

Performance and Risk Aversion of Funds with Benchmarks:
A Large Deviations Approach ¹

F. Douglas Foster and Michael Stutzer

Professor of Finance, University of New South Wales, Sydney, AU.

and

Professor of Finance and Director, Burrige Center for Securities Analysis and Valuation
University of Colorado, Boulder, CO.

ABSTRACT

This paper provides a simple method of ranking mutual funds' probabilities of outperforming a benchmark portfolio. We show that ranking fund performance in this way is identical to ranking each fund's portfolio with a generalized entropy, equivalent to an expected generalized power utility index that uses a risk aversion coefficient specific to that fund. When the return differential between fund and benchmark portfolio (log gross) returns follows a Gaussian process, this ranking is equivalent to using a simple modification of the Information Ratio (1998). We develop and apply feasible parametric and nonparametric estimators to rank the performance of the small fraction of mutual funds that (from the results of an hypothesis test) could outperform the S&P 500 index in the long run, and to estimate the fund-specific risk aversion coefficients required for the ranking. We also argue that an auxiliary hypothesis that fund *managers* attempt to maximize the outperformance probability is no less plausible than an extant alternative behavioral hypothesis, and is more parsimoniously parametrized.

1 Introduction

As noted by Roll [23, p.13], “ Today’s professional money manager is often judged by total return performance relative to a prespecified benchmark, usually a broadly diversified index of assets.” He argues that “This is a sensible approach because the sponsor’s most direct alternative to an active manager is an index fund matching the benchmark.” A typical example, of more than just professional interest to academic readers, is the following statement by the TIAA-CREF Trust Company:

Different accounts have different benchmarks based on the client’s overall objectives...Accounts for clients who have growth objectives with an emphasis on equities will be benchmarked heavily toward the appropriate equity index – typically the S&P 500 index – whereas an account for a client whose main objective is income and safety of principal will be measured against a more balanced weighting of the S&P 500 and the Lehman Corporate/Government Bond Index.[29, p.3]

How should plan sponsors and the investors they represent evaluate the performance of a fund like this? William Sharpe [25, p.32] asserts that

The key information an investor needs to evaluate a mutual fund is (i) the fund’s likely future exposures to movements in major asset classes, (ii) the likely added (or subtracted) return over and above a benchmark with similar exposures, and (iii) the likely risk vis-à-vis the benchmark.

Procedures for implementing this recommendation will differ, depending on the quantitative framework used for measuring “return over and above a benchmark” in (ii) and “risk vis-á-vis the benchmark” in (iii). Let $R_p - R_b$ denote a portfolio p ’s “return over and above a benchmark”, b . The natural generalization of mean-variance efficiency *relative to a benchmark the investor wants to beat* is Roll’s [23] Tracking Error Variance (TEV)-efficiency, resulting from minimization of the tracking error variance $Var[R_p - R_b]$ subject to a constraint on the desired size of $E[R_p - R_b] > 0$. The tracking error variance measures the “risk vis-á-vis the benchmark”. The most common scalar performance measure consistent with TEV efficiency is the Information Ratio [12], defined as:

$$(1) \quad E[R_p - R_b] / \sqrt{Var[R_p - R_b]}$$

Note that (1) reduces to the textbook Sharpe Ratio when the benchmark portfolio is a constant return riskfree asset.² A different scalar performance measure is the expected exponential (a.k.a. CARA) utility:

$$(2) \quad E \left[-e^{-\gamma(R_p - R_b)} \right].$$

When $R_p - R_b$ is normally distributed, maximization of (2) yields a specific TEV-efficient portfolio that depends on the fixed $\gamma > 0$ used to evaluate portfolios in the opportunity set. These results have motivated Brennan [4], Gomez and Zapatero [11] and Becker, Ferson, et al. [2] to assume that (2) is a fund manager’s criterion when ranking normally distributed portfolios. Of course, only one value of γ will lead to the choice of the portfolio that maximizes the Information Ratio (1).

But several questions arise when considering the general legitimacy of using either (1) or (2) to rank portfolios relative to a designated benchmark:

1. The TEV logic underlying both (1) and (2) is single-period. Is there an appropriate performance measure for those wanting to beat the designated benchmark over a long horizon?

2. When differential returns are not normally distributed, should (1) or (2) be modified?
3. An advantage of (1) is the absence of critical, exogenous preference parameters, like γ in (2).

When differential returns are not normally distributed, can a performance measure analogous to (1) be found that obviates the need to know exogenous parameters like γ in (2)?

This paper argues that all three questions can be affirmatively answered by ranking portfolios in accord with an index of the probability that they will outperform the benchmark over typical long-term investors' time horizons. Section 2 provides the answer to the first question by using the Gärtner-Ellis Large Deviations Theorem [7, Chap.2] to show that the appropriate performance measure is the following function:

$$(3) \quad D_p \equiv \max_{\gamma} - \lim_{T \rightarrow \infty} \frac{1}{T} \log E \left[e^{-\gamma (\sum_{t=1}^T (\log R_{pt} - \log R_{bt}))} \right].$$

We will also show how (3) can be viewed as a generalized entropy. Section 2 also answers the second and third questions, by showing that ranking a portfolio p in accord with the generalized entropy (3) is equivalent to ranking it in accord with the asymptotic expected *power* utility of the ratio of wealth invested in the portfolio to wealth that would be earned by alternatively investing in the benchmark. This utility has relative risk aversion equal to $1 + \gamma_p$, where γ_p denotes the maximized value of γ in (3), and hence does not need to be exogenously specified. However, when approximation of time averaged log gross returns by arithmetic averaged net returns is reasonable (e.g. when $Var[R_p - R_b]$ is small) and when $R_p - R_b$ is IID, section 2.3 shows that the single-period exponential, TEV-based criterion (2) does arise from our criterion *without* assuming normality, by substituting our maximizing fund-specific γ_p from (3) for the fixed γ in (2). We also show that that under the additional assumption of normality, this substitution of γ_p for γ in (2) reduces to the Information

Ratio (1). *In this sense, the outperformance probability hypothesis nests the better-known criteria (1) and (2) as special cases, and hence is not subject to critiques commonly made of different probability-based criteria, e.g. expected utility maximization subject to “safety-first”, Value-At-Risk (VAR) constraints.*[1]. Section 2.3 contains the appropriate modifications of the Information Ratio that arise under non-IID normality. Then, we develop some historical returns-based estimators of the general performance index in section 3. Section 3.1.1 applies them to rank the relatively few mutual fund portfolios that, according to standard hypothesis tests, could outperform the S&P 500 in the long run. Section 4 develops some consequences of the auxiliary hypothesis that fund managers act (either now or eventually) as-if they maximized the outperformance probability. We argue that a recent and comprehensive study of an alternative fund manager behavioral hypothesis – expected exponential utility maximization with a fixed managerial risk aversion coefficient – provides little empirical evidence in favor of that alternative. In light of this, we argue that the well-established scientific principles of Popperian falsifiability and Ockham’s Razor weigh in favor of the outperformance probability maximization hypothesis, unless and until future empirical evidence weighs in favor of something else, because the hypothesis eliminates the free risk aversion parameter in conventional expected utility hypotheses. Finally, we show if the outperformance probability maximization hypothesis is true, econometric estimates of managerial risk aversion in tests of (what would then be) the misspecified expected utility maximization hypothesis are subject to a Lucas critique [20]. Section 5 concludes with several future research topics that are directly suggested by our findings.

While some other connections to the portfolio choice and asset pricing literatures are made in the following section, the most closely related papers use other outperformance probability criteria, and are now summarized. Unlike our approach, Stutzer [26] is not based on the probability that

the fund’s cumulative return should exceed the benchmark’s, and did not contain an empirical ranking of mutual funds. But that paper’s method was adopted for a time by Morningstar, Inc. to produce its “Global Star Ratings” of mutual funds [16], so our alternative approach should be of interest to performance analysts like them, as well as the fund managers they rank. Our approach also permits a stochastic benchmark, generalizing the constant growth rate benchmark used in the constantly rebalanced portfolio choice model of Stutzer [28]. The framework of that portfolio choice paper was used by Pham [21], to model optimal dynamic portfolio choice when a risky asset’s returns are generated by the process adopted in Bielecki, Pliska, and Sherris [3]. Finally, Browne [6, Sec. 4] formulated a related, but more specific criterion for optimal dynamic portfolio choice. After imposing restrictive portfolio and benchmark parametric price process restrictions, he characterized the portfolio that maximized “the probability of beating the benchmark by some predetermined percentage, before going below it by some other predetermined percentage.” Our analysis differs from his, by not assuming that agents are constrained to use such floors and ceilings, specific time horizons, nor specific parametric return processes.

2 An Index of Outperformance Probability

Ex-ante, wealth at some future time T arising from initial wealth W_0 invested in some portfolio strategy p will be denoted $W_T^p = W_0 \prod_{t=1}^T R_{pt}$, where R_{pt} denotes the random gross return from the strategy between times $t - 1$ and t . Note that the validity of this expression does not depend on the length of the time interval between $t - 1$ and t , nor the particular times t at which the random gross returns are measured. Similarly, an alternative investment of W_0 in a different portfolio b , dubbed the “benchmark”, yields $W_T^b = W_0 \prod_{t=1}^T R_{bt}$. Taking logs and subtracting shows that:

$$(4) \quad \log W_T^p - \log W_T^b = \log \frac{W_T^p}{W_T^b} = \sum_{t=1}^T \log R_{pt} - \sum_{t=1}^T \log R_{bt}.$$

From (4), we see that the portfolio strategy p outperforms the benchmark b when and only when the sum of its log gross returns exceeds the benchmark's. Dividing both sides of (4) by T yields the following expression for the difference of the two continuously compounded growth rates of wealth:

$$(5) \quad \log W_T^p/T - \log W_T^b/T = \frac{1}{T} \log \frac{W_T^p}{W_T^b} = \frac{1}{T} \sum_{t=1}^T (\log R_{pt} - \log R_{bt}).$$

Suppose one wants to rank portfolios according to the rank ordering of their respective probabilities for the event that $W_T^p > W_T^b$. Using (5), this *outperformance event* is the event that (5) is greater than zero – that is. the portfolio's continuously compounded growth rate of wealth exceeds that of the benchmark. Hence one desires a rank ordering of the probabilities

$$(6) \quad Prob \left[\frac{1}{T} \sum_{t=1}^T (\log R_{pt} - \log R_{bt}) > 0 \right].$$

which is equivalent to ordering the complimentary probabilities from lowest to highest, i.e. *we seek to rank a portfolio strategy inversely to its underperformance probability*:

$$(7) \quad Prob \left[\frac{1}{T} \sum_{t=1}^T (\log R_{pt} - \log R_{bt}) \leq 0 \right].$$

Of course, the rank ordering of portfolio strategies via (7) could depend on the exact value of the investor's horizon T . Because it is difficult for performance analysts to determine an exact value of an investor's horizon (when one exists), and because short horizon investors may have different portfolio rankings than long horizon investors, let us try to develop a ranking of (7) that will be valid

for all T greater than a suitably large T . This is similar in spirit to the motivation behind choice of an infinite horizon for the investor’s objective in most consumption-based asset pricing models (for a survey, see Kocherlakota [19]), or in many portfolio choice models (e.g., in Grossman and Vila [13]). Supporting evidence in Stutzer [28] shows that portfolios with relatively low underperformance probabilities (7) for suitably large T often also have relatively low underperformance probabilities for small T (or even *all* T) as well. Hence shorter term investors may also make use of the results.

To produce this ranking, we use the following two step procedure. First, reasonably assume that investors are not interested in portfolio strategies that (almost surely) will not even beat the benchmark when given an infinite amount of time to do so; in that event, they would prefer investing in the benchmark to investing in the portfolio. Hence, one should restrict the ranking to portfolios for which the underperformance probability (7) approaches zero as $T \rightarrow \infty$. More formally, one need only rank portfolio strategies p for which

$$(8) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (\log R_{pt} - \log R_{bt}) > 0.$$

Inequality (8) requires that the so-called *ergodic mean* of the log portfolio gross returns exceeds the benchmark’s (a.e.). In the familiar special case of IID returns, (8) requires that the portfolio’s expected log gross return exceeds the benchmark’s. In fact, if one weren’t worried about the probability of underperforming at finite horizons T , applying the law of large numbers to the limit in (8) shows that the highest performing fund would be the one with highest expected log return, i.e. the “growth optimal” fund that maximizes the expected log utility of wealth. It will almost surely *asymptotically* generate more wealth than the benchmark and all other funds will. But this “time diversification” argument in favor of log utility is irrelevant, because as Rubinstein [24] so effectively demonstrated, significant underperformance probabilities persist over time spans well in

excess of typical investors' retirement horizons. As he summarized: "The long run may be long indeed"!

So we will now quantify this downside, and formulate alternative utility-based formulations that reflect it. To do so, the second step of our procedure seeks a rank ordering index for the underperformance probability (7) of portfolios satisfying (8), i.e. those portfolios whose underperformance probabilities decay to zero as $T \rightarrow \infty$. Fortunately, the powerful, yet simply stated Gärtner-Ellis Large Deviations Theorem [7, Chap.2] is tailor-made for this purpose. This theorem shows that those portfolios' underperformance probabilities (7) will decay to zero as $T \rightarrow \infty$, at a portfolio-dependent exponential rate. As a result, the underperformance probability of a portfolio with a higher decay rate will approach zero faster as $T \rightarrow \infty$, and hence its complementary out-performance probability will approach 1 faster, so the portfolio will have a higher probability of outperforming for suitably large T . Again, it is important to emphasize that while this is formally an asymptotic criterion, in practice it will produce a ranking that applies to much shorter investor horizons T as well (see Stutzer [28] for substantial evidence establishing this). In summary, *the underperformance probability's rate of decay to zero as $T \rightarrow \infty$ is our proposed ranking index for portfolios. A portfolio whose underperformance probability decays to zero at a higher rate will be ranked higher than a portfolio with a lower decay rate.*

Direct application of the Gärtner-Ellis Large Deviations Theorem [7, Chap.2] shows that the decay rate of the underperformance probability (7) is

$$(9) \quad D_p \equiv \max_{\theta_p} - \lim_{T \rightarrow \infty} \frac{1}{T} \log E \left[e^{\theta_p \sum_{t=1}^T (\log R_{pt} - \log R_{bt})} \right]$$

Under the restriction (8), the maximizing θ_p in (9) will be negative (see Stutzer [28]), so without loss of generality one may substitute a value $-\gamma_p$, where $\gamma_p > 0$. Hence the rank ordering index is

the decay rate:

$$(10) \quad D_p = \max_{\gamma_p > 0} - \lim_{T \rightarrow \infty} \frac{1}{T} \log E \left[e^{-\gamma_p \sum_{t=1}^T (\log R_{pt} - \log R_{bt})} \right]$$

An expected utility interpretation of (10) is found by first substituting (4) into it and simplifying to yield:

$$(11) \quad D_p = \max_{\gamma_p > 0} \lim_{T \rightarrow \infty} -\frac{1}{T} \log E \left[\left(\frac{W_T^p}{W_T^b} \right)^{-\gamma_p} \right].$$

There are two differences between the right hand side of (11) and a power utility of wealth with exogenous degree of relative risk aversion $1 + \gamma$. First, the argument of the power function in (11) is not the wealth invested in the portfolio strategy, but instead is the ratio of it to the wealth that would have been earned if invested in the benchmark. This ratio is the state variable in the formulations of Browne [6]. It is analogous to the argument in the period utility of an “external habit formation”, consumption-based criterion [8, p.327]. Instead of our ratio of individual wealth to wealth created from an exogenous benchmark investment, its argument is the ratio of individual consumption to an exogenous benchmark function of aggregate (past) consumption (“keeping up with the Joneses”). In fact, when generalizing this argument to model other forms of this consumption externality, Gali [10, Footnote 2] noted that “such a hypothesis may be given an alternative interpretation: agents in the model can be thought of as professional “portfolio managers” whose performance is evaluated in terms of the return on their portfolio relative to the rest of managers and/or the market.” While our infinite horizon criterion for pure investment (11) is perhaps more reminiscent of the asymptotic growth of expected utility criterion $J_p = \lim_{T \rightarrow \infty} \frac{1}{T} \log E[(W_T^p)^\alpha]^{1/\alpha}$ used by Grossman and Zhou [14] and Bielecki, Pliska and Sherris [3], it should be possible to adapt the reasoning leading to (11) to analyze the consumption/investment problem with consumption

externalities. The meaning of the degree of risk aversion in all these models is not the usual aversion to mean preserving spreads of distributions of wealth or consumption, but rather of distributions of wealth or consumption *relative to a benchmark*.

The second and more unusual difference between our criterion and all these others is that the curvature parameter γ_p on the right hand side of (11) is determined by maximization of expected utility, and is hence dependent on the stochastic process for $\log R_{pt} - \log R_{bt}$. Hence outperformance probability maximizing portfolio choice may be reconciled with expected utility maximizing portfolio choice, by requiring that *both* the portfolio p *and* the degree of risk aversion $1 + \gamma_p$ be chosen to maximize D_p in (10) or (11).

Applicability of the Gärtner-Ellis Theorem requires that one maintain the assumptions that the limit in (10) exists (possibly as the extended real number $+\infty$) for all $\gamma > 0$, and is differentiable at any γ yielding a finite limit. Many log return processes adopted by analysts will satisfy these hypotheses, as will be demonstrated by example.

2.1 Entropic Interpretation

One can also show that (10) is a generalization of a suitably minimized value of the Kullback-Leiber Information Criterion or relative entropy $I(P \parallel \mu)$ between probability distributions P and μ . To see this, note that when the process $\log R_{pt} - \log R_{bt}$ is IID with the identical distribution of $\log R_p - \log R_b$ denoted by μ , (10) simplifies to $D_p \stackrel{IID}{=} \max_{\gamma_p} -\log E_{\mu}[e^{-\gamma_p(\log R_p - \log R_b)}]$. Kullback's Lemma [18, Proposition 2.2] shows that:

$$(12) \quad D_p \stackrel{IID}{=} \max_{\gamma_p} -\log E_{\mu}[e^{-\gamma_p(\log R_p - \log R_b)}] = \min_{\{P: E_P[\log R_p - \log R_b] = 0\}} I(P \parallel \mu)$$

Hence the underperformance probability decay rate to zero is the constrained minimum value of the relative entropy, over the set of probability distributions that make the expected log portfolio return equal to the benchmark's expected log return. As such, (10) or (11) may be viewed as the value of a generalized entropy appropriate for general non-IID excess log returns.

2.2 Time Varying Gaussian Log Returns

In order to both illustrate the calculation (10) and to relate it to the Information Ratio, suppose that for each time t , $\log R_{pt} - \log R_{bt}$ is normally distributed, so that each partial sum $\sum_{t=1}^T (\log R_{pt} - \log R_{bt})$ in (10) is a normally distributed random variable with mean $\sum_t E[\log R_{pt} - \log R_{bt}]$ and variance $Var[\sum_t (\log R_{pt} - \log R_{bt})]$. But (10) is just -1 times the time average of the log moment generating functions of these normally distributed random variables, evaluated at the maximizing $-\gamma_p$. Hence in the Gaussian case, (10) is just the quadratic function

$$(13) \quad D_p = \max_{\gamma} \lim_{T \rightarrow \infty} \frac{\sum_{t=1}^T E[\log R_{pt} - \log R_{bt}]}{T} \gamma - \frac{1}{2} \frac{Var[\sum_{t=1}^T (\log R_{pt} - \log R_{bt})]}{T} \gamma^2.$$

This first term in (13) will exist whenever the ergodic mean (8) exists, while the second term will exist whenever the analogous ergodic variance exists. These are standard assumptions to make in econometric estimation. Setting the first derivative of (13) equal to zero and solving for the maximizing γ yields:

$$(14) \quad \gamma_p = \frac{\lim_{T \rightarrow \infty} \sum_{t=1}^T E[\log R_{pt} - \log R_{bt}]/T}{\lim_{T \rightarrow \infty} Var[\sum_{t=1}^T (\log R_{pt} - \log R_{bt})]/T}$$

which from assumption (8) is positive, as asserted earlier. Now substitute (14) back into (13) and rearrange to obtain the following underperformance probability decay rate in the Gaussian case:

$$(15) \quad D_p = \frac{1}{2} \left(\frac{\lim_{T \rightarrow \infty} \sum_{t=1}^T E[\log R_{pt} - \log R_{bt}]/T}{\sqrt{\lim_{T \rightarrow \infty} \text{Var}[\sum_{t=1}^T \log R_{pt} - \log R_{bt}]/T}} \right)^2.$$

Hence the Gaussian performance index (15) depends on the ratio of the long run mean excess log return to its long run standard deviation, and is hence a generalization of the usual Information Ratio. This differs from the Gaussian (i.e. second order) approximation of the performance criterion in Grossman and Zhou [14] and Bielecki, Pliska, and Sherris [3], which (the latter paper shows) is the long run mean minus an exogenous risk aversion parameter times the long run variance.

2.3 Familiar Performance Measures as Approximations

The single period, exponential utility performance measure (2) has been rationalized by its consistency with Roll's (op.cit.) TEV-efficiency when the difference in returns $R_{pt} - R_{bt}$ is IID normal. To obtain something akin to (2) by an approximation of our index (10), first approximate a log gross return $\log R$ by its net return $R - 1$, i.e. substitute $R_{pt} - R_{bt}$ for $\log R_{pt} - \log R_{bt}$ in (10). Now under the restriction that the difference in the time series of equity returns is produced by a serially independent process, one obtains the following performance measure:

$$(16) \quad \max_{\gamma_p} - \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \log E \left[e^{-\gamma_p (R_{pt} - R_{bt})} \right]$$

which under the additional restriction that the independent distributions are identically distributed reduces to the single-period performance measure:

$$(17) \quad -\log E \left[e^{-\gamma_p (R_p - R_b)} \right]$$

that rank orders portfolios in the same way as the single-period expected exponential utility:

$$(18) \quad E \left[-e^{-\gamma_p (R_p - R_b)} \right].$$

Note that (18) is based on the exponential function used in the TEV-based index (2), despite the fact that the argument for it made no use of normality. But (18) is not the same as (2), because (18) uses the portfolio-dependent, expected utility maximizing γ_p when ranking portfolio p , rather than some constant value of γ used for all p . However, if we impose the additional restriction that is used to rationalize the TEV hypothesis, i.e. that $R_p - R_b$ is normally distributed, we can substitute the Gaussian (quadratic) log moment generating function into the equivalent problem (17) and solve to yield the following maximizing γ_p

$$(19) \quad \gamma_p = E[R_p - R_b] / \text{Var}[R_p - R_b].$$

Substituting (19) back into that log moment generating function and rearranging yields

$$(20) \quad \frac{1}{2} \left(\frac{E[R_p - R_b]}{\sqrt{\text{Var}[R_p - R_b]}} \right)^2$$

which is half the squared Information Ratio (1). Hence we see that under the log return approximation, and the IID normal process restriction used to rationalize the TEV hypothesis, use of the fund-specific maximizing γ_p (19) transforms the exponential utility (2) into a parameter-free performance measure (20), that ranks the funds in the same order as the Information Ratio (1)! In fact, when the differential return itself is normally distributed (rather than the differential *log gross* return), Stutzer [27, Proposition 2] shows that the Information Ratio (1) ranks portfolios in accord with their respective returns' probabilities of outperforming the benchmark return *on average* over the horizon T , rather than their respective *cumulative* returns outperforming the benchmark cumulative returns *at T*. Hence our general performance measure (10) or (11) may be viewed as a generalization of the TEV-based performance measures (1) and (2), to be used when the approximation of log returns by net returns, the IID assumption, and the normality assumption are

unwarranted.

Finally, to see what happens when we maintain the IID assumption without either the approximation of log returns by net returns nor the normality assumption, one can apply the IID restriction directly to the difference of log gross returns in (10), producing the following index:

$$(21) \quad \max_{\gamma} -\log E \left[\left(\frac{R_p}{R_b} \right)^{-\gamma} \right]$$

which yields the same rank ordering of portfolios as the expected power utility of the return ratio

$$(22) \quad E \left[- \left(\frac{R_p}{R_b} \right)^{-\gamma_p} \right],$$

instead of the expected exponential utility of the return difference (18).

3 Nonparametric Estimation of the Performance Measure

The simplest estimator of the performance index (10) arises when one makes the additional assumptions that the differential log gross return process $X_{pt} \equiv \log R_{pt} - \log R_{bt}$ is independently and identically distributed (IID). As argued earlier, one need only rank the fund portfolios p that almost surely will outperform the benchmark, i.e. that satisfy (8). Under the IID assumption, (8) requires that $E[X_p] > 0$, and (10) reduces to

$$(23) \quad D_p = \max_{\gamma > 0} -\log E \left[e^{-\gamma X_p} \right]$$

Given an observed time series of past observations, denoted $X_p(t)$, $t = 0, \dots, T$, one can consistently estimate (23) by substituting its sample analog, which is just:

$$(24) \quad \hat{D}_p^{IID} = \max_{\gamma > 0} -\log \frac{1}{T} \sum_{t=1}^T e^{-\gamma X_p(t)}.$$

But what if X_{pt} is not IID, necessitating a good estimator for the general rate function (10) rather than its IID special case (23). An argument in Kitamura and Stutzer [18, pp. 168-169] showed that the estimator in Kitamura and Stutzer [17] provides the basis for an analog estimator of the general rate function (10). Specifically, one must select a “bandwidth” integer $K > 0$ used to smooth the series $X_p(t)$ (via a two-sided average of radius K about each $X_p(t)$), resulting in the general rate function estimator:

$$(25) \quad \hat{D}_p^K = \max_{\gamma > 0} -\frac{1}{2K+1} \log \frac{1}{T-2K} \sum_{t=K+1}^{T-K} e^{-\gamma \frac{\sum_{j=-K}^K X_p(t-j)}{2K+1}}$$

Of course, an analyst who was truly confident that some specific parametric stochastic process generated $X_p(t)$ could attempt to either directly calculate a formula for (10) (as was done for Gaussian processes in section 2.2) and then estimate it, or could use the parametric process to construct a direct simulation estimator of (10).

3.1 Empirical Results

It is useful to compare our fund performance and risk aversion coefficient estimates with results from other recent studies. To foster this comparison, we examined mutual funds during the 228 months starting in January 1976 and ending in December 1994. This coincides with the estimation period used in Becker, Ferson, et al. (op.cit), and is almost identical to the estimation period used in the mutual fund performance study of Wermers [30]. We also followed Wermers in using the CRSP Survival-Bias Free US Mutual Fund Database, originated by Mark Carhart [9]. After describing our results and quantifying the sampling error present in them, we will use them to re-examine some of the claims in Wermers, which of course were based on different performance criteria. We will then compare our fund-specific risk aversion coefficient estimates to the fund

manager risk aversion coefficient estimates reported in Becker, Ferson, et.al. We follow them in adopting the most commonly cited mutual fund benchmark, the S&P 500 index.³

3.1.1 Fund Performance

In accord with the rationale presented in section 2, one should only rank funds (if any) that would asymptotically outperform the S&P 500, i.e. one should only rank funds that satisfy (8). The S&P 500 benchmark monthly returns $\log R_b(t)$ were subtracted from the corresponding monthly returns of each of equity mutual funds returns $\log R_p(t)$ to produce the historical time series $X_p(t)$, and used to conduct the usual *one-way* paired difference of means tests of the null hypothesis $H_0 : E[X_p] = 0$ versus the alternative hypothesis $H_1 : E[X_p] > 0$. The test statistic is $\overline{X_p} \equiv \sum_{t=1}^{228} X_p(t)/228$.

Performance analysts will not be surprised to find out that the null hypothesis was rejected in favor of the alternative ($t > 1.65$) for only 32 of the 347 CRSP mutual funds whose returns persisted from January 1976 to December 1994. Now if one were testing an hypothesis about whether or not a *typical* fund beat the S&P 500, a survival bias would result from examining only those funds that survived over that entire period. But that hypothesis is *not* being tested here. Even though the procedure here examines only those 347 funds that were skillful or lucky enough to survive those 228 months, a standard hypothesis test concludes that only 32 of the 347 could asymptotically outperform the S&P 500, strengthening our conclusion that relatively few funds will do so.

Summary statistics and performance rankings for those 32 funds are reported in Table 1, ranked in order of their estimated performance index values $\hat{D}_p^{IID}\%$ given by (24). This estimator might be problematic for funds whose X_{pt} are serially correlated. A standard test of the hypothesis that the 1st through 6th autocorrelation correlation coefficients are all zero was rejected at the 5% level for only 8 of the 32 funds, and even those had low autocorrelation coefficients. Not surprisingly,

Mutual Funds with $E[X_p] > 0$						
CRSP #	\bar{X}_p %	SDev %	Skew	Kurt	\hat{D}_p % (24)	$1 + \hat{\gamma}_p$
36450	.80	2.4	-.66	5.94	4.95	13.5 (5.5)
17960	.50	2.1	-.42	3.62	2.77	12.6 (4.1)
17010	.33	1.7	-.26	0.64	1.81	12.8 (4.2)
05730	.41	2.4	-.16	1.00	1.41	8.9 (2.9)
31250	.47	2.9	.36	1.51	1.38	7.9 (2.5)
19760	.42	2.6	.17	1.76	1.37	8.5 (2.8)
31820	.29	1.8	-.09	0.71	1.34	10.9 (3.8)
18290	.43	2.7	-.01	2.26	1.28	8.0 (2.7)
17920	.35	2.3	.33	3.25	1.23	9.0 (3.4)
16010	.25	1.6	-.02	2.31	1.19	11.4 (4.6)
13520	.26	1.7	.06	0.64	1.15	10.9 (4.1)
18350	.27	1.9	.92	4.15	1.07	10.2 (3.9)
08162	.40	2.8	-.065	5.11	0.97	5.8 (2.8)
37600	.33	2.4	.16	.80	0.95	7.8 (2.9)
31380	.37	2.7	-.79	6.43	0.94	6.9 (3.1)
08161	.55	4.1	-.72	5.40	0.86	5.1 (1.9)
06570	.27	2.1	-.09	1.34	0.83	8.2 (3.4)
14350	.26	2.0	-.37	2.30	0.82	8.3 (3.6)
04013	.44	3.5	-.01	0.86	0.81	5.7 (2.0)
17280	.03	2.3	-.30	1.63	0.81	7.4 (3.0)
14240	.38	3.0	-.74	4.04	0.80	6.1 (2.5)
09330	.47	3.7	-.51	1.52	0.78	5.3 (1.9)
16290	.27	2.2	-.39	2.19	0.76	7.6 (3.3)
08340	.24	2.0	.01	1.29	0.76	8.2 (3.5)
15700	.23	1.9	-.57	5.07	0.74	8.4 (4.1)
18790	.30	2.5	-.26	4.02	0.72	6.8 (3.0)
08620	.29	2.4	-.45	2.35	0.72	6.9 (3.0)
18070	.35	3.1	.19	3.45	0.66	5.8 (2.4)
08480	.35	3.1	-.30	.40	0.65	5.7 (2.2)
15270	.23	2.1	.42	3.35	0.64	8.3 (3.4)
16820	.30	2.7	-.28	0.97	0.63	6.2 (2.6)
18050	.33	3.0	-.12	1.27	0.60	5.6 (2.3)

Table 1: The 32 out of 347 CRSP mutual funds that could (on the basis of a standard hypothesis test) statistically significantly outperform the S&P 500 index during January 1976-December 1994. Rankings are in order of the nonparametrically estimated decay rate \hat{D}_p from (24). No fund's implied degree of relative risk aversion $1 + \hat{\gamma}_p$ from (24) is implausibly high, and all have standard errors (in parentheses) that are generally much lower than the estimates themselves, unlike the risk aversion estimates reported in Becker, Ferson, et al. (1999).

use of the alternative estimator \hat{D}_p^K given by (25) gave the exact same performance ranking of the funds, for each value of the “bandwidth” K tested (i.e. $K = 3, 6,$ and 12), and hence are not reported in the table. ⁴ A (naive) analyst who is only concerned with the terminal wealth in the funds at the end of the ranking period, i.e. the cumulative return over the ranking period, would *not* rank the funds by $\hat{D}_p\%$. Instead, that analyst would rank the funds in order of the average difference in log returns over the rating period (5), listed in the second column of Table 1 as $\bar{X}_p\%$. But $\bar{X}_p\%$ ranks the funds very differently than $\hat{D}_p\%$, which is used by analysts concerned with eventual underperformance that did not happen during the ranking period. However, the two rankings do agree on the top-ranked fund 36450. Perhaps not surprisingly, this is the Fidelity Magellan fund. Our estimates show that its portfolio strategy resulted in the least probability (7) of underperforming the S&P 500 benchmark, because the probability of underperforming it decays to zero as $T \rightarrow \infty$ at the highest estimated rate $\hat{D}_p^{IID} = 4.95\%$ per month. Using the well-known, compounding “Rule of 72” approximation, the underperformance probability (7) will eventually be cut in half about every $72/4.95 \approx 15$ months. While the bottom-ranked fund 18050 should also eventually outperform the S&P 500 (due to rejection of its null $E[X_p] = 0$ with $t > 1.65$), its probability of underperformance is estimated to die off much slower, i.e. it will eventually halve only every $72/0.60 = 120$ months.

We designed and conducted a bootstrap resampling study, to study the likely impact of sampling error on the stability of our nonparametrically estimated rankings. We resampled the 228 months (with replacement) 10,000 times, to construct alternative possible $X_p(t)$ series for each of the 32 funds that could (on the basis of the nonparametric hypothesis test) outperform the S&P 500. After each of the 10000 replications, we estimated the 32 funds performance index values (24) and ranked them. In each replication, we followed Brown and Goetzmann’s [5] in classifying fund performance

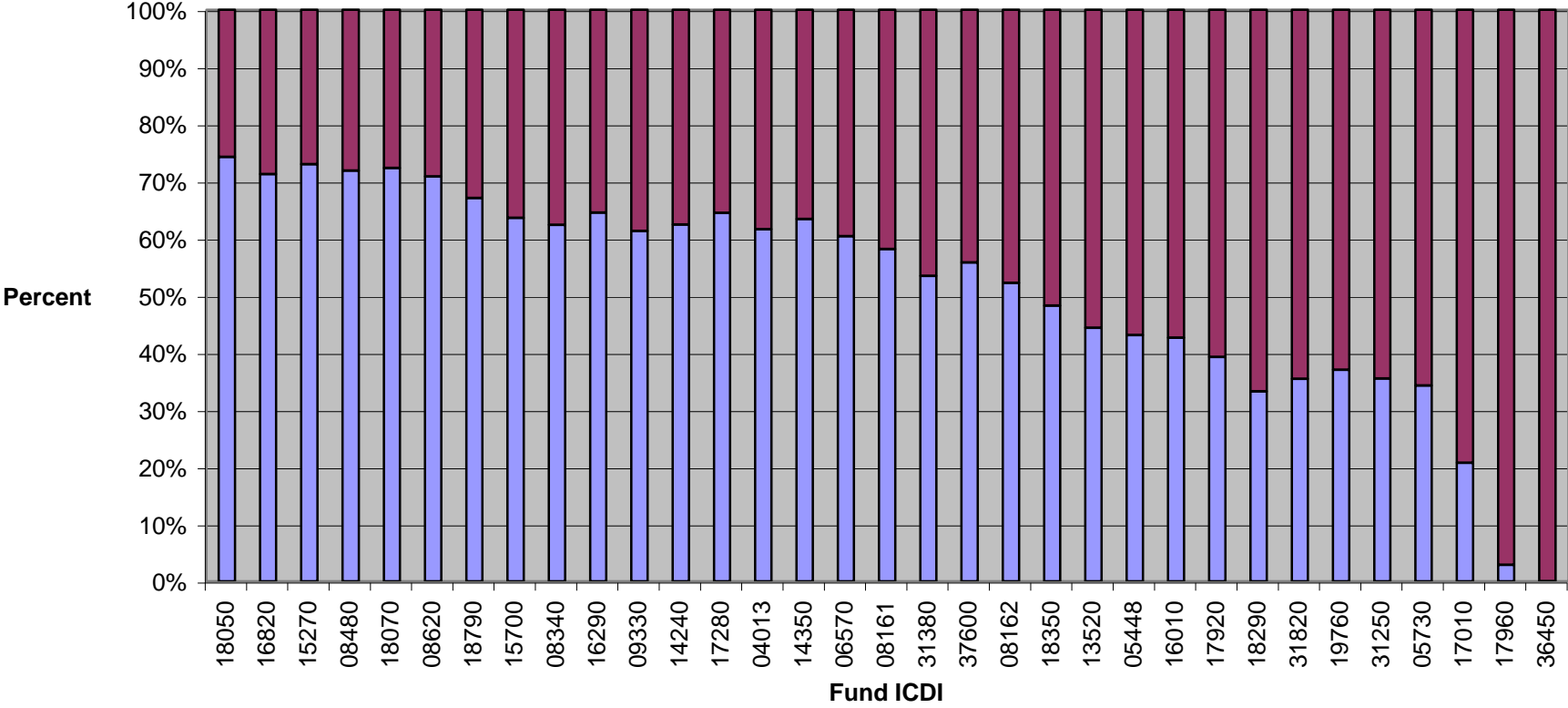
as either above (“high”) or below (“low”) the median performance for that replication. Figure 1 shows the nonparametric results for the funds listed in Table 1. The first panel of Figure 1 clearly shows a high degree of stability where it is most needed, i.e. at the top end of the ranking. The highest rated fund (Fidelity Magellan) stayed above the median in virtually all replications. More detail is provided in the second panel of Figure 1, which lists each fund’s four transition probabilities of moving from above (below) the median to above (below) the median over successive replications. The panel confirms that it is more common for low funds to stay low and high funds to stay high in successive replications; as expected, the top and bottom ranked funds experience the most stability in this sense.

It is fruitful to re-examine one of the findings highlighted in Wermers’ [30] study of mutual fund performance, conducted over an almost identical period. Wermers (op.cit, p. 1686) asked the question: “Do higher levels of mutual fund trading result in higher levels of performance?” Wermers attempted to answer this question by constructing a hypothetical portfolio implemented by annually shifting money into funds that had relatively high turnover during the previous year. He concludes (op.cit, p. 1690): “Although these high-turnover funds have negative (but insignificant) characteristic-adjusted net returns, their average unadjusted net return over our sample period significantly beats that of the Vanguard Index 500 fund.”

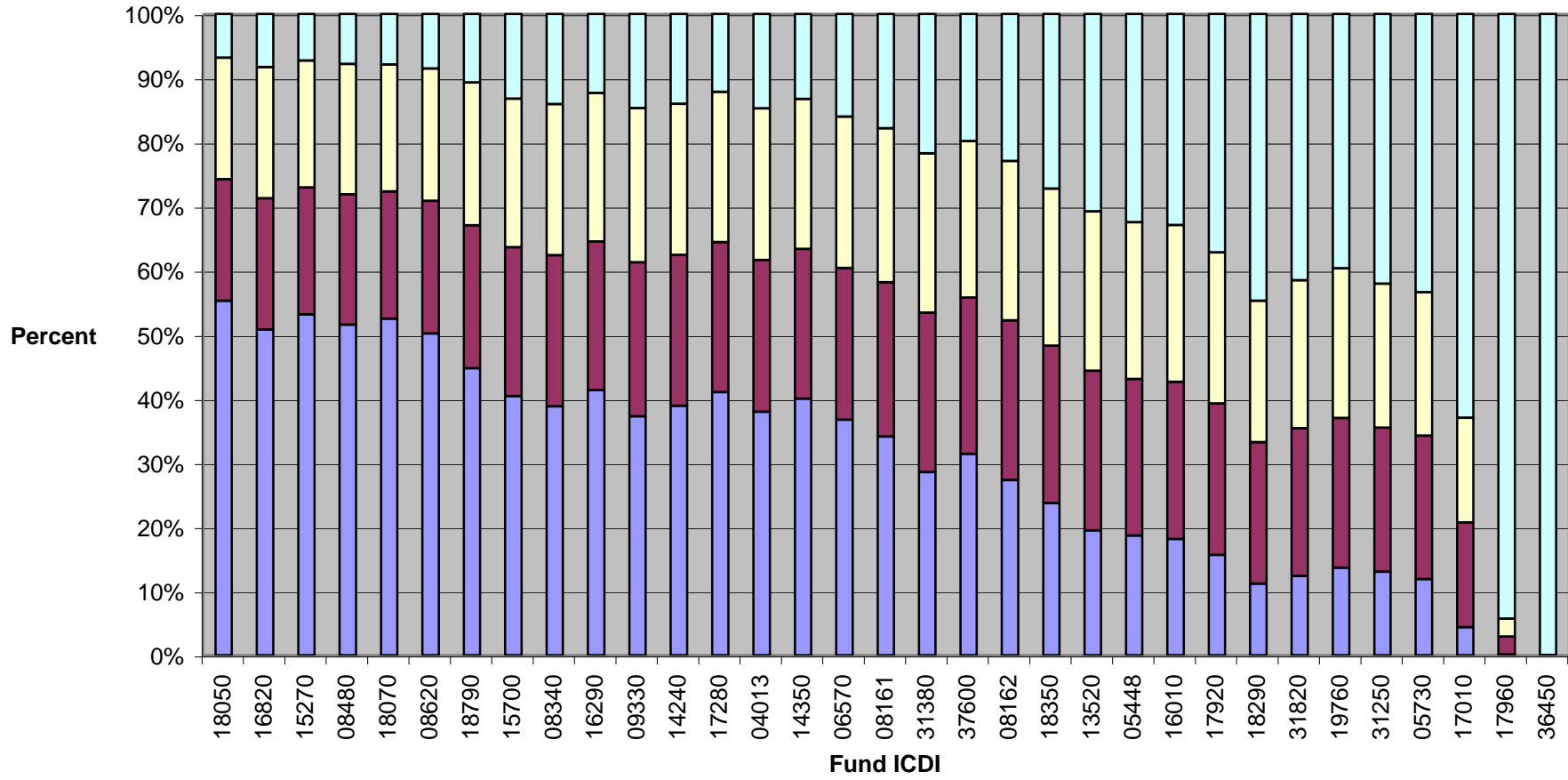
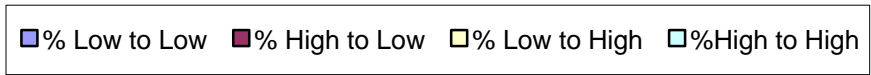
There are two questions left unanswered by Wermers’ conclusion. First, we have shown that fund investors, who want their invested wealth to exceed that which would have accrued in the S&P 500 stocks, must not use the average net return as a performance measure. Rather, they must use the average $\log(1 + \text{net return})$, akin to the geometric average, which can be significantly lower when returns are volatile. Second, Wermers examined a *strategy* of annually moving money from low turnover funds to high turnover funds. Significant load payments may pile up while doing this,

Fund Categorization

% Low Rated % High Rated



Persistence in Mutual Fund Rankings



but even if they did not, investors may also want to know whether or not it pays to buy-and-hold individual mutual funds that have relatively high average turnover.

As Wermers notes, annual turnover rates are reported for each fund. So we compiled the median of the 19 years' turnover rates for each of the 32 funds that could (on the basis of the standard hypothesis test) asymptotically outperform the S&P 500. These ranged from a high of 181% to a low of 15%; half of these 32 funds had median annual turnover below 65%. Half of the 347 funds that had the full 228 monthly returns reported for our data period had median annual turnover below 43%, as reported in Table 2. So it is fair to say that the 32 outperforming funds had somewhat higher turnover than the rest.

Also, half of the 32 funds in Table 2 had median annual stock allocations that exceeded 85%, compared to 80% for the 347 funds. Moreover, the 32 funds had expense ratios that were similar to the 347 funds; half of each group had a median annual expense ratio below 89 basis points. Hence, the outperformance of the 32 funds does not appear to be associated with atypical stock allocations or expense ratios.

The final column of Table 1 lists the estimates of the fund-specific coefficients of risk aversion $1 + \gamma_p$ from (24), and their bootstrapped standard errors (in parentheses). No fund's coefficient is implausibly large, nor are the estimates imprecise, i.e. the corresponding standard errors are considerably smaller than the estimates. Table 1 shows that there is positive, but not perfect, correlation between a fund's degree of risk aversion and its outperformance probability ranking. To help see why, let us approximate the index by substituting the net return difference for the log return difference, and assume that the net return difference is IID Gaussian. Then, (19) shows that the manager's degree of risk aversion will be high whenever the ratio $E[R_p - R_b]/Var[R_p - R_b]$ is high. In this case, the approximated performance index is (20), i.e. half the squared or Information Ratio,

Mutual Funds with $E[X_p] > 0$						
CRSP #	Fund	Turnover	Expense %	Stock %	Bond %	Cash %
36450	Fidelity Magellan	1.26	1.08	92	2	1
17960	Fidelity Destiny I	0.75	0.68	NA	NA	NA
17010	NY Venture	0.59	1.02	91	2	5
05730	Mutual Shares	0.67	0.74	67	21	13
31250	Lindner Growth	0.21	0.89	80	6	14
19760	Sequoia	0.32	1.00	66	0	27.2
31820	Guardian Park Ave.	0.48	0.72	82	10	7
18290	Acorn Inv. Trust	0.32	0.82	85	3	9
17920	Van Kamp-Amer Pace/A	0.45	1.01	86	0	14
16010	Neub & Berm Partners Fund	1.81	0.95	73	2	25
13520	Fidelity Equity Income	1.07	0.72	68	26	5
18350	Pioneer II	0.26	0.81	90	1	8
08162	20th Cent.Select	0.95	1.01	99	0	1
37600	SteinRoe Special	0.61	0.96	84	1	9
31380	Evergreen Fund/Y	0.48	1.13	92	0	8
08161	20th Cent.Growth	1.05	1.01	99	0	1
06570	Templeton Growth/I	0.15	0.75	87	4	10
14350	AMCAP	0.18	0.73	84	0	14
04013	Spectra	NA	NA	NA	NA	NA
17280	Nicholas	0.27	0.86	89	0	11
14240	Weingarten Equity	1.09	1.19	96	0	4
09330	Loomis-Sayles Cap. Dev.	1.29	0.79	98.6	0	1.4
16290	IDS New Dimensions/A	0.70	0.74	87	0	13
08340	Windsor	0.34	0.53	84	2	10
15700	Van Kamp-Amer Comstock/A	0.61	0.79	85	0	11
18790	Janus	1.63	1.02	71	1	27
08620	Growth Fund of America	0.22	0.76	81	0	16
18070	Value Line Leveraged Growth	1.13	0.91	95	0	3
08480	St. Paul Growth	0.84	1.00	84	0	16
15270	Charter	1.06	1.25	77	0	23
16820	Putnam Voyager/A	0.65	1.10	93	0	6
18050	Van Kamp-Amer Emerg.Grwth/A	0.93	1.05	91	0	8.30
	Median: All 347 Funds	0.43	0.89	80	0	6

Table 2: Comparison of the 32 outperforming funds' characteristics to the median characteristics of all 347 CRSP mutual funds over the period January 1976-December 1994. The 32 outperforming funds had slightly higher turnover and fractional allocation to equities than the 347 funds did, and similar expense ratios. Hence, the outperformance of the 32 funds does not appear to be associated with atypical stock allocations or expense ratios.

which is half the ratio $E[R_p - R_b]^2 / \text{Var}[R_p - R_b]$. So in this case, γ_p differs from D_p only because its numerator $E[R_p - R_b]$ is not squared. So there will be a *tendency* for a fund's performance and risk aversion coefficient to be directly related, but the relationship is not perfect. The most extreme outperformance would occur when $R_p = R_b + c$ with $c > 0$, i.e. when the portfolio return is perfectly correlated with the benchmark, but is always higher by a constant amount. In this case, $\text{Var}[R_p - R_b] = 0$, so the degree of risk aversion is infinite, and shorting the benchmark to purchase the portfolio would then provide an arbitrage opportunity – the ultimate in outperformance! The top-ranked fund 36450 (Fidelity Magellan) certainly did not choose a portfolio that was that good, but its outstanding performance is a noisy signal that its portfolio management acted *as-if* it was highly risk averse to the consequences of underperforming the S&P 500; specifically, its estimated coefficient of risk aversion is $1 + \gamma_p = 13.5$.

4 Outperformance Probability Maximization as a Fund Manager

Behavioral Hypothesis

Becker, Ferson, et al. (op.cit) used funds' returns from the same period we did, to test the hypothesis that fund managers with fixed (but possibly different) coefficients of risk aversion acted as-if they maximized (2). Their Table 5 (op.cit, p. 139) reports summary statistics for the individual asset allocator funds' managers' estimated risk aversion parameters γ . For the "asset allocation" category of funds that they (reasonably) believed to behave most in accord with their model, the mean value of the fund-specific risk aversion parameters is an implausibly high $\gamma = 93.6$ while the median is actually negative, i.e. $\gamma = -13.4$. If so, more than half the fund managers had negative risk aversion, resulting in a strictly *convex* utility (2). Hence they would not have acted as-if

they solved the first order conditions employed by Becker, Ferson, et.al. in their model. Equally implausible and/or imprecise estimates of managerial risk aversion γ were reported for the other fund categories (op.cit., Table 3,pp. 135-136). They accurately conclude that “...the risk aversion estimates are imprecise” (op.cit, p. 145), and reported that the standard Hansen J-Statistic test of their model’s GMM moment restrictions frequently failed.⁵ In contrast, our estimates of fund-specific risk aversion coefficients, listed in Table 1, are neither very high nor imprecisely estimated. So we now examine the possibility that fund managers might instead be trying to maximize the probability of outperforming their respective benchmarks.

4.1 Scientific Principles for Evaluating Managerial Behavioral Hypotheses

Suppose a researcher wants to construct an hypothesis of fund manager behavior based on maximization of some criterion, and then test it using data from a group of managed funds. In order to produce an hypothesis that is both tractable and testable, suppose the researcher adopts the following (not atypical) maintained assumptions:

1. All fund managers in the group evaluate performance relative to the *same* benchmark portfolio.
2. All fund managers in the group perceive the *same* investment opportunity set.
3. All fund managers in the group have the technical ability to maximize the criterion.

Under these maintained assumptions, the outperformance probability maximizing hypothesis predicts that all fund managers in the test group would have chosen the *same* portfolio $p_{\max} \equiv \arg \max_p D_p$, which would be associated with the same coefficient of risk aversion $1 + \gamma_{p_{\max}}$; something unlikely to occur in practice. The same sort of counterfactual implication also arises in the

standard mean-variance hypothesis of individual investor behavior. In the presence of a riskfree asset and the maintained assumptions listed above, the mean-variance hypothesis predicts that all investors in the test group will hold risky assets in the same proportions that the Sharpe Ratio maximizing “tangency” portfolio does, i.e. in the same proportions as the market portfolio in the CAPM.

Now consider the Becker, Ferson, et.al. (op.cit.) hypothesis that managers, each of whom has a fixed risk aversion parameter γ , maximize (2). Under the three maintained assumptions above, the hypothesis’ portfolio choice prediction still depends on each manager’s unobservable risk aversion parameter γ . The hypothesis would not predict that all the managers in the test group choose the same portfolio. Only managers with the same unobservable risk aversion parameter would be predicted to choose the same portfolio. Hence under the three maintained assumptions, the hypothesis need not make the prediction of homogeneous behavior that the outperformance probability maximization hypothesis of fund manager behavior would make (and with the additional presence of a riskless asset, the mean-variance hypothesis of individual behavior would make). But the ability to avoid this was obtained through flexibility arising from the hypothesis’ introduction of an additional free parameter, i.e. a manager’s value of γ . From a scientific viewpoint, is this a strength or a weakness?

Subsequent to Popper’s seminal work [22], scientists have considered a more potentially falsifiable hypothesis to be better than a less potentially falsifiable one, unless empirical evidence clearly favors the latter. This paper showed that the outperformance probability maximization hypothesis is equivalent to maximizing an expected utility over *both* a fund manager’s risk aversion “parameter” and the investment opportunity set. By endogenizing γ , it is eliminated as a free parameter, enabling the hypothesis to make a determinate portfolio choice prediction under the typical theo-

retical conditions (i)-(iii) above. This sharp prediction is of course more potentially falsifiable than the set of γ -dependent predictions made by the conventional hypothesis (2), and is thus favored by Popper's falsifiability principle for evaluating scientific hypotheses. A related principle – the Principle of Ockham's Razor – favors the simplest hypothesis that can potentially make the right prediction. In addition to being a principle accepted by most scientists, Jeffreys and Berger [15] have found a Bayesian rationalization for it, summarized as follows:

Ockham's Razor, far from being merely an *ad-hoc* principle, can under many practical situations in science be justified as a consequence of Bayesian inference....a hypothesis with fewer adjustable parameters has an enhanced posterior probability because the predictions it makes are sharp.[15]

Of course, these principles cannot be used to favor an hypothesis that is obviously counterfactual *about an absolutely critical fact* over an hypothesis that is not. But the homogeneity prediction of the outperformance probability maximization hypothesis was predicated on the three assumptions listed above, and it is highly unlikely that all three assumptions are valid in the real world. Currently, Morningstar separates domestic equity funds into 20 different categories, based on capitalization, style, and sector specializations. Morningstar ranks a fund against the others in its category, rather than against the entire fund universe. Hence it is certainly possible that some of the 32 funds ranked here tried to beat benchmarks other than the S&P 500, perhaps benchmarks specific to their capitalization, style, or sector specialization, thus violating the first assumption listed above. In fact, Becker, Ferson, et.al. also permitted funds in different categories to have different benchmarks; specifically, funds were assumed to have benchmarks that could be different (unobservable) *h*-weighted averages of the S&P 500 and T-Bills. The second assumption listed

above is also unlikely, as it requires that the funds face the same restrictions (if any) on trading, and that the fund managers agree on the forms and parameters of all portfolios' differential (log gross) return processes. The latter is particularly unlikely to be true in practice, due to differences in managers' opinions about the likely future performance of the individual stocks, bonds, etc. that can be used to form their portfolios. Readers doubting this point should watch a randomly chosen episode of the public television show "Wall Street Week", which would also cast doubt on the realism of the third assumption listed above, i.e. that all fund managers have the technical skill to maximize a quantitative criterion function. Hence even if the test group of 32 fund managers examined herein were trying to maximize the probability of outperforming the S&P 500, the likely violations of the second and third assumptions could explain why they chose different portfolios, with differing fund-specific outperformance probabilities and coefficients of risk aversion evidenced in Table 1.

Furthermore, if our outperformance probability hypothesis is correct, the alternative hypothesis – that a manager with a fixed risk aversion parameter γ maximizes (2) – is subject to the critique made in a justly celebrated paper by Robert E. Lucas [20]. He criticized econometric analyses that incorrectly hold parameters of an optimizing agent's decision rule fixed when analyzing the effects of policy-induced changes in the decision making environment. He showed that these parameters would change endogenously when agents optimized more thoroughly than was incorrectly assumed.

Econometric analyses that fix γ as a managerial preference "parameter" are subject to a similar critique. Our hypothesis implies that the optimal portfolio is associated with a portfolio-specific degree of risk aversion that depends on the decision making environment, i.e. the benchmark portfolio and investment opportunity set, via maximization of the performance measure (10) or (11). Plan sponsors and/or their investors, who designate a fund benchmark for management

to beat, are analogous to Lucas' policymakers. Should they designate a tougher benchmark, they should anticipate that the outperformance probability maximizing manager will act as-if he/she had a lower degree of (endogenous) risk aversion. For example, if the Fidelity Magellan fund actually found the portfolio strategy that maximized the probability of outperforming the S&P 500, then the first row in Table 1 shows that its degree of risk aversion was 13.5. But had investors insisted that Magellan designate an even tougher benchmark to beat than the S&P 500, its manager would have acted as-if he/she had a lower degree of risk aversion.

This critique is most starkly illustrated in the case of a single manager contracted to run two separate funds: one for a group of conservative investors who designated the 3 month T-Bill as the benchmark portfolio, and the other for a group of investors who designated the S&P 500 as the benchmark portfolio.⁶ The manager would quickly surmise that it is much easier to outperform a T-Bill benchmark than an S&P 500 benchmark, and that he/she should choose a much more conservative portfolio when managing the former portfolio in order to maximize (minimize) the probability of outperforming (underperforming) it over finite investor horizons. A priori, there is no reason for theorists to rule out the possibility that the manager acted as-if he/she used a higher coefficient of risk aversion when managing the former fund than he/she did when managing the latter. While this explanation for the different choices is unusual, it follows from our power utility criterion (11), which was *derived* from the deeper hypothesis of outperformance maximization. The conventional hypothesis that (2) is maximized *assumes rather than derives* the form of the utility (i.e. exponential), and then imposes the as yet empirically and experimentally unsupported *ad-hoc restriction* that its constant degree of risk aversion is completely exogenous to the manager's benchmark *and* the investment opportunity set used by the manager to beat it.

5 Conclusions

Mutual fund performance is often measured relative to a designated benchmark portfolio. This paper provides performance analysts with a simple way of ranking funds in accord with their respective probabilities of outperforming a benchmark portfolio. We derived a closed form for the ranking index when funds' excess log returns (over the benchmark's) are generated by time-varying Gaussian processes. More generally, the outperformance probability index is (i) a generalization of the familiar constrained minimum value of Kullback-Leibler relative entropy, and (ii) an asymptotic expected generalized power utility, which differs in two ways from the familiar power utility of wealth. First, the argument of the utility function is the ratio of wealth earned in the fund to what would have otherwise been earned from investing in the benchmark. Second and more surprising, the curvature (i.e. risk aversion) parameter value required to evaluate the expected power utility of a fund's portfolio is the value that maximizes the expected utility of that fund's portfolio! Hence the fund performance ranking index uses a power utility whose coefficient of risk aversion varies endogenously from fund to fund.

In order to illustrate the feasibility and plausibility of this approach, we derived simple non-parametric estimators for the performance ranking index and the fund-specific coefficients of risk aversion required to evaluate it. These were used to rank the performance of mutual funds that (based on standard hypothesis tests) could asymptotically outperform the S&P 500. We concluded that only 32 out of 347 funds will be able to asymptotically outperform the S&P 500, even though those 347 funds managed to survive the 19 year test period. Those that outperformed had overall equity allocations and expense ratios that were similar to those that did not. The fund-specific coefficients of risk aversion required to evaluate the relative performance of those 32 funds ranged

between 5.6 and 13.5. The highest ranked fund is Fidelity Magellan, which also had the highest coefficient of risk aversion.

These theoretical and empirical findings should benefit investors and performance analysts who want to rank funds in accord with their probabilities of outperforming a benchmark they want to beat. But for academic readers interested in fund manager *behavior*, we also formulated the hypothesis that a fund manager attempts to maximize the probability of outperforming a designated benchmark. We contrasted this hypothesis with the extant alternative hypothesis that a fund manager with a fixed coefficient of risk aversion attempts to maximize the expected utility of fund returns in excess of the benchmark. We argued that a recent empirical test of this alternative hypothesis provided little evidence in favor of it. In the absence of convincing empirical evidence favoring that alternative, the scientific desiderata of Popperian falsifiability and Ockham's Razor weigh in favor of the outperformance probability maximization hypothesis, which eliminates the assumption that the fund manager has an unobservable, econometrically free curvature (i.e. risk aversion) parameter. Moreover, if the outperformance probability maximization hypothesis is valid, a manager's degree of risk aversion is endogenous and hence will change when the manager is faced with a different benchmark or investment opportunity set. An investor or performance analyst who misspecifies the manager's degree of risk aversion as an econometrically free parameter would then be subject to the celebrated Lucas Critique.

Notes

¹The authors wish to acknowledge helpful comments from Tom Smith, Richard Heaney, Juan-Pedro Gomez, J.C. Duan, Rama Cont, and seminar participants at the Information and Entropy Econometrics Conference in Washington, D.C. , University of Minnesota, University of Colorado, CIRANO Extremal Events Conference in Montreal, French Finance Association, European Financial Management Association, Ecole Polytechnique in Paris, Australian National University, University of Melbourne, University of Queensland, University of Auckland, and University of Otago. Foster acknowledges the support of the Australian Research Council.

²Note that subtracting gross returns in (1) gives the same number as the more common subtraction of net returns.

³To obtain a total return benchmark, we used the CRSP value-weighted index, which includes distributions from the S&P 500 stocks.

⁴We also investigated the possibility that GARCH processes might have generated the excess log fund returns, in which case fitted GARCH models could be used to produce simulation estimates of (10). But stationarity testing of the popular GARCH(1,1) specification failed for all but 6 of the 32 funds.

⁵For readers interested in a detailed analysis of the theoretical and econometric problems in that paper, we have written an appendix, available upon request.

⁶This is not unrealistic; many fund managers run more than one fund.

References

- [1] Suleiman Basak and Alex Shapiro. Value-At-Risk based risk management: Optimal policies and asset prices. *Review of Financial Studies*, 14(2):371–405, 2001.
- [2] Connie Becker, Wayne Ferson, David H. Myers, and Michael J. Schill. Conditional market timing with benchmark investors. *Journal of Financial Economics*, 52(1):119–148, 1999.
- [3] T. R. Bielecki, S.R. Pliska, and M. Sherris. Risk sensitive asset allocation. *Journal of Economic Dynamics and Control*, 24:1145–1177, 2000.
- [4] Michael Brennan. Agency and asset pricing. Finance Working Paper no.6-93, Anderson Graduate School of Management, University of California, Los Angeles, 1993.
- [5] Stephen J. Brown and William N. Goetzmann. Performance persistence. *Journal of Finance*, 50(2):679–698, 1995.
- [6] Sid Browne. Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. *Finance and Stochastics*, 3:275–294, 1999.
- [7] James A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley, 1990.
- [8] John Y. Campbell, Andrew W. Lo, and Craig A. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.
- [9] Mark Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82, 1997.
- [10] Jordi Gali. Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices. *Journal of Money, Credit, and Banking*, 26(1):1–8, 1994.
- [11] Juan-Pedro Gómez and Fernando Zapatero. Asset pricing implications of benchmarking: A two-factor CAPM. *European Journal of Finance*, 2003(forthcoming).
- [12] Thomas H. Goodwin. The Information Ratio. *Financial Analysts Journal*, 54(4):34–43, 1998.

- [13] S. Grossman and J-L. Vila. Optimal dynamic trading with leverage constraints. *Journal of Financial and Quantitative Analysis*, 27(2):151–168, 1992.
- [14] S. Grossman and Z. Zhou. Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276, 1993.
- [15] W.H. Jefferys and J.O. Berger. An application of robust bayesian analysis to hypothesis testing and occam’s razor. *Journal of the Italian Statistical Society*, 1:17–32, 1992.
- [16] Paul Kaplan and Jim Knowles. The Stutzer performance index: Summary of mathematics, rationale, and behavior. Quantitative Research Department, Morningstar, Inc., Chicago, IL, March 2001.
- [17] Yuichi Kitamura and Michael Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997.
- [18] Yuichi Kitamura and Michael Stutzer. Connections between entropic and linear projections in asset pricing estimation. *Journal of Econometrics*, 107:159–174, 2002.
- [19] Narayanan R. Kocherlakota. The equity premium: It’s still a puzzle. *Journal of Economic Literature*, 34(1):42–71, 1996.
- [20] Robert E. Lucas. Econometric policy evaluation: A critique. In Karl Brunner and Allan Meltzer, editors, *The Phillips Curve and Labor Markets: Volume 1, Carnegie-Rochester Conference Series on Public Policy*. North-Holland, 1976.
- [21] Huyên Pham. A large deviations approach to optimal long term investment. *Finance and Stochastics*, 7:169–195, 2003.
- [22] Karl Popper. *The Logic of Scientific Discovery*. Routledge, 1977 (14th Printing).
- [23] Richard Roll. A mean/variance analysis of tracking error. *Journal of Portfolio Management*, 18(4):13–22, 1992.
- [24] Mark Rubinstein. Continuously rebalanced investment strategies. *Journal of Portfolio Management*, 18(1):78–81, 1991.

- [25] William Sharpe. Morningstar's risk-adjusted ratings. *Financial Analysts Journal*, 54(4):21–33, 1998.
- [26] Michael Stutzer. A portfolio performance index. *Financial Analysts Journal*, 56(3):52–61, 2000.
- [27] Michael Stutzer. Fund managers may cause their benchmarks to be priced risks. *Journal of Investment Management*, 1(3):1–13, 2003.
- [28] Michael Stutzer. Portfolio choice with endogenous utility: A large deviations approach. *Journal of Econometrics*, 116:365–386, 2003.
- [29] TIAA-CREF. TIAA-CREF Trust Company's personal touch. TIAA-CREF Investment Forum, September 2000.
- [30] Russ Wermers. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance*, 55(4):1655–1703, 2000.