



Article

The Outperformance Probability of Mutual Funds

Gabriel Frahm * and Ferdinand Huber

Chair of Applied Stochastics and Risk Management, Department of Mathematics and Statistics,
Helmut Schmidt University, 22043 Hamburg, Germany

* Correspondence: frahm@hsu-hh.de

Received: 5 May 2019; Accepted: 19 June 2019; Published: 26 June 2019



Abstract: We propose the outperformance probability as a new performance measure, which can be used in order to compare a strategy with a specified benchmark, and develop the basic statistical properties of its maximum-likelihood estimator in a Brownian-motion framework. The given results are used to investigate the question of whether mutual funds are able to beat the S&P 500 or the Russell 1000. Most mutual funds that are taken into consideration are, in fact, able to beat the market. We argue that one should refer to differential returns when comparing a strategy with a given benchmark and not compare both the strategy and the benchmark with the money-market account. This explains why mutual funds often appear to underperform the market, but this conclusion is fallacious.

Keywords: exchange traded funds; inverse coefficient of variation; mutual funds; outperformance probability; performance measurement; Sharpe ratio

JEL Classification: G11; G19

1. Motivation

The total value of assets under management in open-end funds was 46.7 trillion US dollars at the end of 2018 (ICI 2019). Nowadays, investors and investment advisors have access to an abundant number of performance measures in order to compare investment vehicles with each other. Most performance measures are based on the seminal work of Jensen (1968); Lintner (1965); Mossin (1966); Sharpe (1964 1966); Treynor (1961). The best-known performance measure is the Sharpe ratio, which divides the expected excess return on investment by the standard deviation of the excess return. Most other performance measures that can be found in the literature are based on the same principle, i.e., they divide the return on investment by its risk, where the precise meaning of “return” and “risk” differs from one performance measure to another. A typical example is the expected excess return to value at risk, which goes back to Alexander and Baptista (2003) as well as Dowd (2000). Other examples are the conditional Sharpe ratio, in which case one chooses the expected shortfall, the tail conditional expectation, or similar downside risk measures as a denominator (Baweja et al. 2015; Chow and Lai 2015). The number of possible combinations of numerators and denominators is practically unbounded. Depending on whether one chooses either the expected excess return or the first higher partial moment as a numerator and some lower partial moment, drawdown measure, or some value-at-risk measure as a denominator, the resulting performance measure is called Omega (Keating and Shadwick 2002), Sortino ratio (Sortino et al. 1999), Calmar ratio (Young 1991), Sterling ratio (Kestner 1996), or Burke ratio (Burke 1994), etc. A nice overview of those performance measures can be found in Table 1 of Eling and Schuhmacher (2006). However, it is evident that any list of return-to-risk measures must be far from exhaustive.

In this work, we propose a new performance measure called *outperformance probability* (OP). It differs in many aspects from return-to-risk measures:

1. The OP compares some strategy with a specified benchmark, which need not necessarily be the money-market account.
2. It is a probability. Thus, it is easy to understand also for a nonacademic audience, more precisely, for people who are not educated in statistics or probability theory.
3. The holding period of the investor is considered random. This enables us to compute the performance of an investment opportunity for arbitrary liquidity preferences.

Ad 1: Most financial studies that make use of performance measures compare, e.g., the Sharpe ratio of some strategy, according to Sharpe (1966), with the Sharpe ratio of a benchmark. The Sharpe ratios measure the outperformance both of the strategy and of the benchmark with respect to the money-market account, i.e., the riskless asset. Hence, one needs to calculate *two* performance measures to evaluate the given strategy. By contrast, our approach is based on the idea of analyzing differential returns, not necessarily excess returns.¹ This idea goes back to Sharpe (1994). To be more precise, since we do not use the money market as an anchor point, we need only *one* measure in order to compare two different investment opportunities. We will see that this sheds a completely different light on the question of whether or not one should prefer actively managed funds to passively managed funds. This fundamental question is still discussed in the finance literature and it is a commonplace that most fund managers are not able to beat their benchmarks after accounting for all management fees and agency costs. We will come back to this point in Section 3.1. However, our empirical study reveals that most actively managed funds, in fact, are able to outperform their benchmarks. The reason for this observation might be explained as follows: The difference between the performance of two investment opportunities is not the same as the performance of one investment opportunity compared to another. Simply put, in general, performance measures are not linear. Let R_S be the return on some strategy, R_B be the return on a benchmark, and r be the risk-free interest rate after some holding period. Further, let π be any performance measure. Hence, the performance of the strategy with respect to the money-market account is $\pi(R_S - r)$, whereas $\pi(R_B - r)$ is the corresponding performance of the benchmark. Now, the problem is that, in general,

$$\pi(R_S - r) - \pi(R_B - r) \neq \pi((R_S - r) - (R_B - r)) = \pi(R_S - R_B).$$

We suggest to calculate $\pi(R_S - R_B)$ in order to understand whether or not the strategy is better than its benchmark and not to compare $\pi(R_S - r)$ with $\pi(R_B - r)$. By comparing $\pi(R_S - r)$ with $\pi(R_B - r)$, one investigates the question of which of the two investment opportunities is better able to outperform the *money-market account*. However, the fact that one investment opportunity is better able to outperform the riskless asset than another investment opportunity does *not* imply that the first investment opportunity outperforms the second one. This observation is crucial and we will provide an analytical example in Section 2.4.

Ad 2: We do not claim that the OP is more practicable than other performance measures. Indeed, given the computational capacities that are nowadays available, it is quite easy to compute any other performance measure as well. Instead, we refer to a social problem that is frequently discussed in statistical literacy, namely that a major part of our population has no clear understanding of basic statistics. Hence, most people cannot comprehend the precise meaning of “expected value,” “standard deviation,” or “variance,” etc. Presumably, everybody of us knows that it is hard to explain the difference between the sample mean and the expected value to the layman. The same holds true regarding the empirical variance, sample variance, and population variance, etc. Further, our own experience shows that, at the beginning of their studies, many finance students have problems to understand the distinction between an ex-ante and an ex-post performance measure, which can be

¹ Each excess return is a differential return but not vice versa.

attributed to the fact that they are struggling with statistics and probability theory.² Thus, we are especially concerned about the problem of *financial* literacy, not only regarding an academic audience. According to the OECD (2016), the “overall levels of financial literacy, indicated by combining scores on knowledge, attitudes and behaviour are relatively low” in all 30 countries that participated in the OECD study. More precisely, it is reported that “only 42% of adults across all participating countries and economies are aware of the additional benefits of interest compounding on savings” and that “only 58% could compute a percentage to calculate a simple interest on savings.” Moreover, “only about two in three adults [...] were aware that it is possible to reduce investment risk by buying a range of different stocks” and, in some countries, “no more than half of respondents understood the financial concept of diversification.” The OECD study reveals a very low level of numeracy: Many people even are not able to calculate the balance of an account after an interest payment of 2%. Similar results can be found in the comprehensive review article by Lusardi and Mitchell (2013). For example, many people cannot say whether single stocks or mutual funds are more risky.³ In the light of these findings, we conclude that it is clearly impossible for most people in our population to comprehend the core message of a return-to-risk measure. The OP is a probability and thus it can be much better understood by a nonacademic audience—at least in an intuitive way. Hence, it might help bridging theory and practice.

Ad 3: Performance measures usually presume a fixed holding period. It is well-known that a given performance ratio cannot be extended to other holding periods without making any simplifying—and sometimes quite problematic—assumptions, which are frequently discussed in the literature see, e.g., Lin and Chou (2003); Lo (2002); Sharpe (1994). For example, suppose that the value process $\{S_t\}_{t \geq 0}$ of some asset follows a geometric Brownian motion with drift coefficient $\mu \in \mathbb{R}$ and diffusion coefficient $\sigma > 0$.⁴ We may assume without loss of generality that the instantaneous risk-free interest rate per year is zero. Hence, the (excess) return on the asset at time $T > 0$ is given by

$$R = \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) T + \sigma W_T \right) - 1,$$

which means that the expected return is $E(R) = e^{\mu T} - 1$. Further, the variance of the return amounts to $\text{Var}(R) = (e^{\sigma^2 T} - 1) e^{2\mu T}$. We conclude that the Sharpe ratio of the strategy is

$$\text{Sh} = \frac{1 - e^{-\mu T}}{\sqrt{e^{\sigma^2 T} - 1}}.$$

Figure 1 depicts the Sharpe ratio of two different assets depending on the holding period T . Asset 1 possesses the parameters $\mu_1 = 0.1$ and $\sigma_1 = 0.2$, whereas Asset 2 has the parameters $\mu_2 = 0.2$ and $\sigma_2 = 0.3$. We can see that the Sharpe ratio is essentially determined by the time of liquidation. Interestingly, the Sharpe ratio is not linear and even not monotonic in time. If we use the Sharpe ratio as a performance measure, it can happen that we prefer Asset 1 for shorter holding periods but Asset 2 for longer holding periods.⁵ Hence, the optimal choice between Asset 1 and Asset 2 heavily depends on the investment horizon. That is, in order to give a clear recommendation, we must know the holding period of the investor. However, most investors do not liquidate their strategies after a fixed period of time. More precisely, the time of liquidation is not known to them in advance, which means that it represents a random variable. We take the individual liquidity preference of the investor into

² It is a matter of fact that some students do not understand even the distinction between parameter and estimator after attending a statistics course.

³ According to Lusardi and Mitchell (2013, pp. 15–16), we can expect that a large number of people cannot understand the question at all because they are unfamiliar with stocks, bonds, and mutual funds.

⁴ From now on, we will omit the subscript “ $t \geq 0$ ” for notational convenience.

⁵ In the given example, the critical time point is, approximately, $T = 5$ years.

account by specifying a holding-time distribution. Our approach is time-continuous, whereas most other performance measures are based on a one-period model. This clearly distinguishes the OP from other performance measures and we think that this is one of our main contributions to the literature.

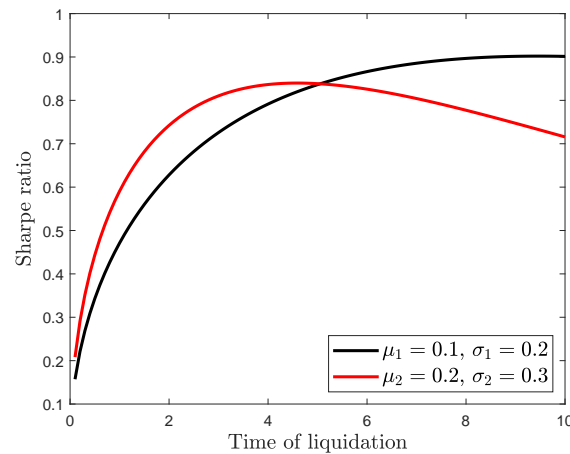


Figure 1. Sharpe ratios as a function of time.

In order to test the OP of different trading strategies, we selected 10 actively managed mutual funds that try to beat the S&P 500 or the Russell 1000 stock-market index. We decided to choose these funds because they focus on growth stocks in the US, have a large amount of assets under management, and their issuers enjoy a good reputation. The question of whether or not it is worth investing in actively managed funds at all is essential because if they do not outperform their benchmarks, market participants might be better advised to buy exchange traded funds (ETFs). More precisely, they should prefer so-called index ETFs, which aim at tracking a stock-market index. Analogously, if it turns out that the mutual funds are not able to outperform the bond market, it could be better to buy some US treasury-bond ETFs.⁶ Further, it could happen that the money-market account is preferable and, in the worst case, it is even better to refrain from financial markets at all.

What are the merits of the OP compared to other performance measures?

1. The first one is conceptual:

- (a) A simple thought experiment reveals that comparing two performance measures with one another, where each one compares an investment opportunity with the money-market account, can lead to completely different conclusions than evaluating only one performance measure that compares the given investment opportunities without taking the money-market account into consideration at all. The former comparison refers to the question of which of the two investment opportunities is better able to outperform the money-market account, whereas the latter comparison refers to the question of whether the first investment opportunity is able to outperform the second. In general, performance measures are not linear and so the former comparison does not imply the latter, i.e., an investment opportunity that is better able to produce excess returns than another investment opportunity need not be better than the other.
- (b) Most performance measures presume that the holding period of the investor is fixed. This assumption is clearly violated in real-life because investors usually do not know, in advance, when they will liquidate all their assets. We solve this problem by incorporating any holding-time distribution, which specifies the individual liquidity preference of the

⁶ The reason why we focus on ETFs is discussed in Section 3.1.

investor. It can be either discrete or continuous and it can have a finite or infinite right endpoint. Any fixed holding period can be considered a special case, which means that we can treat one-period models, too.

2. The second one is empirical:

- (a) The natural logarithm of the assets under management of the mutual funds that are taken into consideration are highly correlated with their inverse coefficient of variation (ICV). The ICV is a return-to-risk measure, which is based on differential log-returns but not (necessarily) on excess log-returns, and in the Brownian-motion framework it is the main ingredient of the OP. This means that capital allocation and relative performance are strongly connected to one another, which suggests that market participants take differential (log-)returns *implicitly* into account when making their investment decisions.
- (b) We emphasize our results by comparing the p -values of the differences between the Sharpe ratios of all mutual funds and the Sharpe ratios of the given benchmarks. The p -values indicate that it is hard to distinguish between the former and the latter Sharpe ratios. For this reason, we cannot say that any fund is better than the benchmark by comparing two Sharpe ratios with one another. A completely different picture evolves when considering the p -values of the ICVs of all funds with respect to their benchmarks. Those p -values are much lower and *economically* significant, i.e., it turns out that most funds are able to beat their benchmarks.

The rest of this work is organized as follows: In Section 2 we present and discuss the OP. To be more precise, Section 2.1 contains our basic assumptions and our general definition of the OP. In Section 2.2 we investigate its theoretical properties, whereas in Section 2.3 we derive the statistical properties of our maximum-likelihood (ML) estimator for the OP. Section 2.4 contains a general discussion of the OP compared to other performance measures. Further, Section 3 contains the empirical part of this work, in which we investigate the question of whether or not mutual funds are able to beat their benchmarks. More precisely, in Section 3.1 we discuss some general observations related to the performance of mutual funds and index ETFs, whereas Section 3.2 contains our empirical results for different holding-time distributions. Our conclusion is given in Section 4.

2. The Outperformance Probability

2.1. Basic Assumptions and Definition

Let $\{S_t\}$ be the value process of some trading strategy and $\{B_t\}$ be the value process of a benchmark. We implicitly assume that $\{S_t\}$ and $\{B_t\}$ are, almost surely, positive. The strategy starts at time $t = 0$ and stops at some *random* time $T > 0$. Hence, T can be considered the time of liquidation and is referred to as the holding period. Throughout this work, time is measured in years. Moreover, we suppose that $S_0 = B_0 = 1$ without loss of generality.

Definition 1 (Outperformance probability). *The OP is defined as $\Pi = P(S_T > B_T)$.*

Hence, Π is the probability that the value of the strategy will be greater than the value of the benchmark when the investor liquidates his strategy. The time of liquidation, T , is considered a random variable and so the law of total probability leads us to

$$\Pi = \mathbf{E}_T \left[P(S_t > B_t \mid T = t) \right].$$

The performance measure Π ranges between 0 and 1. In the case that $\Pi = 1$, the strategy is always better than the benchmark, where “always” means “with probability 1,” i.e., almost surely. This can be considered a limiting case, which usually does not appear in real life because otherwise

the investor would have a weak arbitrage opportunity (Frahm 2016).⁷ By contrast, if we have that $\Pi = 0$, the benchmark turns out to be always *at least as good* as the strategy, which does not mean that the benchmark is always *better*.⁸ More generally, if the strategy outperforms its benchmark with probability Π , the benchmark outperforms the strategy with probability *less than or equal to* $1 - \Pi$. Therefore, the OP is not symmetric. Finally, if Π is near 0.5, there is no clear recommendation in favor or against the strategy. However, if a fund manager claims to be better than the benchmark, we should expect Π to be greater than 0.5.

We make the basic assumption that the time of liquidation does not depend on whether or not the strategy outperforms the benchmark, viz.

$$P(S_t > B_t | T = t) = P(S_t > B_t), \quad \forall t \geq 0.$$

Put another way, the decision of the investor whether or not to terminate the strategy at time t does not depend on its performance relative to the benchmark. It follows that

$$\Pi = \mathbf{E}_T[P(S_T > B_T)] = \int_0^\infty P(S_t > B_t) dF(t),$$

where F represents the cumulative distribution function of the holding period T .

2.2. Theoretical Properties

In this work, we rely on the standard model of financial mathematics, i.e., we suppose that $\{S_t\}$ and $\{B_t\}$ follow a 2-dimensional geometric Brownian motion. This means that the stochastic processes obey the stochastic differential equations

$$\frac{dS_t}{S_t} = \mu_S dt + \sigma_S dW_{S_t} \quad \text{and} \quad \frac{dB_t}{B_t} = \mu_B dt + \sigma_B dW_{B_t},$$

where $\mathbf{E}(dW_{S_t}dW_{B_t}) = \rho dt$ with $-1 \leq \rho \leq 1$. It follows that

$$\begin{bmatrix} \log(S_t) \\ \log(B_t) \end{bmatrix} \sim \mathcal{N}(\gamma t, \Sigma t), \quad \forall t \geq 0,$$

with

$$\gamma = \begin{bmatrix} \gamma_S \\ \gamma_B \end{bmatrix} = \begin{bmatrix} \mu_S - \sigma_S^2/2 \\ \mu_B - \sigma_B^2/2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_S^2 & \sigma_S \sigma_B \rho \\ \sigma_B \sigma_S \rho & \sigma_B^2 \end{bmatrix}.$$

The numbers γ_S and γ_B are called the growth rates of $\{S_t\}$ and $\{B_t\}$, respectively (Frahm 2016). Thus, we have that

$$X := \log\left(\frac{S_1}{B_1}\right) \sim \mathcal{N}\left(\mu_S - \mu_B - \frac{\sigma_S^2 - \sigma_B^2}{2}, \sigma_S^2 + \sigma_B^2 - 2\sigma_S \sigma_B \rho\right).$$

Hence, the random variable X represents the *relative* log-return on the strategy with respect to its benchmark after the first period of time. We implicitly assume that

$$\sigma_S^2 + \sigma_B^2 - 2\sigma_S \sigma_B \rho > 0$$

⁷ More precisely, the strategy would *dominate* the benchmark in the sense of Merton (1973).

⁸ $P(X > Y) = 0$ does not imply that $P(X < Y) = 1$ but $P(X \leq Y) = 1$.

in order to avoid the case in which X is degenerate, i.e., constant with probability 1. In our Brownian-motion framework, this happens if and only if $\Pi \in \{0, 1\}$.

It follows that

$$P(S_t > B_t) = P(X > 0) = \Phi \left(\sqrt{t} \cdot \frac{\mu_S - \mu_B - (\sigma_S^2 - \sigma_B^2)/2}{\sqrt{\sigma_S^2 + \sigma_B^2 - 2\sigma_S\sigma_B\rho}} \right),$$

where Φ is the cumulative distribution function of the standard normal distribution. This leads us to

$$\Pi = \int_0^\infty \Phi \left(\sqrt{t} \cdot \frac{\mu_S - \mu_B - (\sigma_S^2 - \sigma_B^2)/2}{\sqrt{\sigma_S^2 + \sigma_B^2 - 2\sigma_S\sigma_B\rho}} \right) dF(t),$$

which demonstrates that the OP essentially depends on the diffusion coefficient σ_S and not only on the drift coefficient μ_S of the considered strategy.

For example, if we choose the money-market account as a benchmark, we obtain $\mu_B = r$ and $\sigma_B = 0$, where r is the instantaneous risk-free interest rate per year. In this standard case, the OP amounts to

$$\Pi = \int_0^\infty \Phi \left(\sqrt{t} \cdot \frac{\mu_S - r - \sigma_S^2/2}{\sigma_S} \right) dF(t).$$

In general, the benchmark is risky and we can see that the inverse coefficient of variation of the strategy with respect to its benchmark, i.e.,

$$\text{ICV} := \frac{\mathbf{E}(X)}{\mathbf{Std}(X)} = \frac{\mu_S - \mu_B - (\sigma_S^2 - \sigma_B^2)/2}{\sqrt{\sigma_S^2 + \sigma_B^2 - 2\sigma_S\sigma_B\rho}},$$

plays a crucial role when calculating the OP. Note that the ICV refers to the *differential* log-return $X = \log(S_1) - \log(B_1)$, not to the excess log-return $\log(S_1) - r$.

Now, we are ready for our first theorem. Its proof has already been given throughout the previous explanations and so it can be skipped.

Theorem 1 (Outperformance probability). *Under the aforementioned assumptions, the OP is given by*

$$\Pi = \int_0^\infty \Phi \left(\sqrt{t} \text{ICV} \right) dF(t). \tag{1}$$

Hence, Π is strictly increasing in ICV. This means that the higher the ICV the higher the OP.⁹ However, under the given assumptions, the OP can never be 0 or 1.

For example, suppose that the holding period is *fixed* and equal to $T > 0$. In this case, the OP simply amounts to $\Pi = \Phi(\sqrt{T} \text{ICV})$. By contrast, if the holding period is uniformly distributed between 0 and $M > 0$, we obtain

$$\Pi = \frac{1}{M} \int_0^M \Phi \left(\sqrt{t} \text{ICV} \right) dt.$$

Further, if the holding period is exponentially distributed with parameter $\lambda > 0$, we have that

$$\Pi = \lambda \int_0^\infty \Phi \left(\sqrt{t} \text{ICV} \right) e^{-\lambda t} dt,$$

⁹ Thus, if a strategy has a higher OP than another strategy, given some holding-time distribution, the same holds true for any other holding-time distribution.

and if we assume that T has a Weibull distribution with parameters $\kappa > 0$ and $\gamma > 0$, we obtain

$$\Pi = \frac{\kappa}{\gamma} \int_0^\infty \Phi(\sqrt{t} \text{ICV}) \left(\frac{t}{\gamma}\right)^{\kappa-1} e^{-(t/\gamma)^\kappa} dt.$$

Note that for $\kappa = 1$ and $\gamma = \lambda^{-1}$, the Weibull distribution turns into an exponential distribution with parameter λ . It is evident that the spectrum of possibilities to capture the distribution of T is almost infinite. In this work, we restrict to the aforementioned holding-time distributions. Obviously, F need not be continuous or even differentiable. The first holding-time distribution, which asserts that T is a constant, is deterministic, which poses no problem at all, too.

Figure 2 depicts the densities of the uniform, exponential, and Weibull distribution for different parameterizations. We have chosen all parameters such that the mean holding period equals 5 years. Moreover, Figure 3 contains some curves that quantify the OP depending on the ICV. On the left-hand side we can see how Π depends on ICV if the holding period T is considered a constant, which is the usual assumption in the finance literature, whereas on the right-hand side it is assumed that T is uniformly distributed between 0 and M , i.e., the maximum number of years that the investor maintains the given strategy.

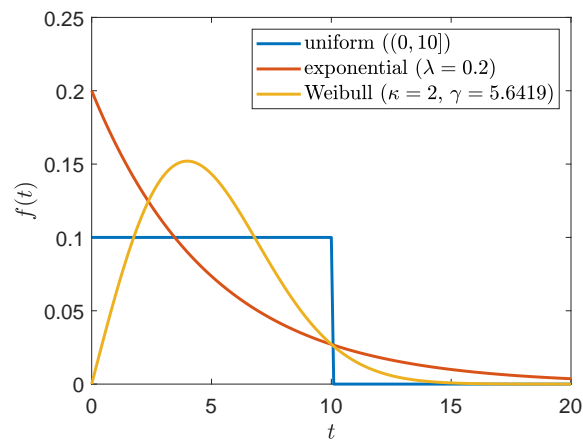


Figure 2. Holding-time distributions with mean 5.

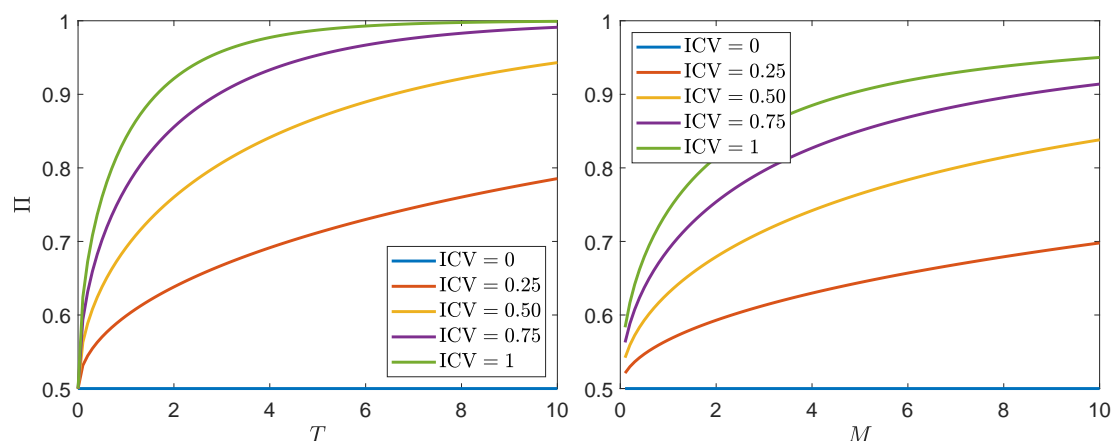


Figure 3. Outperformance probability (OP) depending on the inverse coefficient of variation (ICV) and the (maximal) holding period. On the **left-hand** side, the holding period T is considered fixed, whereas on the **right-hand** side, T is supposed to be uniformly distributed between 0 and M years.

2.3. Statistical Inference

Here, we propose a parametric estimator for Π , based on the assumption that the value processes $\{S_t\}$ and $\{B_t\}$ follow a 2-dimensional geometric Brownian motion. Let

$$X_t = \log\left(\frac{S_t}{S_{t-\Delta}}\right) - \log\left(\frac{B_t}{B_{t-\Delta}}\right), \quad t = \Delta, 2\Delta, \dots, n\Delta,$$

be the difference between the log-return on the strategy and the log-return on the benchmark at time t . The period between $t - \Delta$ and t , i.e., $\Delta > 0$, is fixed. Throughout this work, we assume that one year consists of 252 trading days and so we have that $\Delta = 1/252$.

The ML-estimator for the ICV is

$$\widehat{\text{ICV}}_n = \frac{1}{\sqrt{\Delta}} \frac{\frac{1}{n} \sum_{t=\Delta}^{n\Delta} X_t}{\sqrt{\frac{1}{n} \sum_{t=\Delta}^{n\Delta} (X_t - \frac{1}{n} \sum_{t=\Delta}^{n\Delta} X_t)^2}}.$$

Our first proposition asserts that the ML-estimator for the ICV is strongly consistent.

Proposition 1. *We have that*

$$\widehat{\text{ICV}}_n \xrightarrow{\text{a.s.}} \text{ICV}, \quad n \rightarrow \infty.$$

Proof. The strong law of large numbers implies that

$$\frac{1}{n} \sum_{t=\Delta}^{n\Delta} X_t \xrightarrow{\text{a.s.}} \gamma_S - \gamma_B \quad \text{and} \quad \frac{1}{n} \sum_{t=\Delta}^{n\Delta} \left(X_t - \frac{1}{n} \sum_{t=\Delta}^{n\Delta} X_t\right)^2 \xrightarrow{\text{a.s.}} \sigma_S^2 + \sigma_B^2 - 2\sigma_S\sigma_B\rho$$

as $n \rightarrow \infty$. Hence, the continuous mapping theorem immediately reveals that $\widehat{\text{ICV}}_n$ converges almost surely to ICV as the number of observations, n , grows to infinity. \square

The next proposition refers to the asymptotic distribution of $\sqrt{n}(\widehat{\text{ICV}}_n - \text{ICV})$.

Proposition 2. *We have that*

$$\sqrt{n}(\widehat{\text{ICV}}_n - \text{ICV}) \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{1}{\Delta} + \frac{\text{ICV}^2}{2}\right), \quad n \rightarrow \infty.$$

Proof. The ICV can be treated like a Sharpe ratio and thus $1 + \text{ICV}^2/2$ is the asymptotic variance of $\sqrt{n}(\widehat{\text{ICV}}_n - \text{ICV})$ in the case that $\Delta = 1$ (Frahm 2018, p. 6). \square

Hence, provided that the sample size n is large, the standard error of $\widehat{\text{ICV}}_n$ is

$$\text{Std}(\widehat{\text{ICV}}_n) \approx \sqrt{\frac{1}{n\Delta} + \frac{\text{ICV}^2}{2n}}.$$

Our empirical study in Section 3 reveals that the ICV does not exceed 1 and, since the sample size n is large, we can ignore the term $\text{ICV}^2/(2n)$.¹⁰ Note also that $m := n\Delta$ corresponds to the sample length measured in years. Thus, we obtain the following nice rule of thumb:

$$\text{Std}(\widehat{\text{ICV}}_n) \approx \frac{1}{\sqrt{m}}. \tag{2}$$

¹⁰ This term can be attributed to the estimation error regarding the standard deviation of X (Frahm 2018).

The ML-estimator for Π is obtained, in the usual way, by substituting ICV with $\widehat{\text{ICV}}_n$.

Definition 2 (ML-estimator). *The ML-estimator for Π is given by*

$$\widehat{\Pi}_n = \int_0^\infty \Phi(\sqrt{t} \widehat{\text{ICV}}_n) dF(t).$$

The following theorem asserts that $\widehat{\Pi}_n$ is strongly consistent.

Theorem 2. *We have that*

$$\widehat{\Pi}_n \xrightarrow{\text{a.s.}} \Pi, \quad n \rightarrow \infty.$$

Proof. From Proposition 1 we already know that $\widehat{\text{ICV}}_n$ converges almost surely to ICV as $n \rightarrow \infty$. According to Equation (1), we have that

$$\Pi: \text{ICV} \mapsto \int_0^\infty \Phi(\sqrt{t} \text{ICV}) dF(t)$$

and it is clear that Π is continuous in ICV. Hence, the continuous mapping theorem reveals that $\widehat{\Pi}_n$ converges almost surely to Π as the number of observations, n , grows to infinity. \square

The next theorem provides the asymptotic distribution of $\sqrt{n}(\widehat{\Pi}_n - \Pi)$, which can be used in order to calculate confidence intervals or conduct hypotheses tests. For this purpose, we need only assume that the expected value of \sqrt{T} is finite.¹¹ In fact, in most practical applications we even have that $\mathbf{E}(T) < \infty$, which holds true also for all holding-time distributions that are considered in this work. Hence, the moment condition is always satisfied.

Theorem 3. *Assume that $\mathbf{E}(\sqrt{T}) < \infty$. We have that*

$$\sqrt{n}(\widehat{\Pi}_n - \Pi) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad n \rightarrow \infty,$$

with

$$\sigma^2 = \left(\int_0^\infty \sqrt{t} \phi(\sqrt{t} \text{ICV}) dF(t) \right)^2 \left(\frac{1}{\Delta} + \frac{\text{ICV}^2}{2} \right),$$

where ϕ is the probability density function of the standard normal distribution.

Proof. Note that

$$0 \leq \frac{\Phi(\sqrt{t} x_1) - \Phi(\sqrt{t} x_0)}{x_1 - x_0} \leq \frac{d\Phi(\sqrt{t} x)}{dx} \Big|_{x=0} = \sqrt{t} \phi(0)$$

for all $x_0 < x_1$ and $t \geq 0$. Further, we have that

$$\int_0^\infty \sqrt{t} \phi(0) dF(t) = \phi(0) \mathbf{E}(\sqrt{T}) < \infty.$$

Thus, we can apply the dominated convergence theorem, i.e.,

$$\frac{d\Pi(\text{ICV})}{d\text{ICV}} = \frac{d}{d\text{ICV}} \int_0^\infty \Phi(\sqrt{t} \text{ICV}) dF(t) = \int_0^\infty \sqrt{t} \phi(\sqrt{t} \text{ICV}) dF(t).$$

¹¹ This is true whenever $\mathbf{E}(T^k) < \infty$ for any $k \geq \frac{1}{2}$.

Now, by applying the delta method and using Proposition 2, we obtain

$$\sqrt{n} (\hat{\Pi}_n - \Pi) \xrightarrow{d} \left(\int_0^\infty \sqrt{t} \phi(\sqrt{t} \text{ICV}) dF(t) \right) \mathcal{N} \left(0, \frac{1}{\Delta} + \frac{\text{ICV}^2}{2} \right), \quad n \rightarrow \infty,$$

which leads us to the desired result. \square

Hence, provided that n is sufficiently large, we can apply Theorem 3 in order to approximate the standard error of $\hat{\Pi}_n$ by

$$\text{Std}(\hat{\Pi}_n) \approx \frac{1}{\sqrt{m}} \int_0^\infty \sqrt{t} \phi(\sqrt{t} \widehat{\text{ICV}}_n) dF(t).$$

Figure 4 quantifies the standard error of $\hat{\Pi}_n$ as a function of m , i.e., the years of observation, for different ICVs. Note that $\text{Std}(\hat{\Pi}_n)$ is an even function of ICV and so we may focus on $\text{ICV} \geq 0$. On the left-hand side it is assumed that T equals 5 years, whereas on the right-hand side T is supposed to be uniformly distributed between 0 and 10 years. Obviously, the specific choice of the holding-time distribution is not essential for the standard errors. They are high even if the sample length, m , is big. This is a typical problem of performance measurement (Frahm 2018). Further, we can see that the standard error of the ML-estimator essentially depends on the ICV. To be more precise, the higher the ICV the lower the standard error.

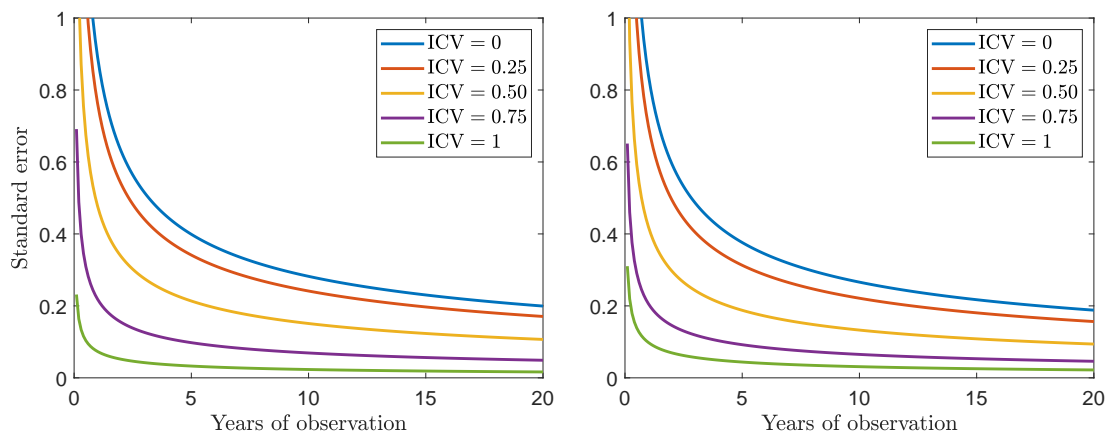


Figure 4. Standard error of $\hat{\Pi}_n$ depending on the ICV. On the **left-hand** side, the holding period T is considered fixed and is equal to 5 years, whereas on the **right-hand** side, T is supposed to be uniformly distributed between 0 and 10 years.

A typical null hypothesis is $H_0: \Pi \leq \Pi_0 \in (0, 1)$, which can be tested by means of the p -value

$$p_n = \Phi \left(- \frac{\hat{\Pi}_n - \Pi_0}{\text{Std}(\hat{\Pi}_n)} \right).$$

More precisely, given some significance level α , H_0 can be rejected if and only if $p_n < \alpha$, in which case Π turns out to be significantly larger than Π_0 .

2.4. Discussion

Classical performance measures are based on the Capital Asset Pricing Model (CAPM). The most prominent performance measures are the Sharpe ratio (Sharpe 1966), the Treynor ratio (Treynor 1961), and Jensen’s Alpha (Jensen 1968). Let R be the discrete return on some strategy after a predefined period of time. Further, let r be the discrete risk-free interest rate after the given holding period and R_M be the return on the market portfolio.

Consider the linear regression equation

$$R - r = \alpha + \beta(R_M - r) + \varepsilon,$$

where α is the Alpha, β is the Beta, and ε represents the unsystematic risk of the strategy. Further, let σ_ε^2 be the variance of ε .

An actively managed fund aims at beating its benchmark. Let us suppose that the benchmark is the market portfolio. If the fund manager performs a strategy with $\beta = 1$, we have that

$$R - r = \alpha + (R_M - r) + \varepsilon$$

and so we obtain the differential return $R - R_M = \alpha + \varepsilon$. Hence, if we want to know whether or not the fund is able to beat the market portfolio, we should analyze the differential return $R - R_M$ and not the excess return $R - r$. More precisely, we should calculate the *generalized* Sharpe ratio according to Sharpe (1994), i.e.,

$$\text{Sh}^{1994} = \frac{\mathbf{E}(R - R_M)}{\mathbf{Std}(R - R_M)} = \frac{\alpha}{\sigma_\varepsilon},$$

and not the *ordinary* Sharpe ratio (Sharpe 1966), i.e.,

$$\text{Sh}^{1966} = \frac{\mathbf{E}(R) - r}{\mathbf{Std}(R)},$$

which is based on the CAPM (Lintner 1965; Mossin 1966; Sharpe 1964).

What happens if the fund's Beta does not equal 1? In this case we should compare the fund with a benchmark strategy that possesses the Beta of the fund. Let β be the proportion of equity that is invested in the market portfolio and $1 - \beta$ be the proportion of equity that is deposited in the money market. The corresponding (benchmark) return is thus

$$R_B = \beta R_M + (1 - \beta)r = r + \beta(R_M - r).$$

Hence, we obtain the differential return $R - R_B = \alpha + \varepsilon$, which leads us to the same result as before, i.e., $\text{Sh}^{1994} = \alpha/\sigma_\varepsilon$.

Why is it so important to consider differential returns instead of excess returns if we want to understand whether or not a given strategy is able to outperform its benchmark? The most simple illustration goes like this: Consider two random variables X and Y . The probability that X is positive is $P(X > 0)$ and the probability that Y is positive is $P(Y > 0)$. Suppose that we want to know the probability that X exceeds Y . Then we would calculate $P(X > Y) = P(X - Y > 0)$ but not $P(X > 0) - P(Y > 0)$ and the same holds true for performance measurement.

This means that in order to judge whether or not a strategy is able to outperform its benchmark we should not compare one ordinary Sharpe ratio with another. Ordinary Sharpe ratios are calculated on the basis of excess returns and so they answer the question of how good a given investment opportunity is able to beat the *money-market account*. Analogously, by comparing $P(X > 0)$ with $P(Y > 0)$ we answer the question of whether it is more probable that X or that Y exceeds 0. However, in performance measurement we want to know whether or not a strategy is able to outperform its benchmark, which is a completely different question and it has nothing to do with the money-market account. Analogously, we have to calculate the probability that X exceeds Y and not compare $P(X > 0)$ with $P(Y > 0)$.

This shall be illustrated, in financial terms, by a simple thought experiment. Suppose, for the sake of simplicity but without loss of generality, that the risk-free interest rate is zero. Further, let us assume that the return on some mutual fund is $R = R_M + \varepsilon$, where ε is a positive random variable that

is uncorrelated with R_M .¹² Hence, the fund *always* yields a better return than the market portfolio and so it is clearly preferable. The ordinary Sharpe ratio of the market portfolio is

$$Sh_M^{1966} = \frac{\mu_M}{\sigma_M},$$

where $\mu_M > 0$ is the expected return on the market portfolio and $\sigma_M > 0$ represents the standard deviation of the return on the market portfolio. Thus, we have that $Sh_M^{1966} > 0$. By contrast, the ordinary Sharpe ratio of the strategy is

$$Sh_S^{1966} = \frac{\mu_M + \mu_\varepsilon}{\sqrt{\sigma_M^2 + \sigma_\varepsilon^2}},$$

where $\mu_\varepsilon > 0$ is the expected value and $\sigma_\varepsilon^2 > 0$ is the variance of ε . We conclude that also Sh_S^{1966} is positive.

Now, depending on μ_ε and σ_ε^2 , we can construct situations in which the fund appears to be better than the market portfolio and other situations in which it appears to be worse. To be more precise, we can find a positive random variable ε with a fixed expectation and an arbitrarily high variance. For example, suppose that $\varepsilon = \gamma + \delta\zeta$, where $\gamma, \delta > 0$ and ζ is a random variable such that $P(\zeta = 0) = 1 - p$ and $P(\zeta = 1/p) = p > 0$. Hence, we have that $E(\varepsilon) = \gamma + \delta$ and $Var(\varepsilon) = \delta^2(1 - p)/p$, which means that $Sh_S^{1966} \searrow 0$ as $p \searrow 0$. Put another way, it holds that $Sh_S^{1966} < Sh_M^{1966}$ if we make p sufficiently low. Nonetheless, the OP of the fund is

$$\Pi = P(R > R_M) = P(\varepsilon > 0) = 1,$$

which means that the fund outperforms the market portfolio almost surely! It is evident that in such a case every rational investor should prefer the fund.

Why do we come to such a misleading conclusion by comparing the two ordinary Sharpe ratios with one another? The reason is that we analyze the marginal distributions of R and R_M instead of analyzing their joint distribution. In fact, since ε is positive, it is better for the investor to have a large variance σ_ε^2 , not a small one. Thus, it makes no sense to penalize the expected return on the fund by σ_ε^2 , but this is precisely what we do when we calculate Sh_S^{1966} . However, it is worth noting that, in this particular case, it makes no sense either to calculate the generalized Sharpe ratio

$$Sh_S^{1994} = \frac{E(R - R_M)}{Std(R - R_M)} = \frac{\mu_\varepsilon}{\sigma_\varepsilon}.$$

The problem is that ε is positive, which means that σ_ε is the wrong risk measure in that context.¹³ However, irrespective of the parameters μ_ε and σ_ε , we always have that $\Pi = 1$, which clearly indicates that the fund is better than its benchmark. We could imagine plenty of other examples that are less striking, but the overall conclusion remains the same: One should use differential (log-)returns instead of excess (log-)returns when comparing some strategy with a benchmark.

The main goal of this work is to provide a performance measure that is better understandable for a nonacademic audience. The OP is based on the ICV, which compares the log-return on the strategy with the log-return on the benchmark. Thus, it is compatible with Sharpe's (Sharpe 1994) principal argument and so we hope that it can serve its purpose in the finance industry. Most investors want to know whether they can beat a risky benchmark, not the money-market account. It is worth emphasizing that the OP is based on a continuous-time model and we do not make any use of capital market theory. For this reason, the OP cannot be compared with performance measures that are based

¹² This is not a linear regression equation because $E(\varepsilon) > 0$.

¹³ This kind of situation cannot happen in our Brownian-motion framework.

on a one-period model without making any simplifying assumption and thus sacrificing the merits of continuous-time finance. This holds true, in particular, for the classical performance measures mentioned above. The reason why we use log-returns instead of discrete returns is because our model is time-continuous. More precisely, it is build on the standard assumption that the value processes follow a geometric Brownian motion. It is worth noting that return-to-risk measures can change essentially when substituting log-returns with discrete returns or vice versa—even if we work with daily asset (log-)returns.

The classical performance measures are based on the CAPM and so they are vulnerable to a model misspecification. On the contrary, the OP is not based on any capital-market model—it is a probabilistic measure. Thus, it is model-independent although we assume that the value processes of the strategy and of the benchmark follow a 2-dimensional geometric Brownian motion. This assumption is made only for statistical reasons and it has nothing to do with the OP itself (see Definition 1). Hence, it is possible to find nonparametric estimators for the OP, which are not based on any parametric model regarding the value processes of the strategy and of the benchmark. We plan to investigate such kind of estimators in the future.

Nonetheless, in this work we deal with the ML-estimator for Π and this estimator is not unimpeachable:

- It is based on a parametric model, namely the geometric Brownian motion. While this model is standard in financial mathematics, we may doubt that value processes follow a geometric Brownian motion in real life. The assumption that log-returns are independent and identically normally distributed contradicts the stylized facts of empirical finance.
- We assume that the time of liquidation does not depend on whether or not the strategy outperforms the benchmark. This assumption might be violated, for example, if investors suffer from the disposition effect, i.e., if they tend to sell winners too early and ride losers too long (Shefrin and Statman 1985).
- Also the holding-time distribution follows a parametric model, which can either be true or false, too. Indeed, we have to choose some model for T but need not necessarily know its parameters. However, in order to estimate the parameters of F , we would need appropriate data and then the statistical properties of the ML-estimator might change essentially.

The reader might ask whether we want to suggest a better alternative to return-to-risk measures from a pure theoretical perspective or to provide a more illustrative performance measure for the asset-management industry. The answer is that we aim at *both*. However, it is worth emphasizing that we do not try to convince the audience that return-to-risk measures are useless or fallacious *per se*. In fact, the OP itself is based on a return-to-risk measure, namely the ICV, and we will see that this is the most important ingredient of the OP. The question is whether one should use *two* return-to-risk measures or just *one* in order to compare a strategy with some benchmark. We decidedly propagate the latter approach and refer to the arguments raised by Sharpe (1994). Simply put, performance measures are not linear and so the difference between the performance of two investment opportunities is not the same as the performance of the first investment opportunity compared to the second. Our own return-to-risk measure is the ICV and we transform the ICV into a *probability* by treating the time of liquidation stochastic. This is done in order to account for the individual liquidity preference of the investor and, especially, for a better practical understanding.

3. Empirical Investigation

3.1. General Observations

Actively managed funds usually aim at beating some benchmark, typically a stock-market index. Our financial markets and computer technology provide many possibilities in order to pursue that target. For example, we could think about stock picking, market timing, portfolio optimization,

technical analysis, fundamental analysis, time-series analysis, high-frequency trading, algorithmic trading, etc. The list of opportunities is virtually endless, but the problem is that most sophisticated trading strategies produce high transaction costs and other expenses like management fees and agency costs. The principal question is whether or not actively managed funds are able to beat their benchmarks after taking all expenses into account.

Jensen (1968) reports that the 115 (open-end) mutual funds that he considered in his study, which covers the period from 1945 to 1964, “were on average not able to predict security prices well enough to outperform a buy-the-market-and-hold policy.” He states also that “there is very little evidence that any individual fund was able to do significantly better than that which we expected from mere random chance.” By contrast, Ippolito (1989) investigated 143 mutual funds over the period from 1965 to 1984 and finds that “mutual funds, net of all fees and expenses, except load charges, outperformed index funds on a risk-adjusted basis.” This is in contrast to many other studies, but he emphasizes that “the industry alpha, though significantly positive, is not sufficiently large to overcome the load charges that characterize the majority of funds in the sample.” Hence, after taking sales charges or commissions into account, the Alpha of mutual funds disappears. Further, Grinblatt and Titman (1989) considered the sample period from 1974 to 1984. They write that “superior performance may in fact exist, particularly among aggressive-growth and growth funds and those funds with the smallest net asset values.” However, “these funds also have the highest expenses so that their actual returns, net of all expenses, do not exhibit abnormal performance.” Hence, they come to the conclusion that “investors cannot take advantage of the superior abilities of these portfolio managers by purchasing shares in their mutual funds.” Finally, Fama and French (2010) considered the period from 1984 to 2006 and they conclude that “fund investors in aggregate realize net returns that underperform three-factor, and four-factor benchmarks by about the costs in expense ratios,” which means that “if there are fund managers with enough skill to produce benchmark-adjusted expected returns that cover costs, their tracks are hidden in the aggregate results by the performance of managers with insufficient skill.” To sum up, it seems to be better for the market participants to track a stock-market index or to choose the market portfolio.

There exists a large number of passively managed ETFs on the market, which aim at replicating a specific stock-market index like the S&P 500 or the Russell 1000. Elton et al. (1996) vividly explain why it makes sense to compare actively managed funds with index ETFs, and not with stock-market indices, in order to analyze whether or not fund managers are able to beat the market:

“The recent increase in the number and types of index funds that are available to individual investors makes this a matter of practical as well as theoretical significance. Numerous index funds, which track the Standard and Poor’s (S&P) 500 Index or various small-stock, bond, value, growth, or international indexes, are now widely available to individual investors. [...] Given that there are sufficient index funds to span most investors’ risk choices, that the index funds are available at low cost, and that the low cost of index funds means that a combination of index funds is likely to outperform an active fund of similar risk, the question is, why select an actively managed fund?”

Before the first index ETF, i.e., the SPDR S&P 500 ETF (SPY), came up in 1993, authors had to think about how a private investor would replicate a stock index by himself. That is, they had to take all transaction costs, exchange fees, taxes, time and effort, etc., into account. Index ETFs solved that problem all at once. There are no loading charges and, except for minor broker fees and very low transaction costs, private investors do not have to care about anything in order to track a stock index. Many index funds are very liquid and their tracking error seems to be very small. Thus, it is clear why investors usually are not able or willing to replicate a stock-market index by themselves, which is the reason why we treat index ETFs as benchmarks in this work.

Compared to actively managed funds, index ETFs have much lower expense ratios, but their performance might be worse. In order to clarify whether or not the additional costs of an actively managed fund are overcompensated by a better performance, we compare 10 mutual funds with

index ETFs. Most of the funds contain US large-cap stocks and have 1 to 60 billion US dollars of assets under management. In any case, they try to outperform the S&P 500 or the Russell 1000. Hence, our first benchmark is the SPDR S&P 500 ETF. This is almost equivalent to the S&P 500 total return stock index, which serves as a proxy for the overall US stock market. The second benchmark is the iShares Russell 1000 Growth ETF (IWF). That ETF seeks to track the Russell 1000, which is composed of large- and mid-cap US equities exhibiting growth characteristics. In order to reflect the US mid-term treasury-bond market, we chose the iShares 7–10 Year Treasury Bond ETF (IEF). This ETF is composed of US treasury bonds with remaining maturities between 7 and 10 years, including only AAA-rated securities. Our next benchmark is the US money-market account, based on the London Inter-bank Offered Rate (LIBOR) for overnight US-dollar deposits. Our last benchmark is cash, without any interest or risk, i.e., “keeping the money under the mattress.” In this special case, the OP quantifies the probability that the final value of the strategy exceeds its initial value, i.e., $\Pi = P(S_T > 1)$. Hence, to outperform cash just means to generate a positive return, which is the least one should expect when investing money.

Our empirical analysis is based on daily price observations from 2 January 2003, to 31 December 2018. Hence, we take also the financial crisis 2008 into account. The length of that history amounts to $n = 4027$ trading days. All ETFs and mutual funds, together with their symbols, their estimated drift coefficients ($\hat{\mu}$), diffusion coefficients ($\hat{\sigma}$), growth rates ($\hat{\gamma}$), expense ratios (ER), and their assets under management (AUM) are presented in Table 1. It is separated into two parts: The first part shows the benchmarks, whereas the second part contains the mutual funds. We can see that the expense ratios of the mutual funds are up to 90 basis points higher compared to the S&P 500 ETF (SPY). On average, investors pay 67 basis points more for a mutual fund compared to the S&P 500 ETF and 56 basis points more compared to the Russell 1000 ETF (IWF). MSEGX possesses the highest drift coefficient and growth rate,¹⁴ whereas FKGRX has the lowest volatility, i.e., diffusion coefficient, among all mutual funds. PINDX has the lowest drift coefficient and growth rate but the second highest volatility. While the stock-index ETFs are hard to distinguish according to $\hat{\mu}$ and $\hat{\sigma}$, it should be noted that the Russell 1000 ETF has a slightly higher estimated growth rate, $\hat{\gamma}$, compared to the S&P 500 ETF.

Table 1. Basic characteristics of the benchmarks and funds that are taken into consideration, i.e., the estimated drift coefficients ($\hat{\mu}$), diffusion coefficients ($\hat{\sigma}$), growth rates ($\hat{\gamma}$) as well as their expense ratios (ER) and assets under management (AUM).

Benchmark/Fund	Symbol	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	ER (%)	AUM (bn \$)
SPDR S&P 500 ETF	SPY	0.0995	0.1815	0.0830	0.09	264.06
iShares Russell 1000 Growth ETF	IWF	0.1064	0.1774	0.0906	0.20	44.44
iShares 7–10 Year Treasury Bond ETF	IEF	0.0465	0.0656	0.0444	0.15	13.31
Money market	–	0.0100	0.0007	0.0100	–	–
Cash	–	0	0	0	–	–
Fidelity Blue Chip Growth	FBGRX	0.1135	0.1919	0.0951	0.72	26.89
T. Rowe Price Blue Chip Growth	TRBCX	0.1200	0.1974	0.1005	0.70	60.37
Franklin Growth A	FKGRX	0.1064	0.1708	0.0918	0.84	14.96
JPMorgan Large Cap Growth A	OLGAX	0.1105	0.1900	0.0924	0.94	15.15
Vanguard Morgan Growth	VMRGX	0.1094	0.1889	0.0915	0.37	14.79
Morgan Stanley Institutional Growth A	MSEGX	0.1299	0.2164	0.1064	0.89	6.93
BlackRock Capital Appreciation	MAFGX	0.1060	0.1931	0.0873	0.76	3.13
AllianzGI Focused Growth Fund Admin	PGFAX	0.1129	0.1942	0.0940	0.99	1.02
Goldman Sachs Large Cap Growth Insights	GCGIX	0.1015	0.1829	0.0848	0.53	2.16
Pioneer Disciplined Growth A	PINDX	0.0993	0.2077	0.0777	0.87	1.22
∅		0.1110	0.1933	0.0922	0.76	14.66

¹⁴ It holds that $E(S_T) = e^{\mu_S T}$ and so the drift coefficient μ_S can be considered the internal rate of return on the fund.

However, the quantities $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$ are subject to estimation errors and it is well-known that the estimation errors regarding μ and γ are much higher than those regarding σ . Even if we work with daily observations, they cannot be neglected. Hence, Table 1 gives no clear recommendation in favor or against the mutual funds if we compare the quantities of the index ETFs and the mutual funds with each other. On average, the annualized drifts, volatilities, and growth rates of the mutual funds are close to those of the S&P 500 ETF and the Russell 1000 ETF. Does the overall picture change if we evaluate the ordinary Sharpe ratios of the mutual funds?

Table 2 contains the ordinary and the generalized Sharpe ratios of all mutual funds, on an annualized basis, with respect to our two stock-market benchmarks and the bond market. We can see that 9 out of all 10 mutual funds have an ordinary Sharpe ratio that is greater than the (ordinary) Sharpe ratio of the S&P 500 ETF. However, 7 mutual funds possess a Sharpe ratio that is lower than the Sharpe ratio of the Russell 1000 ETF. Hence, we could believe that most fund managers are not able to beat that benchmark. Moreover, except for FKGRX, all funds are inferior to the US treasury-bond ETF. This would suggest that it is better to refrain from the stock market at all and to prefer mid-term US treasury bonds. This conclusion is fallacious.

Table 2. Ordinary vs. generalized Sharpe ratios (annualized).

Symbol	Ordinary	Generalized		
		SPY	IWF	IEF
SPY	0.4933	–	–	–
IWF	0.5433	–	–	–
IEF	0.5579	–	–	–
FBGRX	0.5396	0.2480	0.1608	0.2970
TRBCX	0.5576	0.3570	0.2854	0.3187
FKGRX	0.5649	0.1433	0.0020	0.2914
OLGAX	0.5292	0.1568	0.0788	0.2874
VMRGX	0.5264	0.1954	0.0801	0.2824
MSEGX	0.5544	0.3304	0.2903	0.3366
MAFGX	0.4973	0.1014	−0.0074	0.2628
PGFAX	0.5304	0.2185	0.1338	0.2921
GCGIX	0.5007	0.0436	−0.1372	0.2541
PINDX	0.4302	−0.0028	−0.1008	0.2198
∅	0.5231	0.1792	0.0786	0.2842

Indeed, the overall picture changes completely when we consider the generalized Sharpe ratios. In fact, 9 out of all 10 mutual funds have a positive Sharpe ratio with respect to the S&P 500 ETF, which means that almost all funds were able to beat this benchmark. Moreover, 7 funds have a positive Sharpe ratio with respect to the Russell 1000 ETF and so the majority of the funds have beaten also that benchmark. Finally, since all Sharpe ratios with respect to the US treasury-bond ETF are positive, it is always better to prefer stocks to bonds. Hence, the given results are entirely different and our conclusion turns into the opposite!

The estimated ICV of a fund tells us how the fund manager performed against some benchmark and, according to Equation (1), it essentially determines the (estimated) OP. Table 3 contains the ICVs of the 10 mutual funds with respect to our 5 benchmarks. This table confirms our latter results. This means that 9 out of all 10 mutual funds, in fact, were able to outperform the S&P 500 ETF. Further, 7 funds outperformed the Russell 1000 ETF. The (uniformly) worst performer is PINDX, which is never preferable compared to the index ETFs. The best performer with respect to the index ETFs is TRBCX, whose track record is extraordinarily good. We conclude that most funds are preferable compared to the index ETFs, which was not clear to us after inspecting the ordinary Sharpe ratios in Table 2. It is worth noting that all mutual funds perform much better if we compare them with the S&P 500 ETF

rather than the Russell 1000 ETF. Table 3 reveals also that mutual funds are clearly superior to bonds, the money market, and cash.

Table 3. ICVs of the mutual funds with respect to the benchmarks. The standard errors can be approximated by $1/\sqrt{m} = 0.2502$ (see Equation (2)).

Symbol	Benchmark				
	SPY	IWF	IEF	LIBOR	Cash
FBGRX	0.2133	0.1002	0.2245	0.4429	0.4947
TRBCX	0.3038	0.2066	0.2432	0.4584	0.5088
FKGRX	0.1825	0.0279	0.2307	0.4790	0.5373
OLGAX	0.1339	0.0346	0.2156	0.4335	0.4858
VMRGX	0.1677	0.0239	0.2115	0.4312	0.4839
MSEGX	0.2539	0.1950	0.2503	0.4454	0.4913
MAFGX	0.0669	−0.0638	0.1896	0.4001	0.4516
PGFAX	0.1790	0.0699	0.2182	0.4326	0.4838
GCGIX	0.0378	−0.1656	0.1865	0.4088	0.4632
PINDX	−0.0747	−0.1845	0.1387	0.3259	0.3738
∅	0.1464	0.0244	0.2109	0.4258	0.4774

Clearly, TRBCX is the best performer among all mutual funds and so it is evident why it has, by far, the largest amount of assets under management (over 60 billion US dollars). Figure 5 plots the ICV of each mutual fund, both with respect to the S&P 500 and with respect to the Russell 1000, against the natural logarithm of its assets under management. It turns out that the ICV of a fund is highly correlated with the log-amount of assets under management. To be more precise, the correlation between ICV and $\log(\text{AUM})$ amounts to 71.22% if we refer to the S&P 500 and to 65.67% if we refer to the Russell 1000. However, we can observe two outliers, namely MSEGX and PGFAX, which performed well between 2003 and 2018 compared to their assets under management.

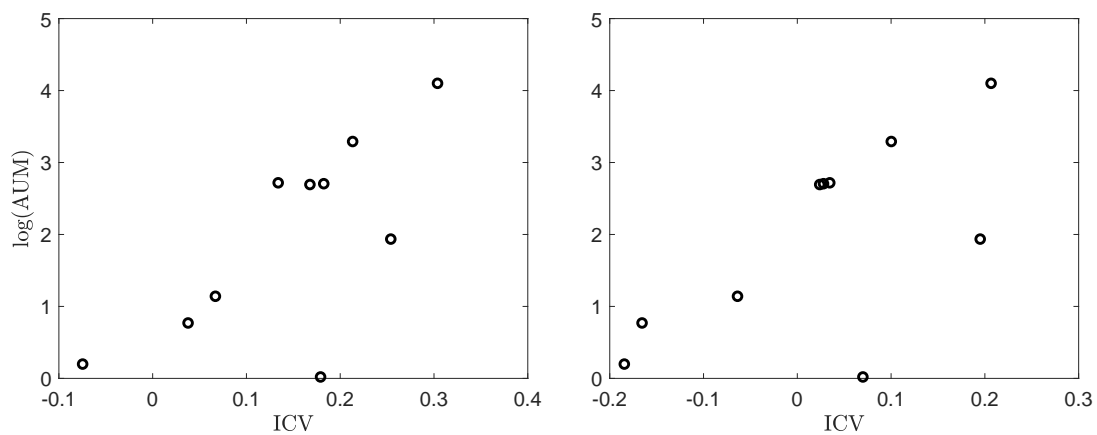


Figure 5. ICV against the natural logarithm of the assets under management of the mutual funds with respect to the S&P 500 (left) and the Russell 1000 (right).

Our results demonstrate that it is highly important to consider *differential* returns and not excess returns if we want to compare some strategy with a specified benchmark. By comparing the ordinary Sharpe ratio of some strategy with the ordinary Sharpe ratio of a given benchmark (see Table 1) one checks which of the two is better able to outperform the *money-market account*. However, this is not what we want to know when we try to understand whether or not a strategy is able to outperform its *benchmark*. For this purpose we have to calculate the generalized Sharpe ratios in Table 2 (Sharpe 1994) or the ICVs in Table 3.

Our arguments shall be clarified also from a statistical point of view: Let \widehat{Sh}_{S_n} be the empirical estimator for the annualized Sharpe ratio of the strategy and \widehat{Sh}_{B_n} be the empirical estimator for the annualized Sharpe ratio of the benchmark. Further, let $\Delta Sh := Sh_S - Sh_B$ be the difference between the annualized Sharpe ratio of the strategy, Sh_S , and the annualized Sharpe ratio of the benchmark, i.e., Sh_B . Then one uses the test statistic $\Delta \widehat{Sh}_n := \widehat{Sh}_{S_n} - \widehat{Sh}_{B_n}$ in order to test the null hypothesis $H_0: \Delta Sh \leq 0$. Our results are presented on the right-hand side of Table 4, which contains also the corresponding p -values for H_0 in parentheses. By contrast, the left-hand side of Table 4 contains the estimated ICVs and their corresponding p -values for the null hypothesis $H_0: ICV \leq 0$.¹⁵ We can see that most p -values on the left-hand side are smaller than those on the right-hand side. This holds true in particular when we compare the funds with the Russell 1000 ETF and the US treasury-bond ETF. Hence, using the ICV instead of comparing two Sharpe ratios with one another in order to test whether a strategy outperforms its benchmark leads to results that are more significant in an *economic* sense. To be more precise, in the majority of cases, the estimated ICVs are positive and their p -values are low enough in order to conclude that most fund managers are able to beat their benchmarks. However, the given results are still insignificant in a *statistical* sense.

Table 4. Test statistics \widehat{ICV}_n and $\Delta \widehat{Sh}_n$ together with their corresponding p -values (in parentheses) for the null hypotheses $H_0: ICV \leq 0$ and $H_0: \Delta Sh \leq 0$, respectively.

Symbol	\widehat{ICV}_n			$\Delta \widehat{Sh}_n$		
	SPY	IWF	IEF	SPY	IWF	IEF
FBGRX	0.2133 (0.1970)	0.1002 (0.3444)	0.2245 (0.1848)	0.0463 (0.2908)	−0.0183 (0.5168)	−0.0038 (0.5233)
TRBCX	0.3038 (0.1123)	0.2066 (0.2044)	0.2432 (0.1654)	0.0643 (0.2193)	−0.0003 (0.5003)	0.0142 (0.4149)
FKGRX	0.1825 (0.2329)	0.0279 (0.4556)	0.2307 (0.1782)	0.0717 (0.1756)	0.0071 (0.4935)	0.0216 (0.3822)
OLGAX	0.1339 (0.2963)	0.0346 (0.4450)	0.2156 (0.1944)	0.0360 (0.3664)	−0.0286 (0.5264)	−0.0141 (0.5706)
VMRGX	0.1677 (0.2513)	0.0239 (0.4619)	0.2115 (0.1989)	0.0331 (0.3315)	−0.0315 (0.5289)	−0.0170 (0.6200)
MSEGX	0.2539 (0.1550)	0.1950 (0.2178)	0.2503 (0.1585)	0.0611 (0.3071)	−0.0035 (0.5032)	0.0110 (0.4575)
MAFGX	0.0669 (0.3946)	−0.0638 (0.6006)	0.1896 (0.2243)	0.0040 (0.4829)	−0.0606 (0.5556)	−0.0460 (0.7294)
PGFAX	0.1790 (0.2371)	0.0699 (0.3899)	0.2182 (0.1915)	0.0371 (0.3403)	−0.0275 (0.5252)	−0.0130 (0.5728)
GCGIX	0.0378 (0.4400)	−0.1656 (0.7460)	0.1865 (0.2280)	0.0074 (0.4587)	−0.0571 (0.5524)	−0.0426 (0.7818)
PINDX	−0.0747 (0.6173)	−0.1845 (0.7696)	0.1387 (0.2897)	−0.0631 (0.7481)	−0.1277 (0.6165)	−0.1131 (0.8875)

3.2. Empirical Results

In this section, we present our estimates of the OPs for different holding-time distributions. More precisely, we consider a fixed, a uniformly distributed, an exponentially distributed, and a Weibull distributed holding period. In order to make the results comparable, we chose the parameters such that the mean holding period equals 5 years for each holding-time distribution.

Our principal question is whether or not actively managed funds are able to beat their benchmarks, which can be answered in a very simple way by estimating their OPs. In the case that $\Pi \leq 0.5$, we can

¹⁵ All p -values have been computed by using the asymptotic results reported by [Frahm \(2018\)](#).

say that the fund manager is not able to beat the specified benchmark.¹⁶ Then it makes not much sense for an investor to buy the fund share because, with probability greater or equal to 50%, the benchmark performs at least as good as the mutual fund. Hence, our null hypothesis is $H_0: \Pi \leq 0.5$. Of course, we have to take estimation errors into account and so we report the estimated OPs along with their standard errors and p -values.

3.2.1. Fixed Holding Period

Let us begin by considering the holding period $T > 0$ fixed. In this case, the ML-estimator for the OP is simply $\widehat{\Pi}_n = \Phi(\sqrt{T} \widehat{ICV}_n)$. Further, its standard error is, approximately,

$$\text{Std}(\widehat{\Pi}_n) \approx \sqrt{\frac{T}{m}} \phi(\sqrt{T} \widehat{ICV}_n).$$

Table 5 contains the results if we assume that the holding period equals 5 years. We can see that 9 out of the 10 mutual funds that are taken into consideration are able to outperform the S&P 500 and 7 funds can outperform the Russell 1000. In fact, this follows also from Table 3 because the OP is greater than 0.5 if and only if $ICV > 0$. Nonetheless, Table 5 provides some additional information. It reveals that the OPs typically range between 0.5 and 0.7 with regard to the index ETFs. An apparent exception is the T. Rowe Price Blue Chip Growth Fund (TRBCX), which has a tremendous OP of 0.7515 with respect to the S&P 500. The worst fund is the Pioneer Disciplined Growth Fund (PINDX), which underperforms both the S&P 500 and the Russell 1000 with a quite small OP of 0.4337 and 0.3400, respectively.

Despite the fact that most funds outperform their stock-market benchmarks, the results are not statistically significant. All p -values that are related to the index ETFs exceed $\alpha = 0.05$. Only TRBCX comes very close to α with a p -value of 0.0779, given that we use the S&P 500 ETF as a benchmark. Of course, this does not mean that we have that $\Pi \leq 0.5$ for any fund. Performance measurement is very susceptible to estimation risk (Frahm 2018) and thus, in that context, we cannot expect to obtain statistically significant results. Nevertheless, we may say that the results are, at least, economically significant. This was surprising to us, since in the finance community it is usually asserted that fund managers, in general, do not better than their benchmarks. This was already discussed in Section 3.1.

All mutual funds outperform the bond market, but these results are not significant either. However, most funds at least outperform the money-market account and cash on a significance level of $\alpha = 0.01$, which can be seen by the p -values at the end of Table 5. The corresponding OPs usually exceed 0.8, which means that the probability of generating some excess return or of making *any* profit at the time of liquidation is greater than 80%. Once again, an exception is PINDX, whose OPs with respect to the money-market account and cash are lower than 80%.

¹⁶ Note that this is precisely the case in which we have that $ICV \leq 0$.

Table 5. Results in the case that *T* equals 5 years.

Fund	SPY			IWF			IEF			LIBOR			Cash		
	OP	Std	<i>p</i>	OP	Std	<i>p</i>	OP	Std	<i>p</i>	OP	Std	<i>p</i>	OP	Std	<i>p</i>
FBGRX	0.6833	0.1992	0.1788	0.5886	0.2176	0.3419	0.6921	0.1967	0.1644	0.8390	0.1367	0.0066	0.8657	0.1210	0.0013
TRBCX	0.7515	0.1772	0.0779	0.6780	0.2006	0.1874	0.7068	0.1925	0.1414	0.8473	0.1320	0.0042	0.8724	0.1168	0.0007
FKGRX	0.6584	0.2053	0.2203	0.5249	0.2227	0.4555	0.6970	0.1954	0.1566	0.8580	0.1257	0.0022	0.8852	0.1084	0.0002
OLGAX	0.6176	0.2134	0.2907	0.5308	0.2225	0.4449	0.6851	0.1987	0.1758	0.8338	0.1395	0.0084	0.8613	0.1237	0.0017
VMRGX	0.6462	0.2080	0.2411	0.5213	0.2228	0.4619	0.6819	0.1995	0.1810	0.8325	0.1402	0.0088	0.8604	0.1243	0.0019
MSEGX	0.7149	0.1899	0.1289	0.6686	0.2029	0.2030	0.7122	0.1908	0.1331	0.8403	0.1359	0.0061	0.8640	0.1220	0.0014
MAFGX	0.5595	0.2207	0.3938	0.4433	0.2209	0.6013	0.6642	0.2040	0.2105	0.8145	0.1496	0.0177	0.8437	0.1340	0.0052
PGFAX	0.6556	0.2060	0.2251	0.5621	0.2204	0.3890	0.6872	0.1981	0.1723	0.8333	0.1398	0.0085	0.8603	0.1243	0.0019
GCGIX	0.5336	0.2224	0.4399	0.3556	0.2084	0.7558	0.6617	0.2046	0.2147	0.8196	0.1470	0.0148	0.8498	0.1305	0.0037
PINDX	0.4337	0.2201	0.6184	0.3400	0.2050	0.7825	0.6217	0.2127	0.2835	0.7669	0.1711	0.0594	0.7984	0.1574	0.0290
∅	0.6254	0.2062	0.2815	0.5213	0.2144	0.4623	0.6810	0.1993	0.1833	0.8285	0.1417	0.0137	0.8561	0.1263	0.0047

3.2.2. Uniformly Distributed Holding Period

Now, we assume that the investor has a holding period that is uniformly distributed between 0 and M years. The ML-estimator for Π is

$$\hat{\Pi}_n = \frac{1}{M} \int_0^M \Phi(\sqrt{t} \widehat{ICV}_n) dt$$

and its standard error can be approximated by

$$\text{Std}(\hat{\Pi}_n) \approx \frac{1}{\sqrt{m}} \frac{1}{M} \int_0^M \sqrt{t} \phi(\sqrt{t} \widehat{ICV}_n) dt.$$

Table 6 contains the results if we assume that the holding period is uniformly distributed between 0 and 10 years. As we can see, the given results are not essentially different compared to those in Table 5. However, it is conspicuous that the OPs are slightly lower. The same effect can be observed also in Figure 3 if we compare the OPs on the left-hand side for $T = 5$ with the OPs on the right-hand side for $M = 10$. The reason is that liquidating a strategy with a positive ICV prematurely lowers the OP. The problem is that later liquidations cannot compensate for earlier ones if the strategy is profitable compared to its benchmark, i.e., $ICV > 0$.

While the OPs in Table 6 are slightly lower compared to those in Table 5, we can verify that all funds are still better than the bond market. The majority of funds are even significantly better than the money-market account and cash on a level of 1%. To sum up, *in general*, it is preferable to invest in mutual funds also if the holding period of the investor is uniformly distributed. However, there are noticeable differences regarding the performance of the mutual funds, which means that investors should be aware that some funds perform much better than others. The question of whether or not the estimated OP of a fund is a *persistent* performance measure is not answered in this study. If it is persistent, i.e., if it can be used as an ex-ante and not only as an ex-post performance measure, investors should clearly prefer funds with a high estimated OP.

Table 6. Results in the case that T is uniformly distributed between 0 and 10 years.

Fund	SPY			IWF			IEF			LIBOR			Cash		
	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p
FBGRX	0.6716	0.1839	0.1754	0.5834	0.2042	0.3414	0.6798	0.1813	0.1607	0.8124	0.1208	0.0049	0.8359	0.1064	0.0008
TRBCX	0.7341	0.1607	0.0726	0.6667	0.1854	0.1843	0.6932	0.1767	0.1371	0.8197	0.1164	0.0030	0.8418	0.1027	0.0004
FKGRX	0.6485	0.1906	0.2179	0.5235	0.2099	0.4555	0.6842	0.1798	0.1528	0.8291	0.1107	0.0015	0.8531	0.0952	0.0001
OLGAX	0.6106	0.1994	0.2896	0.5291	0.2096	0.4449	0.6733	0.1834	0.1723	0.8078	0.1235	0.0063	0.8321	0.1088	0.0011
VMRGX	0.6372	0.1935	0.2392	0.5201	0.2100	0.4619	0.6703	0.1843	0.1777	0.8067	0.1241	0.0067	0.8312	0.1094	0.0012
MSEGX	0.7007	0.1740	0.1244	0.6580	0.1879	0.2002	0.6982	0.1749	0.1286	0.8136	0.1201	0.0045	0.8345	0.1073	0.0009
MAFGX	0.5560	0.2076	0.3937	0.4466	0.2078	0.6014	0.6539	0.1891	0.2078	0.7906	0.1331	0.0145	0.8165	0.1184	0.0037
PGFAX	0.6459	0.1913	0.2228	0.5585	0.2073	0.3888	0.6752	0.1827	0.1688	0.8073	0.1237	0.0065	0.8312	0.1094	0.0012
GCGIX	0.5317	0.2095	0.4399	0.3645	0.1939	0.7577	0.6516	0.1897	0.2122	0.7952	0.1306	0.0119	0.8219	0.1151	0.0026
PINDX	0.4376	0.2069	0.6186	0.3499	0.1902	0.7850	0.6144	0.1987	0.2823	0.7480	0.1545	0.0542	0.7763	0.1407	0.0248
∅	0.6174	0.1917	0.2794	0.5200	0.2006	0.4621	0.6694	0.1841	0.1800	0.8030	0.1258	0.0114	0.8274	0.1113	0.0037

3.2.3. Exponentially Distributed Holding Period

In this section, we discuss the results that are obtained by assuming that the holding period is exponentially distributed. This distributional assumption might be considered more realistic than a fixed or a uniformly distributed holding period. It assigns a high probability to early liquidations and a low probability to late ones. The exponential distribution has no finite right endpoint, which means that T is unbounded above. Hence, this distribution allows for large holding periods. However, the survival probability $P(T > t)$ decreases exponentially with t .

The ML-estimator for the OP is

$$\hat{\Pi}_n = \lambda \int_0^\infty \Phi\left(\sqrt{t} \widehat{\text{ICV}}_n\right) e^{-\lambda t} dt,$$

where $\lambda > 0$ is the parameter of the exponential distribution. The standard error of the ML-estimator can be approximated by

$$\text{Std}(\hat{\Pi}_n) \approx \frac{\lambda}{\sqrt{m}} \int_0^\infty \sqrt{t} e^{-\lambda t} \phi\left(\sqrt{t} \widehat{\text{ICV}}_n\right) dt.$$

We set the parameter λ to 0.2, which means that the mean holding period equals 5 years. The results are given by Table 7. As we can see, the OPs are, once again, slightly lower than those in Table 6 and thus also lower than those in Table 5. The reason is that the exponential distribution assigns earlier liquidations a substantially higher probability than later ones (see Figure 2). The overall conclusion does not differ compared to our previous findings, but it is worth noting that the standard errors and thus also the p -values are somewhat lower than in Tables 5 and 6. Now, we have a p -value of 0.0675 for TRBCX against the S&P 500 ETF, which comes very close to the significance level of 5%. In any case, the mutual funds still outperform the bond market and the riskless investments, i.e., the money market and cash.

Table 7. Results in the case that T is exponentially distributed with $\lambda = 0.2$.

Fund	SPY			IWF			IEF			LIBOR			Cash		
	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p
FBGRX	0.6598	0.1683	0.1712	0.5782	0.1905	0.3407	0.6672	0.1655	0.1562	0.7868	0.1087	0.0042	0.8081	0.0966	0.0007
TRBCX	0.7165	0.1449	0.0675	0.6553	0.1699	0.1803	0.6795	0.1608	0.1322	0.7934	0.1050	0.0026	0.8134	0.0936	0.0004
FKGRX	0.6386	0.1754	0.2147	0.5221	0.1972	0.4555	0.6713	0.1640	0.1481	0.8019	0.1002	0.0013	0.8237	0.0875	0.0001
OLGAX	0.6035	0.1852	0.2881	0.5273	0.1969	0.4448	0.6613	0.1677	0.1681	0.7827	0.1110	0.0054	0.8046	0.0986	0.0010
VMRGX	0.6281	0.1786	0.2365	0.5189	0.1973	0.4619	0.6586	0.1687	0.1735	0.7817	0.1115	0.0058	0.8038	0.0991	0.0011
MSEGX	0.6863	0.1580	0.1193	0.6473	0.1726	0.1966	0.6840	0.1590	0.1236	0.7879	0.1081	0.0039	0.8067	0.0974	0.0008
MAFGX	0.5526	0.1945	0.3934	0.4498	0.1948	0.6016	0.6436	0.1738	0.2044	0.7673	0.1194	0.0126	0.7905	0.1066	0.0032
PGFAX	0.6362	0.1762	0.2197	0.5550	0.1942	0.3886	0.6631	0.1671	0.1645	0.7823	0.1112	0.0056	0.8038	0.0991	0.0011
GCGIX	0.5298	0.1967	0.4398	0.3734	0.1790	0.7603	0.6414	0.1745	0.2089	0.7714	0.1172	0.0103	0.7954	0.1039	0.0022
PINDX	0.4414	0.1937	0.6189	0.3600	0.1750	0.7882	0.6071	0.1843	0.2806	0.7290	0.1389	0.0496	0.7544	0.1262	0.0219
∅	0.6093	0.1771	0.2769	0.5187	0.1867	0.4618	0.6577	0.1685	0.1760	0.7784	0.1131	0.0101	0.8004	0.1009	0.0033

3.2.4. Weibull Distributed Holding Period

Finally, we report the results that are obtained by assuming that the holding period is Weibull distributed. The Weibull distribution generalizes the exponential distribution and allows us to consider unimodal holding-time distributions with mode greater than 0 (see Figure 2). The corresponding ML-estimator for the OP is

$$\hat{\Pi}_n = \frac{\kappa}{\gamma} \int_0^\infty \Phi\left(\sqrt{t} \widehat{\text{ICV}}_n\right) \left(\frac{t}{\gamma}\right)^{\kappa-1} e^{-(t/\gamma)^\kappa} dt$$

and its standard error can be approximated by

$$\text{Std}(\hat{\Pi}_n) \approx \frac{\kappa \gamma^{-\kappa}}{\sqrt{m}} \int_0^\infty t^{\kappa-\frac{1}{2}} e^{-(t/\gamma)^\kappa} \phi\left(\sqrt{t} \widehat{\text{ICV}}_n\right) dt.$$

In our empirical study, the parameters of the Weibull distribution are $\kappa = 2$ and $\gamma = 5.6419$. Once again, this leads us to a mean holding period of 5 years. The results are contained in Table 8. Now, we can observe that the OPs are close to those that are obtained with respect to the index ETFs by assuming that $T = 5$ is fixed (see Table 5). However, the OPs with respect to the US treasury-bond ETF, the money-market account, and to cash are slightly lower than in Table 5 but still higher than in Tables 6 and 7. Nonetheless, we can see that the given results are not very much driven by our specific assumption about the distribution of the holding period. It turns out that the OP is essentially determined by the ICV, not by the holding-time distribution, provided that the considered holding-time distributions have the same mean.

Table 8. Results in the case that T is Weibull distributed with $\kappa = 2$ and $\gamma = 5.6419$.

Fund	SPY			IWF			IEF			LIBOR			Cash		
	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p	OP	Std	p
FBGRX	0.6756	0.1890	0.1764	0.5852	0.2088	0.3415	0.6840	0.1864	0.1619	0.8215	0.1266	0.0056	0.8462	0.1121	0.0010
TRBCX	0.7400	0.1662	0.0744	0.6706	0.1905	0.1853	0.6978	0.1820	0.1385	0.8292	0.1222	0.0035	0.8524	0.1083	0.0006
FKGRX	0.6519	0.1955	0.2186	0.5240	0.2144	0.4555	0.6886	0.1850	0.1540	0.8391	0.1165	0.0018	0.8643	0.1008	0.0001
OLGAX	0.6130	0.2042	0.2899	0.5297	0.2141	0.4449	0.6773	0.1885	0.1734	0.8167	0.1293	0.0072	0.8422	0.1146	0.0014
VMRGX	0.6403	0.1984	0.2398	0.5205	0.2145	0.4619	0.6743	0.1894	0.1787	0.8155	0.1300	0.0076	0.8413	0.1151	0.0015
MSEGX	0.7055	0.1793	0.1259	0.6617	0.1930	0.2011	0.7029	0.1802	0.1301	0.8227	0.1259	0.0052	0.8447	0.1131	0.0012
MAFGX	0.5572	0.2121	0.3937	0.4454	0.2124	0.6014	0.6574	0.1941	0.2086	0.7988	0.1389	0.0158	0.8258	0.1242	0.0043
PGFAX	0.6492	0.1962	0.2235	0.5598	0.2119	0.3889	0.6793	0.1879	0.1699	0.8162	0.1296	0.0073	0.8413	0.1151	0.0015
GCGIX	0.5324	0.2140	0.4399	0.3614	0.1988	0.7571	0.6551	0.1947	0.2129	0.8035	0.1364	0.0130	0.8315	0.1209	0.0031
PINDX	0.4362	0.2115	0.6185	0.3465	0.1951	0.7842	0.6170	0.2034	0.2827	0.7544	0.1601	0.0560	0.7838	0.1465	0.0264
∅	0.6201	0.1966	0.2801	0.5205	0.2053	0.4622	0.6734	0.1892	0.1811	0.8118	0.1316	0.0123	0.8374	0.1171	0.0041

4. Conclusions

We propose the OP as a new performance measure, especially for a nonacademic audience. It differs in many aspects from return-to-risk measures: (i) The OP compares some strategy with a specified benchmark, not necessarily with the money-market account. (ii) It is easy to understand for people who are not educated in statistics or probability theory. (iii) The holding period of the investor is considered random. In our Brownian-motion framework, the OP is essentially determined by the ICV, i.e., the expected annual relative log-return on the strategy with respect to its benchmark divided by the standard deviation of that log-return. By contrast, the choice of the holding-time distribution seems to be secondary, provided that the mean holding period is the same for all distributions.

The ML-estimator for the OP is strongly consistent and, if the square root of the holding period has a finite expectation, it is also asymptotically normally distributed under the usual standardization. The moment condition is very mild and allows us to take almost any holding-time distribution into consideration. Our asymptotic results enable us to calculate confidence intervals or to conduct hypotheses tests. The basic null hypothesis is that the OP does not exceed 50%, which means that the strategy does not outperform its benchmark. Our empirical results reveal that most mutual funds, in fact, are able to beat their benchmarks, although the results are not statistically significant. Nonetheless, they are economically significant. This might be surprising because it is a commonplace in the finance literature that most fund managers cannot beat the market. However, the crucial point is that one should refer to differential returns when comparing some strategy with a given benchmark, not to excess returns.

The best performer is the T. Rowe Price Blue Chip Growth Fund, whereas the worst one is the Pioneer Disciplined Growth Fund. It is evident that market participants prefer to invest their money in those funds that have a high ICV with respect to the S&P 500 or the Russell 1000. More precisely, the ICV with respect to any index ETF is highly correlated with the log-amount of assets under management of the mutual fund. This means that relative performance and capital allocation are strongly connected to one another, which suggests that market participants take differential returns implicitly into account when making their investment decisions.

Author Contributions: G.F. developed the theoretical part, whereas both authors are responsible for the empirical part of this work.

Funding: This research received no external funding.

Acknowledgments: We would like to thank Philipp Adämmer and Alexander Jonen for our helpful discussions on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Alexander, Gordon J., and Alexandre M. Baptista. 2003. Portfolio performance evaluation using value at risk. *Journal of Portfolio Management* 29: 93–102. [\[CrossRef\]](#)
- Baweja, Meena, Ratnesh R. Saxena, and Deepak Sehgal. 2015. Portfolio optimization using conditional Sharpe ratio. *International Letters of Chemistry, Physics and Astronomy* 53: 130–36. [\[CrossRef\]](#)
- Burke, Gibbons. 1994. A sharper Sharpe ratio. *Futures* 23: 56.
- Chow, Victor, and Christine W. Lai. 2015. Conditional Sharpe ratios. *Finance Research Letters* 12: 117–33. [\[CrossRef\]](#)
- Dowd, Kevin. Adjusting for risk: An improved Sharpe ratio. *International Review of Economics and Finance* 9: 209–22. [\[CrossRef\]](#)
- Eling, Martin, and Frank Schuhmacher. 2006. Hat die Wahl des Performancemaßes einen Einfluss auf die Beurteilung von Hedgefonds-Indizes? *Kredit und Kapital* 39: 419–54.
- Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 1996. The persistence of risk-adjusted mutual fund performance. *Journal of Business* 69: 133–57.
- Fama, Eugene F., and Kenneth R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65: 1915–47.

- Frahm, Gabriel. 2016. Pricing and valuation under the real-world measure. *International Journal of Theoretical and Applied Finance* 19. [CrossRef]
- Frahm, Gabriel. 2018. An intersection-union test for the Sharpe ratio. *Risks* 6. [CrossRef]
- Grinblatt, Mark, and Sheridan Titman. 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business* 62: 393–416. [CrossRef]
- ICI. 2019. Worldwide Regulated Open-End Fund Assets and Flows, Fourth Quarter 2018. Available online: https://www.iciglobal.org/iciglobal/research/stats/ww/ci.ww_q4_18.global (accessed on 6 April 2019). [CrossRef]
- Ippolito, Richard A. 1989. Efficiency with costly information: A study of mutual fund performance, 1965–1984. *Quarterly Journal of Economics* 104: 1–23. [CrossRef]
- Jensen, Michael C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23: 389–416. [CrossRef]
- Keating, Con, and William F. Shadwick. 2002. A universal performance measure. *Journal of Performance Measurement* 6: 59–84.
- Kestner, Lars N. 1996. Getting a handle on true performance. *Futures* 25: 44–46.
- Lin, Mei-Chen, and Pin-Huang Chou. 2003. The pitfall of using Sharpe ratio. *Finance Letters* 1: 84–89. [CrossRef]
- Lintner, John. 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.
- Lo, Andrew W. 2002. The statistics of Sharpe ratios. *Financial Analysts Journal* 58: 36–52. [CrossRef]
- Lusardi, Annamaria, and Olivia S. Mitchell. 2013. *The Economic Importance of Financial Literacy: Theory and Evidence*. Technical Report. Cambridge: National Bureau of Economic Research.
- Merton, Robert C. 1973. Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4: 141–83.
- Mossin, Jan. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–83. [CrossRef]
- OECD. 2016. *OECD/INFE International Survey of Adult Financial Literacy Competencies*. Technical Report. Paris: OECD. [CrossRef]
- Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–42. [CrossRef]
- Sharpe, William F. 1966. Mutual fund performance. *Journal of Business* 39: 119–38.
- Sharpe, William F. 1994. The Sharpe ratio. *Journal of Portfolio Management* 21: 49–58. [CrossRef]
- Shefrin, Hersch, and Meir Statman. 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* 40: 777–90. [CrossRef]
- Sortino, Frank, Robert Van Der Meer, and Auke Plantinga. 1999. The Dutch triangle. *Journal of Portfolio Management* 26: 50–58. [CrossRef]
- Treynor, Jack L. 1961. Market value, time, and risk. Available online: <https://dx.doi.org/10.2139/ssrn.2600356> (accessed on 25 June 2019).
- Young, Terry W. 1991. Calmar ratio: A smoother tool. *Futures* 20: 1. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).