JOURNAL OF INVESTMENT MANAGEMENT, Vol. 15, No. 3, (2017), pp. 6–22

www.joim.com

MEASURING PORTFOLIO PERFORMANCE: SHARPE, ALPHA, OR THE GEOMETRIC MEAN?

Moshe Levy[a]

The most popular portfolio performance measures are the Sharpe ratio and alpha. While the Sharpe ratio is optimal under the capital asset pricing model (CAPM) assumptions of normal return distributions and unlimited borrowing at the risk-free rate, we find that it is not well aligned with investors' preferences in more realistic settings. Alpha is a poor measure under both the theoretical and the realistic settings. For investors with typical borrowing constraints, the geometric mean provides an alternative measure that is much better than both the Sharpe ratio and alpha. It may very well be the most important single number to consider in portfolio selection.

1 Introduction

Given a set of mutual funds to choose from, the optimal choice depends in general on the investor's preference. Sharpe's (1966, 1994) seminal work shows that if returns are normally distributed and unlimited borrowing at the risk-free rate is possible, there is a unique performance measure, the Sharpe ratio, that is perfectly aligned with the expected utility for all risk-averse investors. Namely, if Fund A has a higher Sharpe ratio than Fund B, then all risk-averse investors, regardless of their exact preferences, will achieve higher expected utility by choosing Fund A, and potentially borrowing or lending at the risk-free rate.[1] This result is of central importance, as it provides a unique optimal ranking of funds, a simple measure of portfolio performance, and sets a clear objective function for the fund manager. Indeed, the Sharpe ratio is probably the most widely used measure of fund performance among academics and practitioners alike.[2]

While it is clear that the assumptions underlying the optimality of the Sharpe ratio are unrealistic, the Sharpe ratio is nevertheless widely perceived as the best available performance measure. The goal of this paper is to examine this perception: How well does the Sharpe ratio work with realistic borrowing constraints and return distributions? Is there a better alternative?

[a]Jerusalem School of Business, The Hebrew University, Jerusalem 91905, Israel. Tel: 972 25883219. Fax 972 25883101. E-mail: mslm@huji.ac.il.

Many studies have shown that the empirical return distributions deviate from the normal distribution in a statistically significant manner. However, it has been argued that in a portfolio selection context the economic loss incurred by assuming normality is economically very small (Levy and Markowitz, 1979; Kroll *et al.*, 1984; Simaan, 1993). In line with these studies, we also find that the Sharpe ratio is almost perfectly aligned with investors' expected utilities even when the empirical (non-normal) return distributions are employed, as long as unlimited borrowing at the risk-free lending rate is possible.

In contrast, replacing the assumption of unlimited borrowing at the risk-free lending rate with a more realistic setting has a dramatic effect on the usefulness of the Sharpe ratio. Federal Reserve Regulation T restricts borrowing to the level of the investor's own initial capital, i.e. an investor with \$1,000 of his own capital can borrow only up to \$1,000. In practice, investors typically restrict themselves to much lower borrowing levels. For example, Fortune (2000) reports that the aggregate debt is below 10% of the aggregate customer assets, for all brokers examined. For major brokers such as Merrill Lynch, Paine Webber, and Charles Schwab, this figure is below 2.5%. Furthermore, the relatively few investors who do borrow restrict themselves to levels of debt that are typically only 50% of the maximum allowed by Regulation T, i.e. at 50% of their initial capital. The likely reason for this is that, in contrast to the assumptions behind the Sharpe ratio (and the CAPM), the interest rate paid on loans is typically much higher than the risk-free lending rate, and increases with the size of the loan (see, for example, Saunders and Schumacher, 2000).

When borrowing is limited, and the borrowing rate is higher than the lending rate, the Sharpe ratio may lead to sub-optimal choices. Figure 1 illustrates such a case: Fund A has a higher Sharpe ratio than Fund B; thus, if borrowing is unlimited and the borrowing rate is equal to the lending rate, then Fund A yields a higher expected utility. However, when borrowing is limited and the borrowing rate is higher than the lending rate, Fund B may certainly yield higher expected utility than Fund A, as illustrated by the figure. This is a point

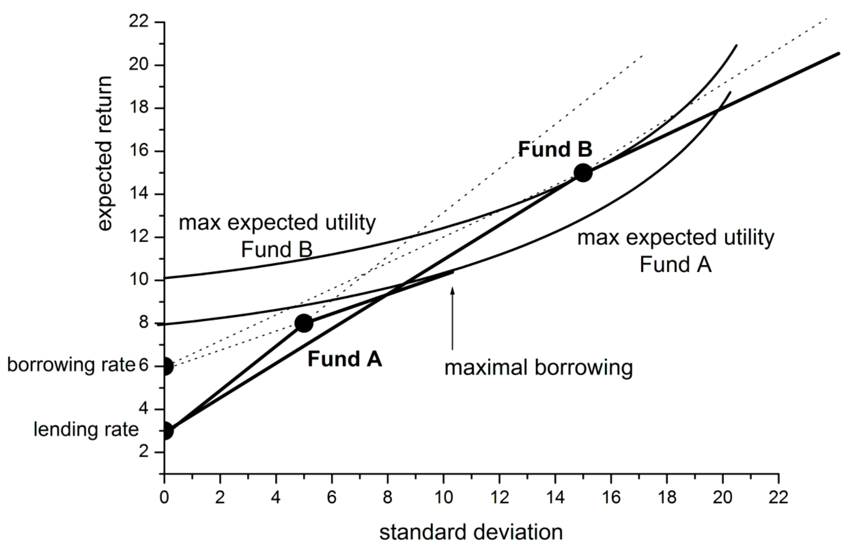


Figure 1 When borrowing is limited, and the borrowing rate is higher than the lending rate, the Sharpe ratio may lead to wrong choices: Fund A has a higher Sharpe ratio, but Fund B yields a higher expected utility. The curved lines are equal-expected-utility lines for a hypothetical investor.

powerfully made by Markowitz (2005). In this realistic setting there is no longer a single performance measure that is perfect for all investors: it is certainly possible that Fund A is better for one investor, while Fund B is better for another. Still, it is extremely important to have a good portfolio performance measure, albeit an imperfect one, for at least three reasons:

1. Fund selection: investors typically do not know their utility functions, and may be unable to intuitively optimize when faced with a large number of potential funds to choose from. A good performance measure allowing for a simple ranking of funds is thus of great importance.
2. Fund manager evaluation: according to what criteria should fund managers be compensated? And closely related to this issue:
3. Setting the target objective function for the manager—setting a target that is closely aligned with the investors' preferences motivates the manager to take actions that maximize his investors' welfare.

How well does the Sharpe ratio perform when borrowing is realistically constrained and is at a higher rate than the lending rate? We empirically find that the Sharpe ratio provides a ranking of funds that is typically very different than the ranking by the investor's expected utility, and that the economic loss implied by choosing a fund based on the Sharpe ratio is very substantial.

Another central problem of the Sharpe ratio is that it produces a ranking that is horizon-dependent (Levy, 1972; Levhari and Levy, 1977). If Sharpe ratios are calculated using monthly returns (as is typically done), but the investor's horizon is, say, 10 years, selecting the fund with the highest monthly Sharpe ratio may lead to a sub-optimal choice, even if unlimited borrowing at the lending rate *is* possible.

Another very popular performance measure employed for fund selection is alpha, initially suggested by Jensen (1968) in the context of the CAPM, and later extended to multiple-factor models. However, it is important to note that alpha is even theoretically *not* intended as a measure for fund selection, but rather only as a guide for *marginal* adjustments of portfolio weights relative to the benchmark portfolio. Even in this regard, alpha is quite problematic, because it is not a good guideline when the adjustments to portfolio weights are small but not infinitesimal (Levy and Roll, 2015).[3] It is thus not surprising that alpha does poorly as a performance measure in the context of fund selection.

Is there a better alternative to the Sharpe ratio and alpha? We find that the Geometric Mean (GM) constitutes such an alternative. The GM has several attractive theoretical properties. First, it coincides with the expected utility of a Bernoulli investor with logarithmic utility. Thus, the fund with the maximal GM is always included in the Second-degree Stochastic Dominance (SSD) efficient set (Levy, 2015). Thus, choosing the fund with the maximal GM cannot lead to absurd Stochastic Dominance violations that are possible with the Sharpe ratio.[4] Second, the GM can be approximated as a function of the mean and variance, increasing in the mean and decreasing in the variance: $GM \cong \mu - \frac{\sigma^2}{2}$, where μ is the mean and σ^2 the variance (Young and Trent, 1969; Markowitz, 2012). Thus, the fund with the maximal GM is typically on, or close to, the mean–variance efficient frontier. Expanding the GM approximation to higher moments reveals that it is increasing in skewness, conforming with the skewness-loving of most standard preferences (Levy, 2015). Finally, if returns are i.i.d. and the horizon approaches infinity, the fund with the highest GM almost surely yields a higher terminal wealth than any other fund. This has led several scholars to suggest that the

GM is the most relevant performance measure (Kelly, 1956; Latane, 1959; Markowitz, 1976), while others show that GM maximization is not perfectly aligned with expected utility maximization for constant relative risk aversion (CRRA) investors unless $\alpha = 1$, i.e. unless preferences are logarithmic (Samuelson, 1971; Merton and Samuelson, 1974). The purpose of this paper is not to rekindle this theoretical debate about investment for the very long run. Rather, we wish to examine the value of the GM as an approximate measure of investors' expected utilities in a realistic setting with limited borrowing, even when the horizon is not necessarily long. We find that the GM provides a much better measure than the Sharpe ratio in this setting. Moreover, under the assumption of i.i.d. returns it provides a ranking that is horizon-independent, which is another advantage relative to the Sharpe ratio.

The findings of this paper suggest that the geometric mean has key advantages over the Sharpe ratio (and certainly over alpha) as a portfolio performance measure. We therefore believe that it is a very important measure that should be at the focus of the investor's attention when selecting a fund.

2 Data and methodology

When the borrowing rate is higher than the lending rate, different investors may rank funds differently. Thus, there is no single performance measure that perfectly reflects the preferences of all investors. Our goal is to examine which performance measure is best aligned with investors' preferences, i.e. we are searching for the best single-number measure for ranking funds. We investigate this issue by looking at a large number of funds with known return distributions, and comparing the fund ranking obtained by direct expected utility maximization (choosing the optimal asset allocation between each fund and the risk-free asset) with the ranking by the candidate performance measure. The more aligned the two rankings, the better the performance measure. We examine the rank-correlation between the two rankings, the probability that the performance measure leads to a choice that is sub-optimal in terms of expected utility, and the certainty-equivalent loss induced by these cases. We perform this analysis for the main preferences suggested in the literature. The performance measures we examine are the Sharpe ratio, alpha, and the geometric mean.

We should note that our analysis takes the return distributions as known. The problem of estimation error is certainly an important one, but it applies to all the performance measures considered. Given that the *ex-ante* return distributions are unavailable in practice, virtually all fund rankings are based on *ex-post* returns, and this is also the approach employed here. This is not to imply that the *ex-post* return distribution is necessarily the best estimate of the *ex-ante* distribution, just that the empirical distribution provides a reasonable example of realistic return moments.

Given a fund with a return distribution $(R_1, p_1; R_2, p_2; \ldots, R_N, p_N)$, where R denotes the total return and p denotes the corresponding probability, the expected utility of an investor investing a proportion x in the fund and a proportion $1 - x$ in the risk-free asset is given by:

$$EU(x) = \sum_{i=1}^{N} p_i U[W_0(xR_i + (1 - x)R_f)], \quad (1)$$

where W_0 is the investor's initial wealth, and R_f is the total risk-free return (i.e. $1 + r_f$). The investor chooses the optimal asset allocation x^* so as to maximize $EU(x)$.

In the setting under which the Sharpe ratio was developed, there are no restrictions on x, and the same risk-free rate R_f applies to borrowing and to lending. In a more realistic setting, borrowing is limited by Regulation T to 100% of the initial wealth, i.e. $x < 2$. In addition, different risk-free rates apply for borrowing and for lending:

$$R_f = \begin{cases} R_f^L & \textit{if } x < 1, \textit{ lending} \\ R_f^B & \textit{if } x > 1, \textit{ borrowing} \end{cases},$$

with $R_f^B > R_f^L$. For a given utility function U, we calculate the maximal expected utility $EU(x^*)$ for each fund, and compare the ranking by expected utility with the ranking by the various candidate performance measures.

The fund's Sharpe ratio is given by $\frac{\sum_{i=1}^{N} p_i R_i - R_f^L}{\sigma(R)}$. Note that this conforms with the practice of employing the risk-free lending rate for calculating the Sharpe ratio. For each fund we also report its Fama–French 5-factor alpha, calculated from monthly returns, and its geometric mean, given by $GM = \prod_{i=1}^{N} R_i^{p_i}$.

The funds we use in our analysis are all U.S. domestic equity funds in the CRSP Survivor-Bias-Free Mutual Fund Database, with complete monthly return records over the 10-year July 2005–June 2015 period.[5] There are 10,145 funds in our sample. Choosing funds with complete return record introduces a selection bias, however, this bias is immaterial to the question addressed in this paper, and indeed, the results are almost identical when funds with incomplete records are also included in the analysis. The empirical return distributions are used, i.e. an equal probability of $p_i = \frac{1}{120}$ is assigned to each one of the 120 historical monthly returns, in line with the standard practice in most applications. The risk-free lending rate, and the Fama–French 5-factor returns (Fama and French, 2016) are taken from Ken French's data library.[6] Saunders and Schumacher (2000) report that the average banks' interest rate margins in the U.S. in 1995 was 4.2%. In January 2016 the prime rate in the U.S. was 3.5%, while the interest rate on deposits was very close to zero. To be conservative, we take the annual borrowing rate as 3.5% higher than the lending rate. For the July 2005–June 2015 period these average rates are $R_f^L = 0.82\%$ and $R_f^B = 4.32\%$. We also report results for other values of the difference between the borrowing and lending rates.

3 Results

Let us first consider the theoretical case where there are no limitations on borrowing, and the borrowing rate is equal to the lending rate. In this case, and if the return distributions are normal, the Sharpe ratio is the optimal performance measure for all investors who prefer more over less, risk-averters and risk-seekers alike (Levy and Levy, 2004; Levy *et al.*, 2012). Figure 2 shows the expected utility of a CRRA investor, $U(W) = \frac{W^{1-\alpha}}{1-\alpha}$, with a typical relative risk-aversion parameter of $\alpha = 1.5$ as a function of the Sharpe ratio, for each one of the 10,145 funds in our sample (we later consider CRRA preferences with other values of α, as well as other preference classes).[7] The alignment between the Sharpe ratio and expected utility is almost perfect: the Spearman rank-correlation is 0.998.

When selecting a fund, investors typically consider a limited menu of funds. Suppose that the investor is presented with a menu of 10 funds, and he chooses the fund with the highest Sharpe ratio. What is the probability that the fund chosen is not the one that actually maximizes his expected utility? If the 10 funds are randomly selected,[8] the answer is 7.7%, and the average certainty-equivalent loss, on an annual basis, is only 0.04% (these are the average results for 1,000 random 10-fund sets).[9] Thus, in the unrealistic case of unlimited borrowing at the lending rate, the Sharpe ratio is an excellent performance measure, even though the distributions employed are the empirical ones, and are obviously not normal. This is consistent with the findings of Levy and Markowitz (1979), Kroll *et al.* (1984), and Simaan (1993), who show that in the context of

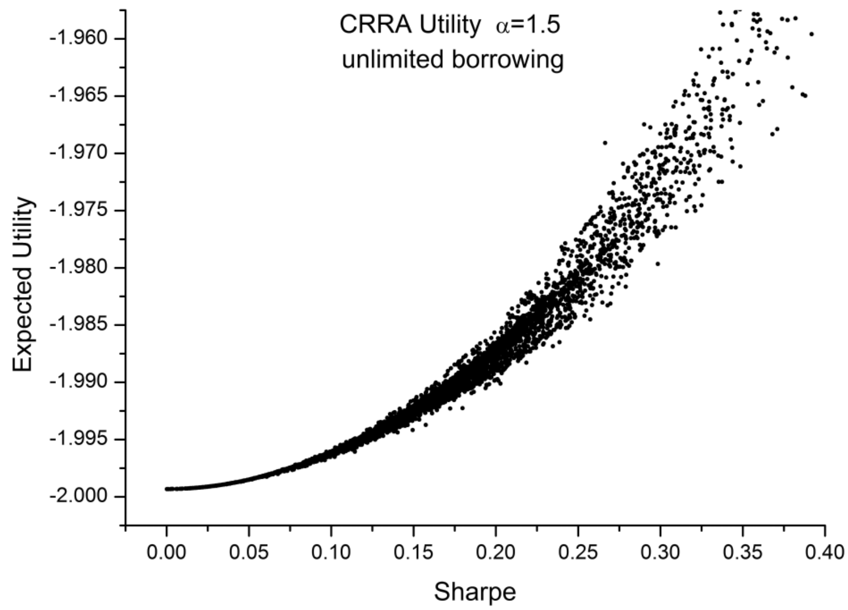


Figure 2 The relationship between expected utility and the Sharpe ratio when borrowing is unlimited, and the borrowing rate is equal to the lending rate.

portfolio selection, assuming normality typically implies only a very small economic loss. Table 1 provides the median parameters of the mutual fund return distributions, and shows that for the vast majority of funds normality can be rejected even at the 1% significance level. Still, our results indicate that the fact that the distributions are not normal does not substantially hinder the performance of the Sharpe ratio.

Table 1 The median parameters of the monthly return distributions. The Jarque–Bera normality test is employed.

	Mutual funds	Hedge funds
Mean	0.59%	0.49%
Standard deviation	4.32%	4.38%
Skewness	−0.75	−0.57
Kurtosis	5.33	5.34
Percentage of funds for which normality if rejected at the 1% significance level	90.6%	91.9%
Percentage of funds for which normality if rejected at the 5% significance level	97.1%	97.2%
Percentage of funds for which normality if rejected at the 10% significance level	98.2%	98.5%

Let us now turn to the realistic case where borrowing is limited to 100%, and the borrowing rate is higher than the lending rate, as detailed above. Figure 3 shows the relationship between the Sharpe ratio and the expected utility in this case. The rank-correlation drops from 0.998 to only 0.372. If the investor is presented with a menu of 10 randomly selected funds and he chooses the one with the highest Sharpe ratio, in 82.5% of the cases this fund is *not* the fund that actually maximizes his expected utility, i.e. in 82.5% of the cases the Sharpe ratio leads to a sub-optimal choice. The average annual certainty-equivalent loss induced by these cases is 4.22% (again, averaged over 1,000 random 10-fund sets). When the menu of funds to choose from consists of 100 randomly drawn funds, the Sharpe ratio leads to a sub-optimal choice in 98% of the cases,

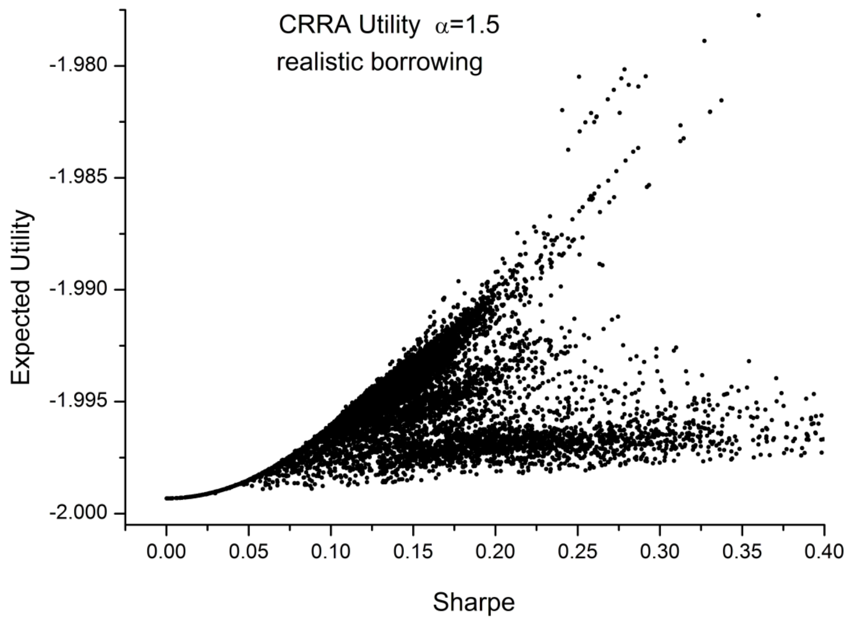


Figure 3 The relationship between expected utility and the Sharpe ratio when borrowing is limited, and the borrowing rate is higher than the lending rate.

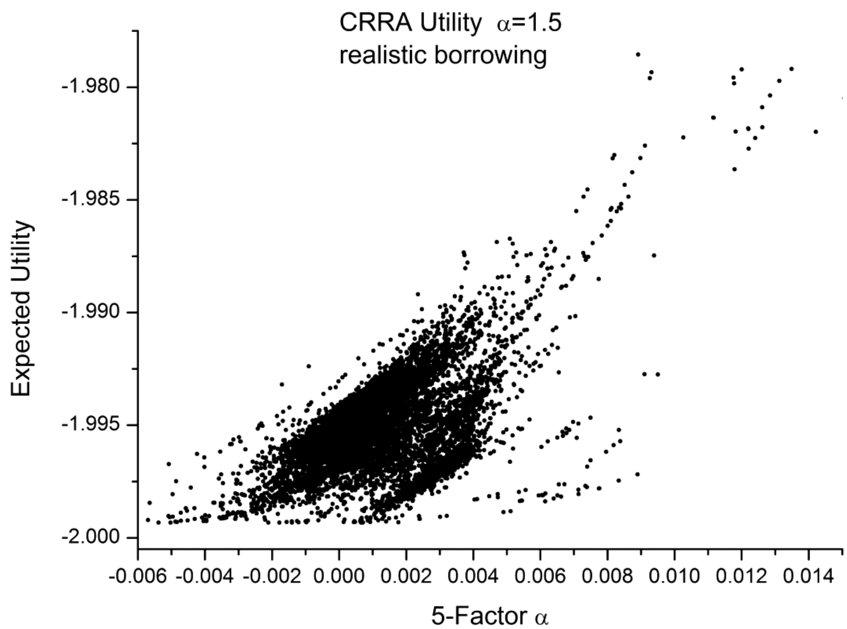


Figure 4 The relationship between the Fama–French 5-Factor alpha and the expected utility. The Spearman rank-correlation is 0.188.

and the average annual certainty-equivalent loss is 9.24%. Thus, employing the Sharpe ratio for fund selection with realistic borrowing conditions induces a very significant economic loss.

Figure 4 presents the same analysis, but when the performance measure employed is the Fama–French 5-factor alpha.[10] The figure reveals that the 5-factor alpha is not better than the Sharpe ratio. In fact, the correlation between the expected utility and the 5-factor alpha is even lower, at 0.188. If the investor is presented with a menu of 10 randomly selected funds and he chooses the one with the highest 5-factor alpha, this fund is *not* the fund that actually maximizes his expected utility in 69.8% of the cases. The average annual certainty-equivalent loss induced by employing alpha is 2.61%. These are rather bleak results about the usefulness of the Sharpe ratio and alpha in a setting with realistic borrowing conditions.

In contrast to the above results, the GM performs much better: Figure 5 depicts the relationship between expected utility and the GM. The rank-correlation is 0.974. If the investor is presented with a menu of 10 randomly selected funds and he chooses the one with the highest GM, this fund is *not* the fund that actually maximizes his expected utility in only 18% of the cases. The average annual certainty-equivalent loss induced by employing the GM instead of direct expected utility maximization is only 0.15%. This is 28 (!) times lower than the average loss induced by employing the Sharpe ratio, and 17 times lower than the loss induced by employing alpha.

As a robustness check, Figure 6 shows the same analysis for hedge funds, rather than mutual funds. In this analysis all U.S. hedge funds in the Thompson-Reuters database with complete monthly return records over the July 2005–June 2015 period are employed. Again, the results are almost identical when funds with incomplete records are also included. The results are for the case of realistic borrowing, and the CRRA utility function with $\alpha = 1.5$. The figure reveals that the GM is better aligned with expected utility than both the Sharpe ratio and 5-factor alpha. For the sake of brevity, in what follows we report results only for mutual funds, but the corresponding results for hedge funds are very similar.

The source of the problem of using the Sharpe ratio when borrowing is limited is illustrated by Figure 7. This figure shows all of the mutual funds in our sample on the mean–variance plane

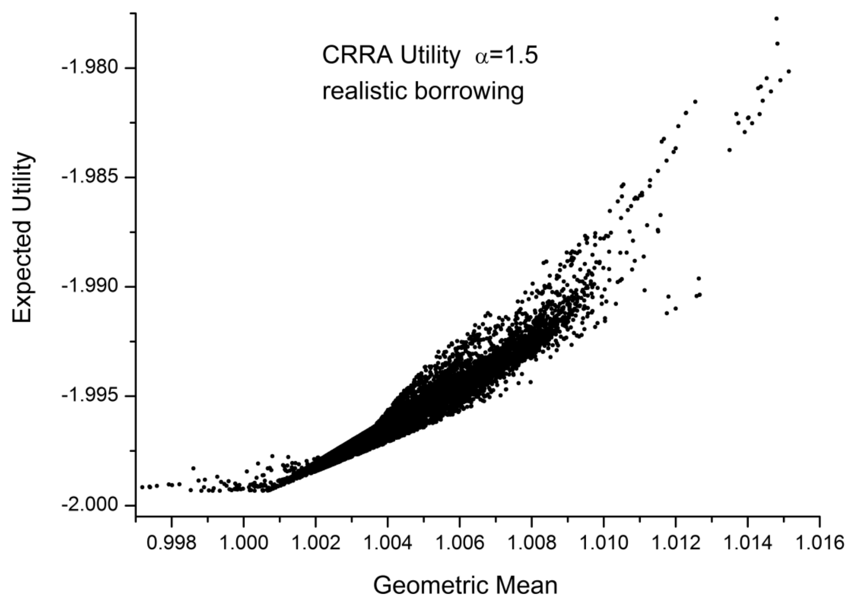


Figure 5 The relationship between the geometric mean and the expected utility. Borrowing is limited, and the borrowing rate is higher than the lending rate. The Spearman rank-correlation is 0.974.

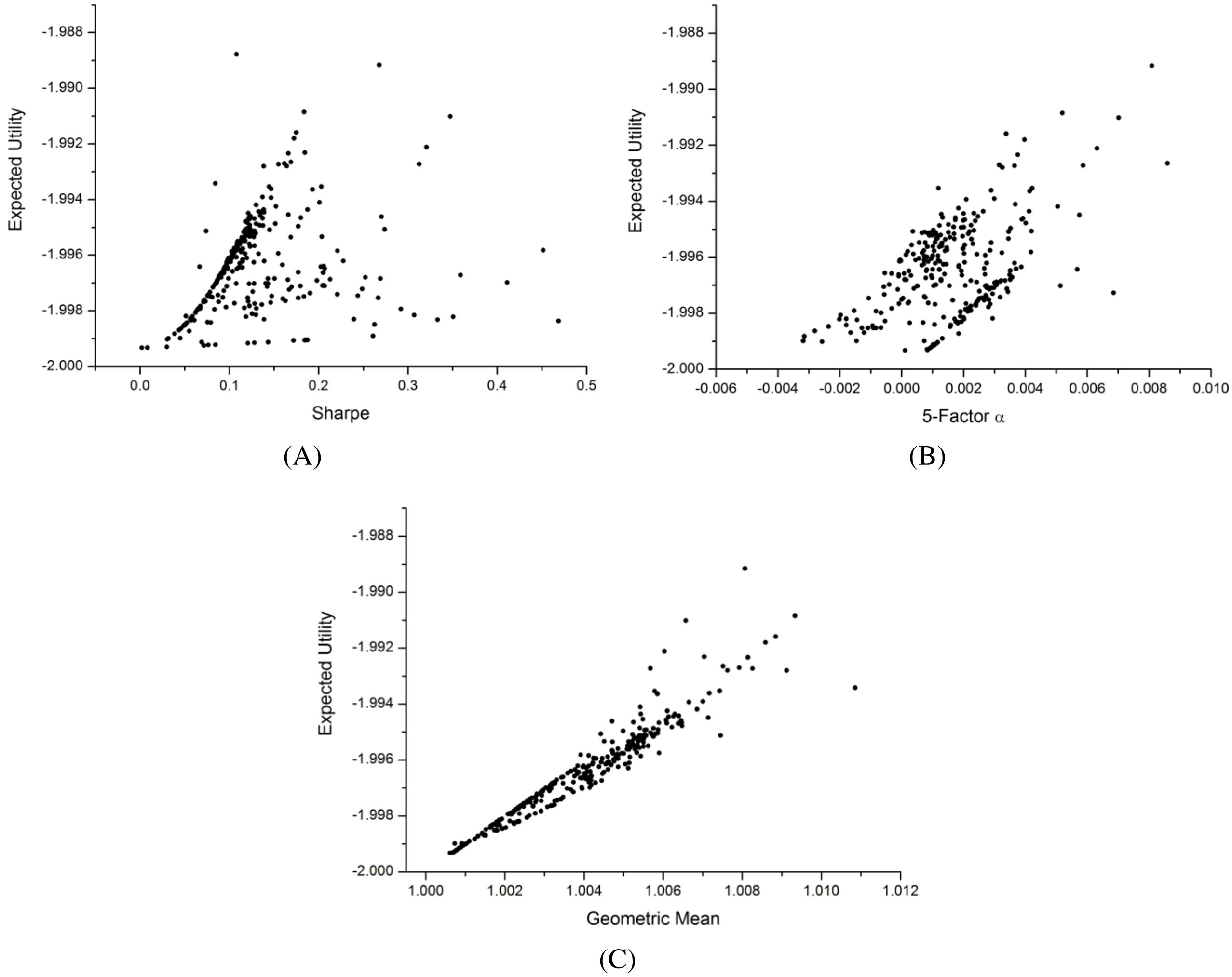


Figure 6 Hedge fund data with realistic borrowing. The expected utility is for a CRRA investor with $\alpha = 1.5$. Panel A: The relationship between the Sharpe ratio and the expected utility–Spearman rank-correlation of 0.36. Panel B: The relationship between the 5-factor alpha and the expected utility–Spearman rank-correlation of 0.42. Panel C: The relationship between the geometric mean and the expected utility–Spearman rank-correlation of 0.96.

(with monthly parameters). The top 10 funds by the Sharpe ratio are depicted by hollow circles, and the top 10 funds by the GM are depicted by stars. The figure reveals that the Sharpe ratio is typically more sensitive to the standard deviation that appears in its denominator, while the GM is more sensitive to the average return. If unlimited borrowing at the lending rate would have been possible, the top Sharpe funds would clearly dominate the top GM funds: one could lever-up the Sharpe funds and obtain portfolios on the straight line directly above the GM funds-portfolios with the same standard deviation as the GM funds, but with a higher mean. However, as the figure reveals, this would require leverage of hundreds of percent, which is clearly unrealistic. When such leverage is not possible, many investors would be better off in terms of expected utility with the GM funds, as the results so far have indicated.

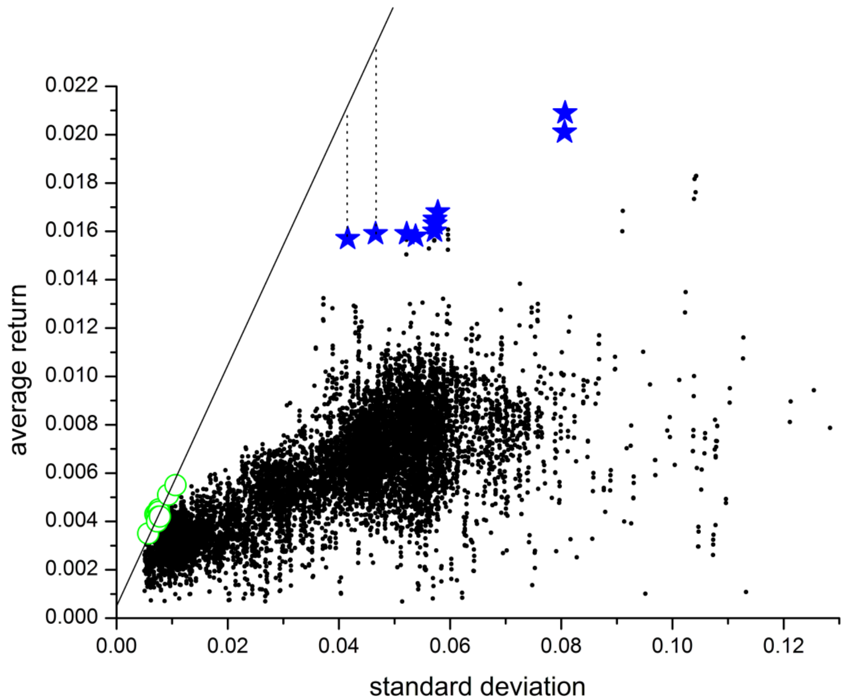
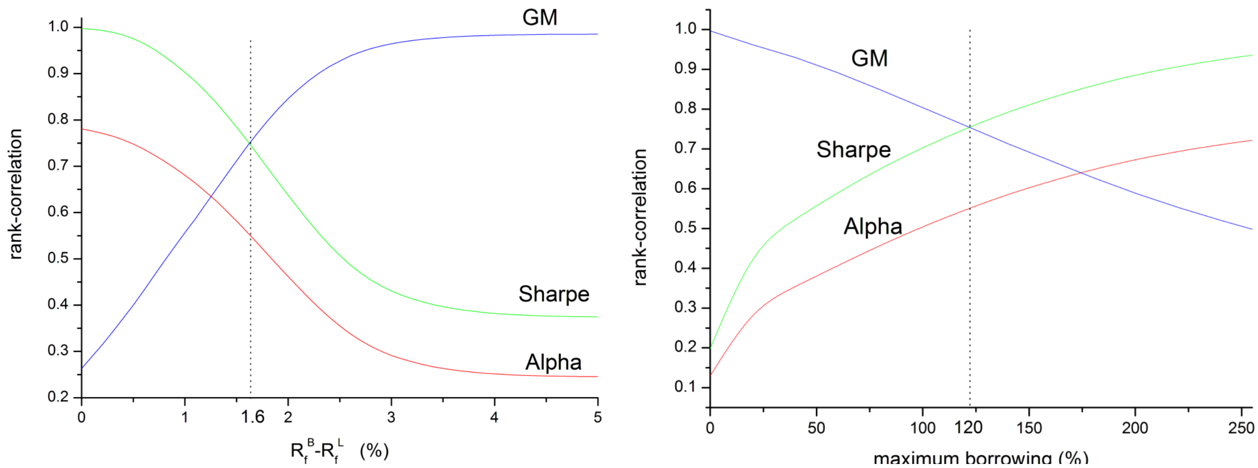


Figure 7 The funds in the mean-standard-deviation plane. The top 10 funds by the Sharpe ratio are depicted by hollow circles. The top 10 funds by the GM are depicted by stars.

There are two differences between the theoretical borrowing setup of the CAPM and realistic borrowing: (i) the fact that the borrowing rate is higher than the lending rate, and (ii) the direct constraint on the maximal amount of borrowing (Reg T). We analyze the effect of each of these factors separately, and display the results in Figure 8. Panel A of the figure shows the correlation between the expected utility and the Sharpe ratio, alpha, and the GM, as a function of the difference between the annual borrowing and lending rates, $R_f^B - R_f^L$, where there is no direct restriction on the amount of borrowing. For the Sharpe ratio, $R_f^B - R_f^L = 0$ corresponds to the case shown in Figure 2 – the correlation is very high (0.998, as reported above). The correlation between the Sharpe ratio and the expected utility decreases monotonically with the difference between the borrowing and lending rates, and it levels-off at a value of about 0.4. In contrast, the correlation between the GM and the expected utility increases with the difference between the borrowing and lending rates. The figure reveals the crossover point at which the GM becomes a better performance measure than the Sharpe ratio: it is at $R_f^B - R_f^L = 1.6\%$. An investor who pays an annual borrowing rate that is 1.6% or more higher than the lending rate should employ the GM rather than the Sharpe ratio, even if there are no restrictions on borrowing (Panel A). Alpha is always inferior to the Sharpe ratio.

Panel B of Figure 8 shows the correlation as a function of the limit on the maximal borrowing, where it is assumed that the borrowing rate is equal to the lending rate. The limit on borrowing is expressed as a percentage of the initial capital, i.e. 100% means borrowing an amount equal to the investor's own initial capital and investing 200% of the initial capital in the fund, i.e. 100% corresponds to the constraint imposed by Reg T. When there are no restrictions at all on borrowing, the

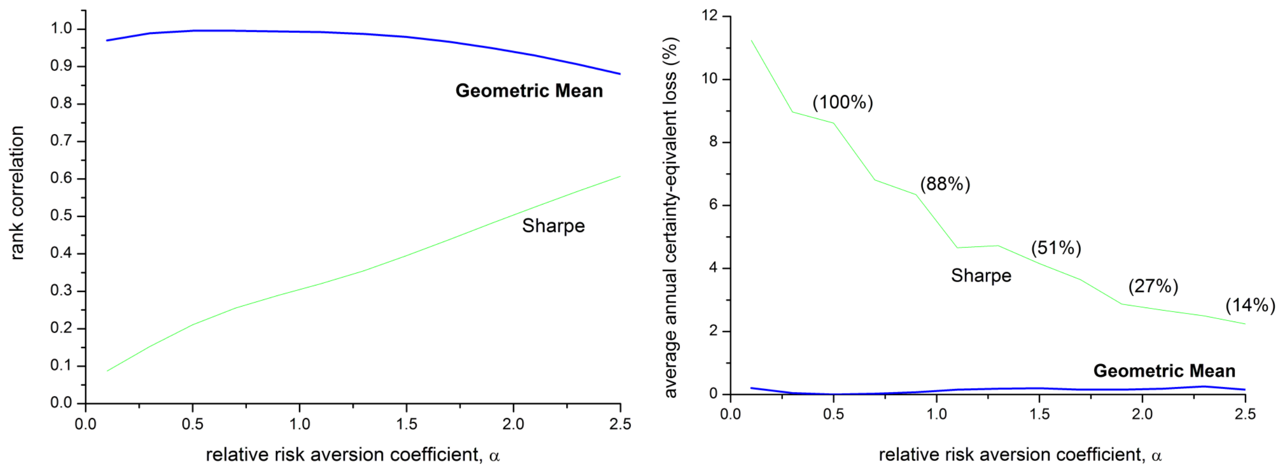


(A) (B)

Figure 8 The correlation between the expected utility of a CRRA investor with relative risk aversion of 1.5 and the three performance measures, as a function of the annual difference between the borrowing rate and the lending rate, with no borrowing constraints (Panel A), and as a function of the borrowing constraint, when the borrowing rate is assumed to be equal to the lending rate (Panel B).

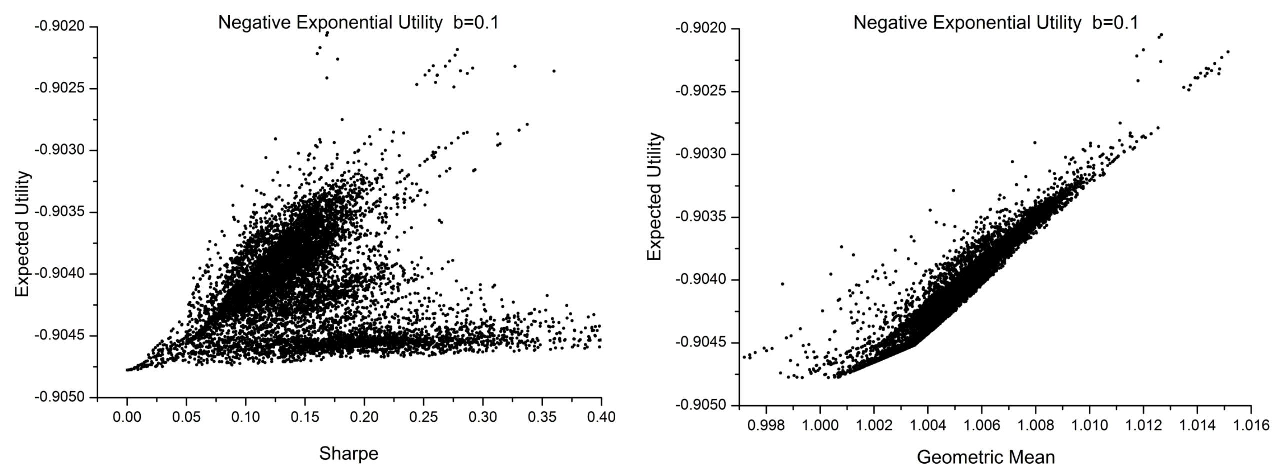
Sharpe ratio is the optimal performance measure. The tighter the borrowing constraint, the worse the Sharpe ratio becomes. In contrast, the performance of the GM declines as maximal borrowing level increases. This is because one can obtain better portfolios by leveraging-up the funds with the highest Sharpe ratios, as illustrated in Figure 7. The crossover between the Sharpe ratio and the GM occurs at approximately 120%: an investor who is restricted to borrow less than 120% of his capital, should employ the GM, even if he borrows at the lending rate. In the realistic case where both effects are present, i.e. borrowing is limited *and* it is at a rate higher than the lending rate, the advantage of the GM becomes even larger.

The results so far are for a specific utility function: CRRA preferences with relative risk-aversion parameter $\alpha = 1.5$. It turns out that these results are quite typical for other preferences as well. Panel A of Figure 9 shows the rank-correlation between the Sharpe ratio and the expected utility for CRRA preferences with different values of the relative risk-aversion parameter, α (thin line). The bold line shows the rank correlation between the GM and the expected utility. The empirical evidence suggests that the relevant range for α is between 1 and 2 (see Arrow, 1971; Tobin and Dolde, 1971; Friend and Blume, 1975; Kydland and Prescott, 1982). The thin line in Panel B of the figure shows the average annual certainty-equivalent loss when the investor is presented with a menu of 10 funds and he chooses the fund with the highest Sharpe ratio (rather than choosing by direct expected utility maximization). The bold lines show the results corresponding to the choice of the fund with the highest GM. Notice that the Sharpe ratio does better as risk-aversion increases, because the more risk-averse the investor, the less relevant borrowing becomes. The numbers in parentheses indicate the optimal leverage in the portfolio with the maximal Sharpe ratio for the values $\alpha = 0.5$, 1, 1.5, 2 and 2.5. The higher the relative risk aversion, the lower the leverage, and therefore the lower the cost induced by employing the Sharpe ratio. Even so, the GM is clearly superior to the Sharpe ratio over the entire range of relevant risk-aversion values.



(A) (B)

Figure 9 The performance of the Sharpe ratio and the geometric mean for different values of the relative risk-aversion parameter. Panel A shows the Spearman rank-correlation between the performance measure and the expected utility. Panel B shows the average annual certainty-equivalent loss when one fund is selected out of a random set of 10 funds by the performance measure. The numbers in parentheses indicate the optimal leverage employed with the fund with the maximal Sharpe ratio (averaged over all 1,000 10-fund sets). As risk-aversion increases, the leverage decreases, and thus the error induced by employing the Sharpe ratio decreases.



(A) (B)

Figure 10 The relationship between expected utility and the performance measures when borrowing is limited, and the borrowing rate is higher than the lending rate for negative exponential utility with initial wealth $W_0 = 1$ and $b = 0.1$. Panel A: the Sharpe ratio –– rank-correlation = 0.086. Panel B: the Geometric Mean –– rank-correlation = 0.969.

Another utility function commonly employed in the literature is the negative exponential. Panel A of Figure 10 shows the relationship between the Sharpe ratio and the expected utility for an investor with negative exponential utility with initial wealth $W_0 = 1$ and an absolute risk-aversion parameter of $b = 0.1$. Panel B shows the corresponding relationship between the GM and the expected utility. The probability of choosing a sub-optimal fund from a menu of 10 funds, and

Table 2 Performance of the Sharpe ratio, 5-factor alpha, and the GM for various preferences.

	Sharpe			5-Factor alpha			Geometric mean		
Preference	Correlation	Wrong choice* (%)	Certainty-equivalent loss** (%)	Correlation	Wrong choice* (%)	Certainty-equivalent loss** (%)	Correlation	Wrong choice* (%)	Certainty-equivalent loss** (%)
CRRA									
$\alpha = 0.5$	0.173	88.5	8.65	0.034	74.2	5.64	0.995	0.7	0.00
$\alpha = 1.0$	0.279	82.9	5.79	0.116	73.3	3.87	0.992	20.2	0.11
$\alpha = 1.5$	0.372	82.5	4.22	0.188	69.8	2.61	0.974	18.0	0.15
$\alpha = 2.0$	0.479	80.3	3.07	0.273	68.1	2.00	0.934	23.4	0.18
$\alpha = 2.5$	0.586	78.6	2.36	0.365	67.4	1.54	0.873	34.2	0.26
Generalized CRRA $\alpha = 1.5$									
$A = 0.5$	0.429	79.7	3.58	0.286	65.4	2.31	0.969	19.6	0.15
$A = 1$	0.344	79.6	4.52	0.225	67.2	2.99	0.989	20.3	0.13
$A = 2$	0.277	82.3	6.35	0.176	67.0	4.20	0.995	11.9	0.07
Negative exponential									
$b = 0.01$	0.058	89.6	11.29	0.056	71.8	7.01	0.957	19.2	0.23
$b = 0.1$	0.086	88.5	10.73	0.068	69.7	7.11	0.969	15.2	0.13
$b = 1$	0.302	82.9	5.71	0.195	66.7	3.55	0.993	16.3	0.13

For each of the performance measures and each preference, the table reports the Spearman rank-correlation between the expected utility and the performance measure for all funds. The table also reports the probability that choosing the fund with the highest ranking according to the performance measure, out of a menu of 10 randomly selected funds, leads to a choice that is sub-optimal by expected utility, and the average annual certainty-equivalent loss induced by these sub-optimal choices. 1,000 random 10-fund sets are drawn for each preference.

*A set of 10 funds is randomly selected. The investor chooses the fund that is ranked first among the 10 according to the performance measure considered (Sharpe, alpha, or the GM). Wrong choice indicates the percentage of times (averaged over 1,000 sets of 10 funds) that the fund chosen is *not* the fund that actually maximizes the investor's expected utility (among the 10).

**In the cases that the wrong fund is selected, as explained above, a certainty-equivalent loss is induced. The average annual certainty-equivalent loss is calculated as detailed in Endnote 9.

the average certainty-equivalent loss induced by such suboptimal choices are reported in Table 2. The table also reports these values, and the rank correlation, for other values of the absolute risk-aversion parameter, b. Also reported in the table are the results for the generalized power utility function, $U(W) = \frac{(W+A)^{1-\alpha}}{1-\alpha}$, suggested by Litzenberger and Rubinstein (1976), Kroll *et al.* (1984), Samuelson (1989), and others. In all cases, the GM is better than both the Sharpe ratio and alpha by all three criteria considered: the rank-correlation, the probability of choosing a sub-optimal fund, and the certainty equivalent loss relative to the optimal choice.

4 The investment horizon

The above results show that when borrowing is limited, and is at a rate higher than the risk-free lending rate, the Sharpe ratio and alpha are very problematic as guidelines for fund selection. Another problem that arises when using the Sharpe ratio is that the ranking by the Sharpe ratio is sensitive to the investor's planned horizon

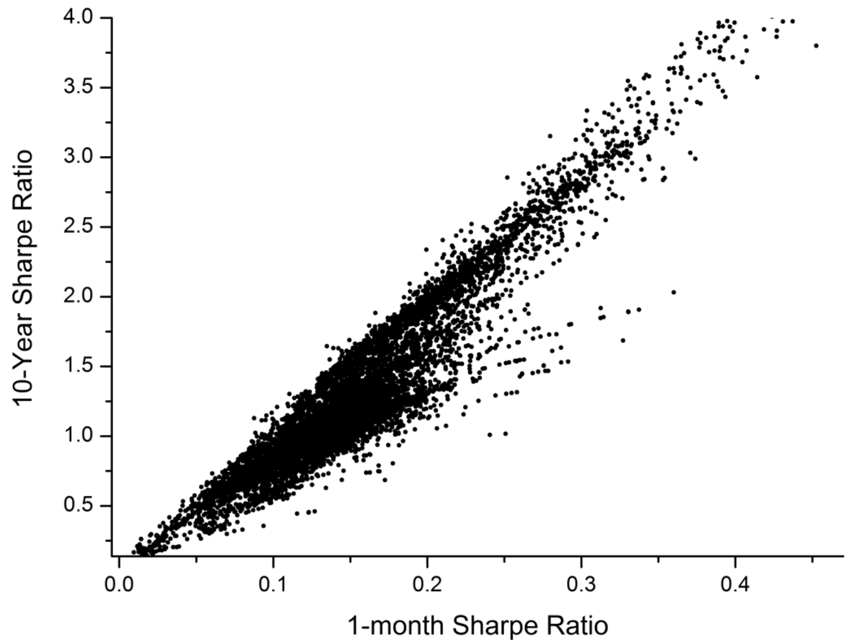


Figure 11 The relationship between the Sharpe ratio calculated with monthly returns, and the Sharpe ratio calculated with 10-year returns. The ranking by the Sharpe ratio is *not* invariant to the horizon.

(Levy, 1972; Levhari and Levy, 1977). For example, suppose that there are no restrictions on borrowing, borrowing is at the lending rate, and funds are ranked by the Sharpe ratio calculated by using monthly returns (as is virtually always done in practice). For an investor who has a 10-year horizon, the relevant ranking is by the Sharpe ratio calculated by using 10-year returns. Figure 11 shows the relationship between the monthly Sharpe ratio and the 10-year Sharpe ratio, which is obtained by randomly drawing 120 monthly returns from the empirical monthly return distribution (with replacement), and simulating 1,000 10-year returns for each fund. It is evident from the figure that even in the unrealistic case of unlimited borrowing at the lending rate, employing the monthly Sharpe may lead investors with a longer investment horizon to suboptimal choices. In contrast, the GM yields a ranking that is independent of the investor's horizon for i.i.d. returns. Under i.i.d. returns, a fund's H-period geometric mean is GM^H, where GM is the 1-period geometric mean.[11] Thus, the ranking is independent of the horizon, H. This means that the investor can safely employ the 1-period geometric mean, regardless of his planned investment horizon.

5 Conclusions

Under a theoretical setting with unlimited borrowing at the same rate as the lending rate and normal return distributions, the Sharpe ratio is the optimal portfolio performance measure for all investors who prefer more to less. This strong result, as well as the simplicity and elegance of the Sharpe ratio, has made it the most widely accepted performance measure, by academics and practitioners alike. It should be noted that even in this theoretical setting one should make sure to use the Sharpe ratio calculated based on returns matching the investor's planned holding period, as the ranking by the Sharpe ratio is *not* invariant to the holding period employed.

Of course, as most individuals who have taken a loan know all too well, the borrowing rate is significantly higher than the lending rate. In addition, Federal Regulation T imposes a constraint on the maximal level of borrowing.

This study shows that in a realistic setting with limited borrowing at a rate higher than the lending rate the Sharpe ratio is not a good performance measure: it often ranks funds differently than the ranking by investors' direct expected utility maximization. Employing the Sharpe ratio as a criterion for fund selection typically leads to a large economic loss. Alpha, which is another very popular performance measure, is even worse in this regard.

Our analysis reveals that the geometric mean is a superior alternative performance measure. It is much better aligned with investors' preferences than the Sharpe ratio or alpha. This is true for a wide range of preferences. In addition, the geometric mean has the advantage of producing a ranking that is invariant to the investment horizon for i.i.d. returns.

Why does the geometric mean perform so well? One reason is that it reflects the preferences of the log utility function, and this turns out to produce a good approximate ranking for other preferences as well. The geometric mean is well-approximated by a function of the mean and the variance, increasing in the mean and decreasing in the variance, and thus portfolios with high geometric means tend to be close to the mean–variance efficient frontier. Additionally, the geometric mean points to the fund with the highest growth rate. Over the long run, this fund almost surely yields a higher terminal wealth than any other fund. This paper shows that the strong early intuitions of Kelly (1956), Latane (1959), and Markowitz (1976) about the importance of the geometric mean have great merit even when the investment horizon is not very long.

The evidence in this paper suggests that the geometric mean should play a much more central role as a performance measure role than it currently does. It may very well be the single most important number to consider in evaluating portfolio performance.

Acknowledgments

I am very grateful to Gifford Fong, the Editor, and to an anonymous referee for their helpful comments and suggestions. I thank Haim Levy and Harry Markowitz for their advice and insights.

Notes

[1] It was later shown that this is true not only for risk-averse investors, but also for all Prospect Theory investors as well, and in fact, for any investor who prefers more to less (Levy and Levy, 2004; Levy *et al.*, 2012).

[2] The popular Morningstar star rating, though based on a somewhat different methodology, is very highly correlated with the Sharpe ratio (Sharpe, 1998). Most academic studies employ the Sharpe ratio, alpha, or both, to measure portfolio performance. For some examples see Fama and French (2002), Phalippou and Gottschalg (2009), Tu and Zhou (2011), Moskowitz *et al.* (2012), and Barroso and Santa-Clara (2015).

[3] Alpha can be employed as the excess return a manager produces relative to the fund's risk exposure. However, Roll (1978) shows that in this setting performance is very sensitive to the benchmark employed.

[4] For example, consider Fund A that yields a return of either 16% or 20% with equal probabilities, and Fund B that yields either 20% or 40% with equal probabilities. Assume for simplicity a risk-free rate of 0%. Then Fund A yields a much higher Sharpe ratio than B ($18/2 = 9$ compared to $30/10 = 3$), although Fund B clearly dominates by First-order Stochastic Dominance (see also Goetzmann *et al.*, 2007). This absurd result does not occur with the GM, as the GM of the total return of Fund A is 1.1798 (or 17.98%), while the GM of Fund B is 1.2961 (or 29.61%).

[5] See Elton *et al.* (2001) for the pros and cons of this database relative to the Morningstar database.

[6] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[7] We include in the analysis only the 10,019 funds with a positive allocation to the fund, i.e. $x^* > 0$ (i.e. all funds with a mean return exceeding the risk-free lending rate, see Arrow, 1971). For all funds with $x^* = 0$ the expected utility is the same – the expected utility from investing 100% at the risk-free lending rate.

[8] In practice, the set of funds will likely not be chosen at random, but rather, funds to be included in the menu are pre-screened by various criteria. Thus, funds with poor performance will typically not be included in the menu. Our setup makes it "easier" for the performance measure to pick the optimal fund, as the differences in performance are typically larger than under the pre-screening setup. This makes our criticism of the Sharpe ratio when borrowing is limited even stronger.

[9] The annual certainty-equivalent loss is calculated as follows. If the investor chooses fund A, that has the highest Sharpe ratio, while the fund that actually maximizes his expected utility is fund B, we calculate the certainty equivalent for fund A, CE_A, as the solution to: $U(CE_A) = \frac{CE_A^{1-\alpha}}{1-\alpha} = EU_A(x_A^*)$, where x_A^* is the optimal asset allocation between fund A and the risk-free asset, and the expected utility $EU_A(x_A^*)$ is given by Equation (1), where without loss of generality we take $W_0 = 1$. Similarly, CE_B is calculated for fund B. $CE_B - CE_A$ is the certainty-equivalent loss in monthly terms. The annual certainty-equivalent loss is calculated as $(CE_B)^{12} - (CE_A)^{12}$.

[10] To avoid confusion, we will use α for the relative risk-aversion parameter, and spell out "alpha" for the 5-factor alpha.

[11] For a 1-period investment we have $\log(GM_{1-period}) = E[\log(R)]$, where the right-hand side is the expectation of the 1-period log total returns. For the H-period investment we have:
$\log(GM_{H-period}) = E[\log(R_1 R_2 \cdots R_H)] = \sum_{i=1}^{H} E[\log(R_i)]$, which under the assumption of i.i.d. returns, becomes: $\log(GM_{H-period}) = H \cdot E[\log(R)] = H \cdot \log(GM_{1-period})$, or, $GM_{H-period} = (GM_{1-period})^H$.

References

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing*. North-Holland.

Barroso, P. and Santa-Clara, P. (2015). "Momentum Has Its moments," *Journal of Financial Economics* **116**(1), 111–120.

Elton, E. J., Gruber, M. J., and Blake, C. R. (2001). "A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases," *The Journal of Finance* **56**(6), 2415–2430.

Fama, E. F. and French, K. R. (2002). "The Equity Premium," *The Journal of Finance* **57**(2), 637–659.

Fama, E. F. and French, K. R. (2016). "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies* **29**(1), 69–103.

Fortune, P. (2000). "Margin Requirements, Margin Loans, and Margin Rates: Practice and Principles," *New England Economic Review*, 19.

Friend, I. and Blume, M. E. (1975). "The Demand for Risky Assets," *The American Economic Review*, 900–922.

Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I. (2007). "Portfolio Performance Manipulation and Manipulation-Proof Performance Measures," *Review of Financial Studies* **20**(5), 1503–1546.

Jensen, M. C. (1968). "The Performance of Mutual Funds in the Period 1945–1964," *The Journal of Finance* **23**(2), 389–416.

Kelly, J. (1956). "A New Interpretation of Information Rate," *IRE Transactions on Information Theory* **2**(3), 185–189.

Kroll, Y., Levy, H., and Markowitz, H. M. (1984). "Mean-Variance versus direct Utility Maximization," *The Journal of Finance* **39**(1), 47–61.

Kydland, F. E. and Prescott, E. C. (1982). "Time to Build and Aggregate Fluctuations," *Econometrica* **50**, 1345–1370.

Latane, H. A. (1959). "Criteria for Choice among Risky Ventures," *Journal of Political Economy* **67**(2), 144–155.

Levhari, D. and Levy, H. (1977). "The Capital Asset Pricing Model and the investment Horizon," *The Review of Economics and Statistics*, 92–104.

Levy, H. (1972). "Portfolio Performance and the Investment Horizon," *Management Science* **18**(12), B-645.

Levy, H. (2015). *Stochastic Dominance: Investment Decision Making Under Uncertainty*. Springer.

Levy, H., De Giorgi, E. G., and Hens, T. (2012). "Two Paradigms and Nobel Prizes in Economics: A Contradiction or Coexistence?," *European Financial Management* **18**(2), 163–182.

Levy, H. and Levy, M. (2004). "Prospect Theory and Mean-Variance Analysis," *Review of Financial Studies* **17**(4), 1015–1041.

Levy, H. and Markowitz, H. M. (1979). "Approximating Expected Utility by a Function of Mean and Variance," *The American Economic Review* **69**(3), 308–317.

Levy, M. and Roll, R. (2015). "Seeking Alpha? It's a Bad Guideline for Portfolio Optimization," *The Journal of Portfolio Management* **42**(5), 107–112.

Litzenberger, R. and Rubinstein, M. (1976). "The Strong Case for the Generalized Logarithmic Utility Model as the Premier Model of Financial Markets," *The Journal of Finance* **31**(2), 551–571.

Markowitz, H. M. (1976). "Investment for the Long Run: New Evidence for an Old Rule," *The Journal of Finance* **31**(5), 1273–1286.

Markowitz, H. M. (2005). "Market Efficiency: A Theoretical Distinction and So What?," *Financial Analysts Journal* **61**(5), 17–30.

Markowitz, H. (2012). "Mean-variance Approximations to the Geometric Mean," *Annals of Financial Economics* **7**(01), 1250001.

Merton, R. C. and Samuelson, P. A. (1974). "Fallacy of the Log-Normal Approximation to Optimal Portfolio Decision-Making over Many Periods," *Journal of Financial Economics* **1**(1), 67–94.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). "Time Series Momentum," *Journal of Financial Economics* **104**(2), 228–250.

Phalippou, L. and Gottschalg, O. (2009). "The Performance of Private Equity Funds," *Review of Financial Studies* **22**(4), 1747–1776.

Roll, R. (1978). "Ambiguity When Performance is Measured by the securities Market Line," *The Journal of Finance* **33**(4), 1051–1069.

Samuelson, P. A. (1971). "The "Fallacy" of Maximizing the Geometric Mean in Long Sequences of Investing or Gambling," *Proceedings of the National Academy of Sciences* **68**(10), 2493–2496.

Samuelson, P. A. (1989). "The Judgment of Economic Science on Rational Portfolio Management: Indexing, Timing, and Long-Horizon Effects," *The Journal of Portfolio Management* **16**(1), 4–12.

Saunders, A. and Schumacher, L. (2000). "The Determinants of Bank Interest Rate Margins: An International Study," *Journal of International Money and Finance* **19**(6), 813–832.

Sharpe, W. F. (1966). "Mutual Fund Performance," *The Journal of Business* **39**(1), 119–138.

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

Sharpe, W. F. (1998). "Morningstar's Risk-adjusted Ratings," *Financial Analysts Journal* **54**(4), 21–33.

Simaan, Y. (1993). "What Is the Opportunity Cost of Mean-Variance Investment Strategies?," *Management Science* **39**(5), 578–587.

Tobin, J. and Dolde, W. (1971). *Wealth, Liquidity, and Consumption* (No. 311). Cowles Foundation for Research in Economics, Yale University.

Tu, J. and Zhou, G. (2011). "Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies," *Journal of Financial Economics* **99**(1), 204–215.

Young, W. E. and Trent, R. H. (1969). "Geometric Mean Approximations of Individual Security and Portfolio Performance," *Journal of Financial and Quantitative Analysis* **4**(02), 179–199.

Keywords: Portfolio performance; Sharpe ratio; alpha; geometric mean; investment horizon