rs o... quilibrium Stock Prices.
porate Finance, and the Theory
Securities under Asymmetric
definite, Heteroskedasticity and
ica, 55, 703–708.
ing Causality Tests. Review of
Asymptotic Distribution of the
. Oxford Bulletin of Economics
. and the Performance of Index
ticipation Units Market. Journal
Root in Time Series Regression.
oot in Time Series Regression.
elationship between Stock Index
s, 13, 645–664.
48, 1–48.
Trading Volume Relations in the
191–220.
and Stock Index Futures Returns.
168.
ex Futures. Review of Financial
Tr... ission of Information in
M... Hypothesis. Journal of
r Correction in Stock Index and
. Journal of Futures Markets, 13,
ance, 46, 1645–1663.

A DOUBLE SHARPE RATIO

Matthew R. Morey and H. D. Vinod

ABSTRACT

Sharpe's (1966) portfolio performance ratio, the ratio of the portfolio's expected return to its standard deviation, is a very well known tool for comparing portfolios. However, due to the presence of random denominators in the definition of the ratio, the sampling distribution of the Sharpe ratio is difficult to determine. Hence, measurement of estimation risk of the Sharpe ratio has often been ignored in practice. This paper uses the bootstrap methodology to suggest a new "Double" Sharpe ratio that allows an investor to make a tradeoff between risk-adjusted performance and estimation risk using the same weighting for estimation risk as the original Sharpe ratio uses for standard deviation. The use of this Double Sharpe ratio along with the original ratio, provides investors with a relatively simple method for measuring risk- and estimation risk-adjusted performance.

INTRODUCTION

The Sharpe (1966) portfolio performance ratio, the ratio of a portfolio's expected return to its standard deviation, is widely used and cited in the literature and pedagogy of finance. Indeed, in a recent finance literature search, over 30 papers published between 1995–1998 cited the Sharpe ratio.[1]

Despite its popularity, the Sharpe ratio suffers from a methodological limitation: because of the presence of random denominators in its definition and the difficulty in determining the sample size needed to achieve asymptotic

Advances in Investment Analysis and Portfolio Management, Volume 8, pages 57–65.
2001 by Elsevier Science Ltd.
ISBN: 0-7623-0798-6

normality, the Sharpe measure does not easily allow for evaluation of its own sampling distribution. As a result, assembling any notion of the estimation risk in the Sharpe index point estimate has been difficult.

We argue in this paper that the inability to easily construct a measure of estimation risk is a weakness of the Sharpe measure, for such information is valuable to investors. To illustrate, consider an investor who uses the Sharpe measure to evaluate portfolios. Such an investor will identify two different portfolios as having very similar performance if they have similar Sharpe point estimates. However, such an appraisal does not at all consider the range of uncertainty behind these point estimates; one portfolio may have little estimation risk on its Sharpe measure while the other may have a great deal of estimation risk. Such knowledge may sway an investor to prefer the portfolio with less estimation risk. Indeed not knowing the estimation risk of the Sharpe ratio is somewhat akin to not knowing the level of significance of the well-known Jensen or single index alpha model. Investors are trained in business schools to associate a positive and significant alpha with more reliable performance than a positive and non-significant alpha, but rarely is the same case made for the Sharpe ratio because of the above-described limitation.

This limitation of the Sharpe ratio has not gone unnoticed in the literature. In fact, several previous attempts have been made to construct methods to measure the estimation risk of the Sharpe ratio. The most well known is by Jobson and Korkie (1981) who use Taylor series (delta) approximations to derive confidence intervals on the Sharpe ratio. More recently work by Vinod and Morey (2000) has documented that a bootstrap methodology can achieve narrower, more reliable confidence intervals on the Sharpe ratio than the Jobson and Korkie method.

However, both of these methodologies still possess a shortcoming in that they provide no standard procedure of trading off performance for estimation risk. For example, consider two funds, one with a high Sharpe ratio and a very wide confidence interval, and another with a lower Sharpe ratio and yet a narrower confidence interval. Which should be preferred? What weighting should the estimation risk be given?

To deal with this issue we propose a modified version of the Sharpe ratio, which we call the Double Sharpe ratio, that, rather than calculating the confidence interval width, *directly* takes into account estimation risk in much the same fashion that the original Sharpe ratio takes into account the portfolio risk. In this way, estimation risk is implicitly in the performance measure with the same weighting as standard deviation receives in the original Sharpe ratio. As such, the Double Sharpe ratio allows investors to examine the risk- and estimation risk-adjusted performance in a relatively simple fashion.

The rest of the paper is organized as follows. Section 1 discusses the Sharpe measure and its methodological problems. Section 2 explains the Double Sharpe ratio. Section 3 explains how we empirically calculate the Double Sharpe Ratio. Section 4 provides a simple empirical example to help understand the use and value of the methodology. Section 5 concludes the paper with some final remarks and more on the practical use of the methodology.

1. THE SHARPE RATIO AND ESTIMATION RISK

Consider the following scenario in which the relative performance of n portfolios is to be evaluated.[2] In this scenario, $r_{i,t}$ represents the excess return from the i-th portfolio in period t, where $i = 1, 2, \ldots, n$. A random sample of T excess returns on the n portfolios is then illustrated by $r_t' = [r_{1t}, r_{2t}, \ldots r_{nt}]$, where $t = 1, 2, \ldots, T$ and where r_t is assumed to be multivariate normal random variable, with mean $\mu = \{\mu_i\}$, $i = 1, 2, \ldots, n$ and a covariance matrix $\Sigma = (\sigma_{ij})$ where $i, j = 1, 2, \ldots n$. It is well-known that the unbiased estimators of the $(n \times 1)$ mean vector and the $(n \times n)$ covariance matrix elements are,

$$\bar{r}_{iT} = \frac{1}{T}\sum_{t=1}^{T} r_{it}, \quad \text{and} \quad S = \{s_{ij}\} = \frac{1}{T-1}\sum_{t=1}^{T} (r_{it} - \bar{r})(r_{jt} - \bar{r})'. \tag{1}$$

These two estimators are then used to form the estimators of the traditional Sharpe performance measure.

The population value of the Sharpe (1966) performance measure for portfolio i is defined as $Sh_i = \frac{\mu_i}{\sigma_i}$, $i = 1, 2, \ldots, n$. It is simply the mean excess return over the standard deviation of the excess returns for the portfolio. The conventional sample-based point estimates of the Sharpe performance measure used in (1) are then

$$\hat{S}h_i = \frac{\bar{r}_i}{s_i} \text{ for } i = 1, 2, \ldots, n. \tag{2}$$

Because of the presence of the random denominator s_i in the definition of (2), the Sharpe ratio does not permit an easy method for evaluating the estimation risk in the point estimate itself. This is because the small-sample distribution of the Sharpe measure is highly non-normal and hence the usual method for computing standard errors can be biased and unreliable.

2. THE DOUBLE SHARPE RATIO

The Sharpe ratio is defined in Eq. (2) as the ratio of the mean excess return to its standard deviation. We define the Double Sharpe ratio by

$$DSh_i = \frac{\hat{S}h_i}{s_i^{Sh}}, \tag{3}$$

where s_i^{sh} is the standard deviation of the Sharpe ratio estimate, or the estimation risk. As is clear in (3), the Double Sharpe penalizes a portfolio for higher estimation risk. One may question the weighting of the estimation risk; however, the implicit weighting of the estimation risk can be modified if desired without affecting our main proposal. For simplicity, our weighting is completely analogous to that implicit in the original Sharpe ratio.

For clarification, a couple of issues should be noted about the Double Sharpe ratio here before moving on. First, we, of course, assume the statistical paradigm is that the observed data are one realization from a random process. Hence, the observed means and variances are a random realization of the true mean and variance. Second, the Double Sharpe ratio deals only with estimation risk and not measurement risk. Estimation risk arises from the sampling variation in the estimates of mean and variance. This is distinctly different from measurement errors which arise from erroneous reports of prices, etc.

3. CALCULATING THE DOUBLE SHARPE RATIO

The question now is how to calculate the estimation risk, s_i^{sh}. To do this we utilize the well-known bootstrap methodology. To understand the basic ideas behind the bootstrap consider the following. Assume that a statistic, $\hat{\beta}$, is based on a sample of size, T. In the bootstrap methodology, instead of assuming the shape of the sampling distribution of $\hat{\beta}$, one empirically approximates the entire sampling distribution of $\hat{\beta}$ by investigating the variation of $\hat{\beta}$ over a large number of pseudo samples obtained by resampling. For the resampling, a Monte-Carlo type procedure is used on the available sample values. This is conducted by randomly drawing, with replacement, a large number of resamples of size T from the original sample. Each resample has T elements, however any given resample could have some of the original data points represented more than once and some not at all. Note that each element of the original sample has the same probability, $(1/T)$, of being in a sample.

The initial idea behind bootstrapping was that a relative frequency distribution of $\hat{\beta}$'s calculated from the resamples can be a good approximation to its sampling distribution. Hence, the underlying distribution of the statistic

can be found when there is not enough data or when the statistic does not behave in the traditional manner as in the case of the Sharpe ratio.

In this paper, the resampling for the Sharpe measure is done "with replacement" of the original excess returns themselves for $j = 1, 2, \ldots, J$ or 999 times. Thus, we calculate 999 Sharpe measures from the original excess return series. The choice of the odd number 999 is convenient, since the rank-ordered 25-th and 975-th values of estimated Sharpe ratios arranged from the smallest to the largest, yield a useful 95% confidence interval. It is from these 999 Sharpe measures that we calculate s_i^{sh}.

The choice of 999 is rather standard in the bootstrap literature,[3] however the question of whether the results are robust to different resample sizes is obviously important. The answer to this question comes from research by Davison and MacKinnon (2000). They conducted elaborate experiments and concluded that the size of the resamples should be at least 399, but recommended a size of 1000 if computational costs were not high. They found that once the resampling size was above 399, the results were robust across larger samples. We found this to be true with our study as well.

4. AN EMPIRICAL EXAMPLE

As an illustration, we have calculated the Sharpe and Double Sharpe ratios for the 30 largest growth mutual funds (as of January 1998 in terms of overall assets managed).[4] Table 1 reports the results for each fund over the time period 1987–1997 with following column headings: (i) the excess mean monthly return;[5] (ii) the standard deviation of the excess monthly returns; (iii) the Sharpe ratio; (iv and v) the mean and standard deviation of the bootstrapped Sharpe ratios; (vi and vii) the lower [0.025] and upper [0.975] confidence values of the bootstrapped Sharpe value; (viii) the 95% confidence interval width; and (ix) the Double Sharpe ratio.

The results yield several interesting findings. First, the Sharpe ratios vary widely among the 30 growth funds. They range from 0.1346 (Fidelity Trend) to almost twice that level, 0.2683 (Fidelity Contrafund).

Second, the bootstrap distribution, which incorporates random variation in estimates of both the numerator and denominator, is generally a good approximation to the true sampling distribution of the estimate. In our case, the sampling distribution representing the estimation risk is non-normal with positive skewness. This is why the means of the bootstrapped Sharpe ratios (col. iv) are always slightly higher than the point estimates of Sharpe ratios (col. iii), which ignore the estimation risk altogether. Also, the standard deviation of the bootstrapped Sharpe ratios are quite variable, suggesting that

Table 1. Characteristics of the 30 Largest Growth Funds.

Fund Name	Excess Mean Monthly Return (%)	Standard Dev. of excess returns	Sharpe Ratio	Mean of Boot-strapped Sharpe Ratios	Standard Dev. of Boot-strapped Sharpe Ratios	Lower Confidence Value (0.025) for Sharpe Ratio	Upper Confidence Value (0.975) for Sharpe Ratio	95% Confidence interval width	Double Sharpe Ratio
AIM Value A	1.1150	4.529	0.2461	0.2569	0.1037	0.0672	0.4727	0.4055	2.372
AIM Weingarten A	0.9252	4.894	0.1890	0.1959	0.0948	0.0174	0.3818	0.3644	1.995
Amcap	0.8209	4.407	0.1863	0.1904	0.0967	0.0107	0.3867	0.3759	1.927
Amer Cent-Growth	0.9158	5.849	0.1566	0.1629	0.0942	−0.0149	0.3513	0.3662	1.662
Amer Cent-Select	0.7191	4.642	0.1549	0.1607	0.0949	−0.0247	0.3529	0.3777	1.632
Brandywine	1.1890	6.244	0.1904	0.1973	0.0964	0.0179	0.3912	0.3733	1.975
Davis NY Venture	1.0990	4.313	0.2547	0.2624	0.0986	0.0779	0.4715	0.3936	2.583
Fidelity Contrafund	1.2360	4.609	0.2683	0.2919	0.1225	0.0513	0.5319	0.4806	2.190
Fidelity Destiny I	1.1490	4.746	0.2421	0.2568	0.1041	0.0640	0.4694	0.4054	2.326
Fidelity Destiny II	1.2000	4.954	0.2423	0.2524	0.1076	0.0465	0.4788	0.4323	2.253
Fidelity Growth	1.0560	5.306	0.1991	0.2074	0.1026	0.0225	0.4231	0.4007	1.940
Fidelity Magellan	0.9955	4.644	0.2144	0.2234	0.1037	0.0402	0.4479	0.4077	2.068
Fidelity OTC	0.9595	5.069	0.1893	0.2026	0.1064	0.0125	0.4313	0.4189	1.780
Fidelity Ret. Growth	0.8455	4.761	0.1776	0.1907	0.1029	−0.0006	0.4053	0.4059	1.726
Fidelity Trend	0.7113	5.286	0.1346	0.1459	0.0989	−0.0309	0.3443	0.3751	1.361
Fidelity Value	0.8130	4.209	0.1932	0.2118	0.1136	0.0098	0.4486	0.4388	1.700
IDS Growth A	1.0230	5.371	0.1905	0.2017	0.0945	0.0214	0.3818	0.3605	2.017
IDS N.Dimensions	1.0680	4.399	0.2428	0.2447	0.0944	0.0652	0.4355	0.3703	2.572
Janus	0.9459	3.822	0.2475	0.2510	0.0964	0.0803	0.4649	0.3847	2.568
Janus Twenty	1.0580	5.113	0.2069	0.2092	0.0980	0.0245	0.4149	0.3904	2.112
Legg Mas. Val. Prim	0.9320	4.629	0.2013	0.2153	0.1049	0.0207	0.4288	0.4081	1.919
Neuberg&Ber Part	0.8707	3.786	0.2300	0.2385	0.1017	0.0464	0.4505	0.4041	2.261
New Economy	0.9237	4.450	0.2076	0.2194	0.0997	0.0409	0.4199	0.3790	2.081
Nicholas	0.8691	3.821	0.2274	0.2339	0.1035	0.0435	0.4512	0.4078	2.197
PBHG Growth	1.2530	7.079	0.1770	0.1813	0.0864	0.0064	0.3502	0.3438	2.047
Prudential Equity B	0.8303	4.224	0.1966	0.2087	0.1106	0.0140	0.4246	0.4106	1.777
T. Rowe Growth	0.7829	4.324	0.1811	0.1911	0.1038	0.0013	0.4032	0.4020	1.745
Van Kampen Pace	0.7598	4.519	0.1681	0.1885	0.1003	0.0004	0.4048	0.4044	1.627
Vanguard U.S. Growth	0.8838	4.459	0.1982	0.2017	0.0987	0.0125	0.4107	0.3982	2.009
Vanguard/Prime	0.9986	5.156	0.1937	0.2089	0.0967	0.0282	0.3956	0.3674	2.003

the estimation risk is variable. The lowest is PHBG with a value of 0.0864, the highest, more than 40% higher than PBHG, is the Fidelity Contrafund with a value of 0.1225.

Third, four of the 30 funds have lower bound confidence values that are negative. This indicates that these funds have Sharpe ratios which are not significantly different from 0 (at the 5% level) when estimation risk is considered. Moreover, the results on the confidence interval width show that it is not entirely intuitive which funds will have narrower confidence intervals; the fund with the highest standard deviation of returns actually has the narrowest confidence interval (PBHG Growth).

Fourth, the Double Sharpe ratios provide a different set of ranking of the funds. Indeed, the Spearman rho rank correlation test of Sharpe ratio rankings to the Double Sharpe rankings is 0.197 and the test does not reject the null hypothesis that the correlation of the two sets of rankings is 0. *Hence, it is not the case that funds with high Sharpe ratios will necessarily have low estimation risk*. Such results indicate that an investor who uses the Sharpe ratio should use the Double Sharpe ratio also. Only the latter incorporates the estimation risk. If the fund scores well in its Sharpe ratio and Double Sharpe ratio rankings, the investor can be assured that the fund has performed relatively well and is not subject to high estimation risk. Again, the weighting of the estimation risk in the Double Sharpe could be questioned. Since the investor is using the implicit weighting in the Sharpe ratio anyway, we start by weighting the estimation risk in a similar fashion. More sophisticated investors could of course adjust the weighting to reflect their individual preferences for representing the importance of estimation risk.

5. CONCLUSIONS

It is often the case that financial researchers assume that investors know the true mean, variance, covariance, etc. of stock returns. But in reality, investors rarely have this information. For example, to apply the elegant framework of modern portfolio theory, investors must estimate expected stock returns using whatever information is currently available. However, since there is often much random noise in financial markets, the observational statistics used by investors need not coincide with true parameters. This is increasingly being referred to as *estimation risk*.

The notion of estimation risk is becoming more and more of an issue in finance. In a recent paper, Lewellen and Shanken (2000) document that not accounting for estimation risk can lead to different empirical findings in regards to stock market return predictability and market efficiency. In this paper, we showcase how one of the most widely taught and used performance metrics, the Sharpe ratio, does not easily account for this estimation risk. This limitation of the Sharpe ratio is often ignored in spite of the fact that the Sharpe ratio is regularly taught in investments classes right along side with the Jensen and 4-index alpha models, which of course examine the underlying estimation error by also investigating the significance of the alpha.

To deal with this issue we propose a modified version of the Sharpe ratio, which we call the Double Sharpe ratio. Rather than calculating the confidence interval width, it *directly* takes into account estimation risk in much the same fashion that the original Sharpe ratio takes into account the portfolio risk. In this way, estimation risk is implicitly in the performance measure with the same weighting as standard deviation receives in the original Sharpe ratio. As such, the Double Sharpe ratio allows investors to examine the risk- and estimation risk-adjusted performance in a relatively simple fashion.

Since computational abilities have improved dramatically with the advent of more powerful personal computers, calculating the Double Sharpe ratio is not at all out the realm of most fund managers. The calculation process involves using a simple bootstrap methodology that can be easily programmed in most statistical packages. Indeed, there are many free downloadable bootstrap codes potentially useful for calculating the Double Sharpe ratios. Software packages including S-Plus, SAS, Eviews, Limdep, TSP and GAUSS have public websites, which can be searched for bootstrap options. See Vinod (2000) for references to software web sites and comparisons of numerical accuracy. By using the Double Sharpe ratio in conjunction with the original Sharpe ratio, the fund manager/investor can understand the performance of the portfolio well, in terms of a performance metric that most investors know, as well as getting a grasp of the estimation risk of the portfolio.

We do not claim that the Double Sharpe ratio is always better than others, or that it cannot be further improved. However, we do claim to provide more information regarding the estimation risk to those who still use the Sharpe ratio. Admittedly, the Double Sharpe ratio, since it is based on the Sharpe ratio suffers from the known limitations of the Sharpe ratio, arising from its arbitrary weighting of risk and ignoring negative returns, skewness and kurtosis.[6] However, Sharpe ratio remains one of the most-widely used and taught performance metrics in the finance literature. No one suggests exclusive reliance on the Sharpe ratio. In the same spirit, Double Sharpe ratio is intended to provide a relatively simple and straightforward method of dealing with the estimation risk, to be used in conjunction with other measures.

NOTES

1. According to the EconLit Database, April 1999.
2. The notation in this section is the same as Jobson and Korkie (1981).
3. See Davison and MacKinnon (2000) and Davison and Hinkley (1997).
4. These funds are classified as "Growth" by Morningstar Inc.
5. We subtract the monthly one-month U.S. T-Bill rates from the monthly fund returns to calculate the excess returns.
6. Note that our method also suffers from the same problems as the Sharpe ratio in that there may be some error in calculating the estimation risk due to non-synchronous trading. Kryzanowski and Sim (1990) address ways to correct for such problems.

REFERENCES

Davison, R., & MacKinnon, J. (2000). Bootstrap Tests: How Many Bootstraps. *Econometric Reviews*, *19*, forthcoming.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. New York: Cambridge University Press.

Jobson, J. D., & Korkie, B. M. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *Journal of Finance*, *36*(4), 889–908.

Kryzanowski, L., & Sim, A. B. (1990). Hypothesis Testing with the Sharpe and Treynor Portfolio: Performance Measures Given Non-Synchronous Trading. *Economics Letters*, *32*, 345–352.

Lewellen, J., & Shanken, J. (2000). Estimation Risk, Market Efficiency and the Predictability of Returns. National Bureau of Economic Research, Working Paper No. 7699, May.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, *39*(1), Part 2, January, 119–138.

Vinod, H. D., & Morey, M. R. (2000). Confidence Intervals and Hypothesis Testing for the Sharpe and Treynor Performance Measures: A Bootstrap Approach. *Computational Finance 99*. MIT Press, Chapter 3, 25–39.

Vinod, H. D. (2000). Review of GAUSS for Windows, Including its Numerical Accuracy. *Journal of Applied Econometrics*, *15*, 211–220.