.015	0.105	0.557
AMD	-0.477	0.487	0.024	0.163	0.261
APD	-0.306	0.215	0.013	0.076	-0.021
AT	-0.177	0.171	0.016	0.060	-0.216
BAX	-0.162	0.203	0.011	0.073	-0.029
BK	-0.250	0.287	0.018	0.082	0.154
BMJ	-0.139	0.177	0.015	0.058	0.056
BUD	-0.155	0.258	0.017	0.063	0.324
CC	-0.437	0.415	0.035	0.126	0.115
CPL	-0.137	0.188	0.013	0.050	0.011
CRS	-0.309	0.273	0.013	0.078	-0.127
DCN	-0.256	0.333	0.012	0.080	0.169
DG	-0.265	0.347	0.028	0.121	0.376
GD	-0.277	0.340	0.019	0.091	0.360
KMB	-0.146	0.322	0.017	0.062	0.866
KU	-0.107	0.139	0.012	0.040	-0.140
LTD	-0.431	0.409	0.026	0.129	0.072
MUR	-0.260	0.274	0.015	0.090	0.055
NCC	-0.244	0.216	0.014	0.064	-0.161
ORU	-0.102	0.235	0.011	0.042	0.494

*Sample size of 240.

5 Higher-Dimensional Application

In this section, we report results from applying the SM in a larger-scale application using the data from Haley and Whiteman [2008]. Descriptive statistics concerning these data are presented in table 2. An example of a mean-constrained version of the SM is also presented. Because no risk-free asset was used, following Haley and Whiteman [2008], we set d equal to -0.015 .

In table 3, we report the optimal portfolio weights and moments of the corresponding optimal portfolio return distributions for the SR, PPI, and SM selection rules. The SM portfolio exhibits a somewhat higher mean, a slightly higher standard deviation, more negative skewness, and a lower shortfall probability than the SR and PPI portfolios, but the latter is markedly less negatively skewed than any of its competitors.¹¹ Despite these differences, all the methods

¹¹This last finding is not surprising: preference for positively skewed portfolio return distributions (and, thus, an aversion to negatively skewed portfolio return distributions) is a prominent feature of the PPI method; see Stutzer [2000].

Table 3: Optimal Portfolios: $d = -0.015^*$

Ticker	SR	PPI	SM
ACK	-0.1282	-0.1275	-0.0944
ALK	-0.0067	-0.0313	0.0143
AMD	0.0059	-0.0115	0.0219
APD	0.0170	0.0034	-0.0380
AT	0.0521	0.0017	0.0543
BAX	-0.0998	-0.0729	-0.1167
BK	-0.0182	0.0241	-0.0371
BMV	0.1488	0.1340	0.1169
BUD	0.1283	0.1325	0.1752
CC	0.0505	0.0325	0.0895
CPL	0.0259	0.0458	0.0531
CRS	0.1229	0.1026	0.1571
DCN	-0.0200	-0.0403	-0.0567
DG	0.0276	0.0267	0.0265
GD	0.0214	0.0252	0.0091
KMB	0.0992	0.1874	0.0876
KU	0.4076	0.3958	0.2927
LTD	-0.0663	-0.0575	-0.0453
MUR	0.0501	0.0670	0.0833
NCC	0.0981	0.0647	0.1206
ORU	0.0839	0.0976	0.0862
Portfolio Statistics			
Mean	0.0149	0.0148	0.0164
St.Dev.	0.0312	0.0321	0.0344
Skewness	-0.4353	-0.0706	-0.5697
$\Pr \leq d$	0.1375	0.1458	0.1250
h	—	—	0.0171

*Sample size of 240.

generally agree (with some exceptions) on which stocks to sell short and which to buy long.

5.1 Moment Constraints

The SM outlined above can be easily adapted to incorporate moments of particular concern to the investor. To demonstrate this, we implemented a mean-constrained version of the SM, where we sought to proxy the desire of an investor interested in minimizing the probability of realizing sub- d returns, while seeking a particular level of (expected) return. To accommodate such a desire, we simply augmented the objective (5) by the moment constraint.¹² In table 4,

¹²In this instance, we used MATLAB's `fmincon` routine, though the older `constr` routine also works.

Table 4: Mean-Constrained Optimal Portfolios: $d = -0.015$ and $\bar{R} = 0.025$ *

Variable	SR	PPI	SM
St.Dev.	0.0539	0.0547	0.0589
Skewness	-0.2444	-0.0352	-0.4695
$\Pr \leq d$	0.2292	0.2458	0.1542
h	—	—	0.0171

*Sample size of 240.

we report descriptive statistics of the optimal SR, PPI, and SM portfolios when the portfolio mean, $\bar{R}$, was constrained to equal 0.025.

The results obtained in the unconstrained portfolio return distributions above were obtained here, too. The SM portfolio exhibits a slightly higher standard deviation and mean and somewhat more negative skewness than the SR and PPI portfolios. The differences in shortfall probabilities, however, appear to be nontrivial; e.g., the SR and PPI portfolios have nearly eight percent and ten percent higher chances, respectively, of realizing a sub- d return than does the SM portfolio! (Recall that all portfolios are constrained to deliver the same expected return.) This is a noteworthy result given that the PPI is also a shortfall method, one designed to minimize an upper bound on the shortfall probability.

5.2 Bootstrap Estimates of Sample Variability

Because shortfall may be of particular interest to fund managers whose success is a function of their ability to avoid sub- d returns, sampling errors are potentially important. Therefore, we explored the sampling properties of the various methods using the bootstrap. The financial data summarized in table 2 were resampled 1000 times. For each bootstrap sample, all three portfolio-selection rules were implemented using d of -0.015 and $\bar{R}$ of 0.025. Summary statistics of the bootstrap study are reported in table 5.¹³

As expected, the SM returned shortfall probabilities superior to (i.e., smaller than) those produced by the SR or the PPI. All methods delivered remarkably similar average standard deviations. Also, as expected, the PPI produced portfolios with the highest levels of aver-

¹³Similar results were obtained using other reasonable values for d and $\bar{R}$.

Table 5: Bootstrap Portfolio Summary Results: $d = -0.015$ and $\bar{R} = 0.025^*$

Statistic	SR	PPI	SM
St.Dev.			
Minimum	0.0292	0.0313	0.0308
Maximum	0.0545	0.0688	0.0673
Mean	0.0415	0.0433	0.0468
St.Dev.	0.0046	0.0044	0.0064
Skewness			
Minimum	-0.9976	-0.4819	-1.4727
Maximum	0.5658	1.4303	0.9772
Mean	-0.2287	0.2447	-0.2615
St.Dev.	0.2309	0.2529	0.3722
$\Pr \leq d$			
Minimum	0.0625	0.0792	0.0333
Maximum	0.2625	0.2917	0.1583
Mean	0.1479	0.1691	0.0983
St.Dev.	0.0319	0.0323	0.0209

*Sample size of 240; 1000 bootstrap replications.

age skewness; the average skewness of the SR and SM portfolios compared quite favorably (-0.2287 versus -0.2615), but the latter offered noticeably smaller shortfall probabilities, on average, 0.0983 versus 0.1479 .

6 Summary and Conclusions

We have constructed a simple nonparametric portfolio selection rule based on the time-honored and intuitive notion of minimizing the probability of realizing a return below some predetermined benchmark. We accomplished this by approximating the complementary Heaviside function embedded in the portfolio selection problem by a smooth function, hence sidestepping the inherent difficulty of working with the true population and in-sample shortfall probabilities, equations (1) and (2), respectively.

We analyzed the sampling properties of our new method using Monte Carlo simulations based on a bivariate mixtures of normals and bootstrap methods applied to an higher-dimensional dataset. The results suggest that our method is competitive with the famous Sharpe ratio and the elegant, shortfall-based Portfolio Performance; noteworthy, however, is that our method

delivers substantially lower shortfall probabilities (in some cases as much as 10 percentage points lower, all else held constant) than its competitors without sacrificing much in terms of mean, variance, and skewness.

In summary, our approach appears useful for portfolio selection, especially for investors or fund managers who are concerned about realizing a return below a benchmark. Given the now industry-wide practice of benchmarking a portfolio's performance (e.g., to a Lipper average or one of the many indices), we are hopeful that our approach will find favor among portfolio managers. In addition, our approach builds on the time-honored shortfall minimization objective, is easy to understand, simple to code, quick to converge (even using a basic Excel solver), does not rely on any intrusive distributional assumptions or a formal and parameter laden expected utility structure, and can easily incorporate moment restrictions and no-shorting if so desired.

We investigated a number of additional smoothing-based rules, some of which are depicted in figure 2. However, based on series expansions and simulation exercises, these functions proved inferior to the extreme-value-based approach. Therefore, a particularly interesting avenue for future research would involve analyzing other candidate smoothing functions. Similarly, because h serves a somewhat different purpose here than it does in kernel-smoothed density estimation, further investigation into its determination would be interesting as well.

Acknowledgements

We are grateful to two anonymous referees as well as Kenneth L. Judd for thoughtful comments and useful suggestions. M. Ryan Haley would also like to acknowledge helpful conversations with M. Kevin McGee and the support of the University of Wisconsin Faculty Development Program. Paarsch gratefully acknowledges that some work on the paper was completed while he was a visiting fellow at the Collegio Carlo Alberto in Moncalieri, Italy.

References

- Vijay S. Bawa. Admissible portfolios for all individuals. *Journal of Finance*, 31:1169–1183, 1976.
- Vijay S. Bawa. Safety-first, stochastic dominance, and optimal portfolio choice. *Journal of Financial and Quantitative Analysis*, 13:255–271, 1978.
- Shlomo Benartzi and Richard H. Thaler. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, 110:73–92, 1995.
- Zvi Bodie. On the risks of stocks in the long run. *Financial Analysts Journal*, 51:18–22, 1995.
- Sid Browne. The risk and rewards of minimizing shortfall probability. *Journal of Portfolio Management*, Summer:76–85, 1999.
- Ping Cheng. Comparing downside-risk and mean-variance analysis using bootstrap simulation. *Journal of Real Estate Portfolio Management*, 7:225–238, 2001.
- Peter C. Fishburn. Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, 67:116–126, 1977.
- M. Ryan Haley and Charles H. Whiteman. Generalized safety first and a new twist on portfolio performance. *Econometric Reviews*, 27:457–483, 2008.
- Fred Hanssmann. Probability of survival as an investment criterion. *Management Science*, 15: 33–48, 1968.
- Phillipe Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*, Second Edition. New York: McGraw-Hill, 2000.
- Pavlo Krokhmal, Jonas Palmquist, and Stanislav P. Uryasev. Portfolio optimization with Conditional Value-at-Risk objective and constraints. *Journal of Risk*, 4:11–27, 2002.
- Churlzu Lim, Hanif D. Sherali, and Stanislav P. Uryasev. Portfolio optimization by minimizing Conditional Value-at-Risk via nondifferentiable optimization. *Computational Optimization and Applications*, 46:391–415, 2008.

- Harry M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*, Second Edition. New York: Yale University Press, 1970.
- Harry M. Markowitz. The early history of portfolio theory: 1600–1960. *Financial Analysts Journal*, 55:5–15, 1999.
- Matthew Rabin. Psychology and economics. *Journal of Economic Literature*, 36:11–46, 1998.
- R. Tyrrell Rockafellar and Stanislav P. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.
- Arthur D. Roy. Safety first and the holding of assets. *Econometrica*, 20:431–449, 1952.
- William F. Sharpe. The Sharpe ratio. *Journal of Portfolio Management*, Fall:49–58, 1994.
- Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986.
- Frank A. Sortino and Lee N. Price. Performance measurement in a downside risk framework. *Journal of Investing*, 3:59–65, 1994.
- Michael J. Stutzer. A portfolio performance index. *Financial Analysts Journal*, 56:52–61, 2000.
- Michael J. Stutzer. Portfolio choice with endogenous utility: A large deviations approach. *Journal of Econometrics*, 116:365–386, 2003.