

# Bias and consistency of the maximum Sharpe ratio

## Ross A. Maller

Centre for Financial Mathematics and School of Finance and Applied Statistics, Australian National University, ACT 0200 Australia.

## Robert B. Durand

Department of Accounting and Finance, University of Western Australia, Nedlands, Western Australia.

## Peter T. Lee

Department of Accounting and Finance, University of Western Australia, Nedlands, Western Australia.

We show that the maximum Sharpe ratio obtained via the Markowitz optimization procedure from a sample of returns on a number of risky assets is, under commonly satisfied assumptions, biased upwards for the population value. Thus investment advice, decisions and assessments based on the estimated Sharpe ratio will be overly optimistic. The bias in the estimator is shown theoretically and illustrated using a data set of Spiders and iShares. We obtain bounds on the difference between the sample maximum Sharpe ratio and its population counterpart and show that the sample estimator is consistent for the population value; thus the bias disappears asymptotically under some reasonable assumptions. However, the bias can be significant in finite samples and can persist even in very large samples. We demonstrate this with simulations based on portfolios formed from normally and  $t$ -distributed returns. As expected, the over-optimistic risk–return tradeoff predicted by the procedure is not reflected in corresponding good out-of-sample portfolio performance of the Spiders and iShares.

## 1 Introduction

The Sharpe ratio (Sharpe, 1966; 1994) is a basic tool in finance, especially for performance evaluation. Its use as a summary measure of the (excess) return achieved by a portfolio, as compared with the risk taken, is well established among theorists and practitioners (see, eg, Wu and Xiao, 2002, for a recent review and Emmer *et al.*, 2001, for typical applications). The maximum Sharpe ratio plays an even more fundamental role in the portfolio optimization problem, in which it can be used to locate an optimal point on the efficient frontier – the tangent point – which

---

This work was partially supported by ARC grant DP210572. We thank Professors Philip Brown, Claudia Klüppelberg, Robert Whitelaw and Kent Zumwalt for helpful comments on this work at various stages of its development. We are especially grateful for constructive criticism from a referee and from the Editor, which contributed greatly to our improving the paper. Thanks also to Ken Phillips for his programming help at an early stage.

determines the constitution of that portfolio of risky assets which, in equilibrium, will be held in some proportion by any investor. This mutual fund or separation theorem (Tobin, 1958) is a consequence of the seminal work of Markowitz (Markowitz, 1952), and, ultimately, provides the justification for the capital asset pricing model (CAPM) (Sharpe, 1964).

These aspects of the Sharpe ratio are extensively used in empirical studies for the out-of-sample evaluation of portfolio performance. In this regard it is now well recognized that a portfolio of risky assets determined by Markowitzian allocation on historical data often performs rather poorly when used out-of-sample. There are various reasons for this, and here we add to them a fundamental consideration, by showing that the maximum Sharpe ratio obtained from data – a sample of asset returns – will commonly be biased upwards, as an estimator of the true, unknown, population version. Thus the Markowitz allocation is over-optimistic in the sense that the sample maximum Sharpe ratio overestimates the true maximum Sharpe ratio. We will demonstrate theoretically that this bias disappears in large samples – or, at least, the sample Sharpe ratio is consistent for its population value – if consistent estimators are used (as is usual) for sample mean excess returns and their covariance matrix. However, the bias may be large and may persist even in samples of size of the order of 1,500 days, as we will show by simulations based on the returns from some exchange traded shares. So the consistency property may provide small comfort in practical situations. Our technology in this regard is related to, though with a different focus from, the work of Jorion (1992) on the simulation of portfolio allocations by the Markowitz method.

The paper is set out as follows. In Section 2, we outline the portfolio optimization procedure, prove, under some mild assumptions, that the sample maximum Sharpe ratio is biased (Proposition 1), find bounds for the bias (Proposition 2), and establish the consistency result for the maximum Sharpe ratio (Corollary to Proposition 2). Section 3 reports on the simulations. In a discussion, Section 4, we mention some related results. Section 5 concludes. Proofs and technicalities are in the Appendices.

## 2 Analysis

Given a population of  $N$  risky assets with excess return vector  $\mu$  (an  $N$ -vector) and covariance matrix  $\Sigma$  (an  $N \times N$  positive definite matrix), we form portfolios of the assets with allocations as specified by an  $N$ -vector  $x$ . The maximum Sharpe ratio achievable in the population is the maximized value of the function

$$f(x) = \frac{x'\mu}{\sqrt{x'\Sigma x}} \quad (1)$$

obtained over variations in  $x$ . Here  $x'\mu$  is the expected excess return from a portfolio consisting of the  $N$  assets combined according to the components of the vector  $x$ , with  $\sqrt{x'\Sigma x}$  being the standard deviation of that return. The ratio of these two quantities is the Sharpe ratio of that portfolio – but then we vary  $x$  over a range of possibilities to obtain the portfolio with optimal (maximum) Sharpe

ratio. A usual requirement (to which we shall adhere) is that the components of  $x$  sum to 1 (the total allocation constraint). The maximization of (1) is carried out under the possible extra constraint that  $x$  is also to lie in a specified subset  $C$  of  $N$ -dimensional space. In practice,  $C$  reflects restrictions placed by the investor on the portfolio, such as disallowing short sales (all components of  $x$  non-negative), or, more generally, restricting above and/or below the amounts of certain assets allowed in the portfolio. But in general, the set  $C$  need not be compact or even bounded, in fact it may be the whole of  $N$ -dimensional space, in which case there is no constraint at all on the portfolio allocations (other than the total allocation constraint). This is the situation, in particular, when short sales are allowed.

Denote the maximum value of  $f(x)$  thus obtained by  $SR$ , the population maximum Sharpe ratio.  $SR$  is no longer of the form of a mean return divided by its standard deviation; in general, there is no explicit expression available for it.

Suppose  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  are estimates of  $\mu$  and  $\Sigma$  based on a sample of size  $T$ . The corresponding sample maximum Sharpe ratio is the maximized value, for  $x$  in  $C$  and satisfying the total allocation constraint, of the function

$$\hat{f}_T(x) = \frac{x' \hat{\mu}_T}{\sqrt{x' \hat{\Sigma}_T x}} \tag{2}$$

Call this maximum value  $\widehat{SR}_T$ , and the portfolio allocation which makes it up,  $\hat{x}_T$ . Note that  $\hat{x}_T$  and  $\widehat{SR}_T$  depend on the sample estimates  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  and, hence, have to be treated as random. Practical investment decisions regarding the composition of the optimal portfolio, as well as further theoretical developments, are often based either explicitly or implicitly on  $\widehat{SR}_T$  and  $\hat{x}_T$ . Given the nature of the estimate  $\widehat{SR}_T$  it is not surprising that it is biased, even when  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  are unbiased estimates of  $\mu$  and  $\Sigma$ . What we will show in fact is that  $\widehat{SR}_T$  is commonly biased upwards for  $SR$ , which is the worst (non-conservative) position from a practical point of view.

**PROPOSITION 1** *Suppose that  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  are independent, unbiased estimators of  $\mu$  and  $\Sigma$ , based on a sample of size  $T \geq 2$ , and that  $SR > 0$ . Then, for any collection  $C$  of  $N$ -vectors,  $E(\widehat{SR}_T) > SR$ .*

In Proposition 1, “ $E$ ” denotes expectation. The proof of the proposition is in Appendix A. The case  $SR \leq 0$ , which we exclude in Proposition 1, is of little interest in practice, since it implies no, or a negative, risk premium for the risky assets, at least for portfolios in  $C$ ; if it appeared to be the case in a practical situation, allocation would not proceed by this method anyway.

The assumption of independence of sample mean and covariance matrix in Proposition 1 is of course satisfied if the excess returns are normally distributed. To assess the magnitude of the bias, we carried out some simulations using normally distributed returns, with “true” parameters based on data from iShares and Spiders. It is, however, well accepted in practice that returns tend to be non-normal, with skewed, heavy-tailed leptokurtic distributions. So we also carried out some simulations using a  $t$ -distribution with four degrees of freedom. The simulations are reported in the next section.

To complete this section, we observe that, while Proposition 1 demonstrates the upward bias in the sample estimator of the maximum Sharpe ratio, it gives no idea of the magnitude of the bias. In view of the complicated nature of the estimator it seems difficult to give explicit estimates of this. We can, however, obtain bounds on the absolute difference between the sample statistic and its population counterpart. These can be used to assess the size of the bias in practice, and we also use them to show that the estimator is consistent, provided that it is based on consistent underlying estimators of the mean vector and covariance matrix.

*Proposition 2 Under the conditions of Proposition 1, the following bound holds for any sample values  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$ , and any collection  $C$  of  $N$ -vectors:*

$$\left| \widehat{SR}_T - SR \right| \leq \sqrt{(\hat{\mu}_T - \mu)' \hat{\Sigma}_T^{-1} (\hat{\mu}_T - \mu)} + \sqrt{\mu' \Sigma^{-1} \mu} \left\| \Sigma^{\frac{1}{2}} \hat{\Sigma}_T^{-1} \Sigma^{\frac{1}{2}} - I \right\| \quad (3)$$

In Equation (3),  $\|A\|$  denotes the largest in absolute value of the eigenvalues of a symmetric matrix  $A$ , and  $\Sigma^{1/2}$  is the symmetric square root of  $\Sigma$ . See Appendix B for the proof of Proposition 2.

With some further assumptions and analysis, Proposition 2 enables us to find easily calculable upper bounds for the magnitude of the bias in  $\widehat{SR}_T$ . Suppose the returns are normally distributed, with  $\hat{\mu}_T$  independent of  $\hat{\Sigma}_T$ . Using the Cauchy–Schwarz inequality, the expectation of the first term on the right-hand side of Equation (3) is bounded above by

$$\sqrt{E\left((\hat{\mu}_T - \mu)' \hat{\Sigma}_T^{-1} (\hat{\mu}_T - \mu)\right)} = \sqrt{\frac{(T-1)N}{T(T-N)}} \sim \sqrt{\frac{N}{T}}, \quad \text{for large } T \quad (4)$$

This follows because the expectation in (4) is of  $T$  times a Hotelling’s  $T^2$  random variable, under the normality and independence assumptions. A slightly more involved calculation, omitted here, shows that the median value of the random variable on which the second term on the right-hand side of Equation (3) is based, is bounded above for large  $T$  by the quantity

$$\sqrt{\frac{\mu' \Sigma^{-1} \mu}{T}} \quad (5)$$

This bound depends on the unknown parameters  $\mu$  and  $\Sigma$ . In the data analyzed in the next section, we estimate the values of (4) and (5), and it turns out that (4) gives by far the dominant contribution. We suggest that this will usually be the case, and further, that (4) provides a conservative over-estimate of the magnitude of the bias in  $\widehat{SR}_T$ .

We conclude this section with the following consistency result, which follows from Proposition 2. We remark that the consistency of  $\widehat{SR}_T$  does not follow directly from (1), the consistency of  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$ , and the fact that the supremum functional, when taken over a compact set, is continuous – because in general we

are required to maximize (1) over an unbounded set. This is the case for example when short sales are allowed and the only constraint on the portfolio allocations is that they sum to 1. Indeed, there are cases where the supremum of (1) occurs at infinite values of  $x$ . Nevertheless, due to the “self-normalized” nature of the Sharpe ratio, it remains bounded for all  $x$  and the next Corollary, as mentioned, can be deduced from (3).

**COROLLARY TO PROPOSITION 2** *Suppose  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  are consistent estimators of  $\mu$  and  $\Sigma$ , based on a sample of size  $T$  – that is, as  $T \rightarrow \infty$  they converge (in probability, or almost surely) to their population values. Then, for any collection  $C$  of  $N$ -vectors,  $\hat{SR}_T$  is consistent for  $SR$  as  $T \rightarrow \infty$ .*

### 3 Simulations

To provide empirical support for the two propositions, we consider an international portfolio optimization situation utilizing data on assets readily available to investors. This approach both addresses an allocation problem that is of interest to investors and provides a tractable number of assets for our simulations.

The simulations are based on statistics calculated from the daily continuous excess returns of “Spiders”, which are index shares designed to replicate the S&P 500 index, and the 17 “iShares”, which are assets designed to track Morgan Stanley Capital International country indices. Information is available on these for the period beginning April 1, 1996 and ending March 31, 2002. These  $N = 18$  assets are directly tradable on the American Stock Exchange whereas the share indices they mimic normally cannot be traded. A good description of Spiders may be found in Elton *et al* (2002), and Olienyk *et al* (1999) provide background information concerning iShares. We do not repeat their details here, other than to mention that Spiders and iShares are traded in the same time zone, and thus we can validly calculate correlations between their returns without introducing bias due to cross-country lead/lag relationships in the underlying indices. Such properties mean that iShares are attractive to international finance researchers, and they have recently been used in studies of international portfolio diversification (eg, Olienyk *et al*, 2001). The US Treasury cash middle rate is used as our risk-free rate. All of our data were obtained from Datastream.

We considered the data on a daily basis, quarter by quarter over the period, for a total of 24 quarters. For each quarter, the sample means and covariance matrices of the daily continuously compounded returns among the 18 assets were used to find that portfolio with the highest Sharpe ratio, obtained by maximizing the function in (2). Returns included any dividends received during the period. Short-sales were prohibited, so a quadratic optimization was performed in each quarter.<sup>1</sup> For simulation purposes, the resulting Sharpe ratios and corresponding portfolio

<sup>1</sup> iShares and Spiders permit short selling, so we could also realistically have considered the (easier) optimization problem with short sales allowed. In practice, however, many funds prohibit short selling so we decided to focus on this situation.

**TABLE I** Comparison of “true” and simulated maximum Sharpe ratios – normal distribution

	Sharpe ratio of true maximum Sharpe ratio portfolio	Mean of 10,000 simulated max Sharpe ratios	Median of 10,000 simulated max Sharpe ratios	Std dev of 10,000 simulated max Sharpe ratios
<b>A: Quarterly formation periods – daily returns</b>				
Q2-1996	0.2213	0.3695	0.3631	0.1253
Q3-1996	0.0994	0.2242	0.2194	0.1133
Q4-1996	0.3104	0.4576	0.4520	0.1286
Q1-1997	0.0919	0.2621	0.2567	0.1126
Q2-1997	0.4669	0.6169	0.6103	0.1409
Q3-1997	0.1193	0.2474	0.2434	0.1167
Q4-1997	0.0537	0.1312	0.1276	0.1139
Q1-1998	0.4464	0.5186	0.5125	0.1363
Q2-1998	0.1785	0.2491	0.2452	0.1217
Q3-1998	-0.1401	-0.0105	-0.0037	0.1198
Q4-1998	0.3341	0.4495	0.4456	0.1294
Q1-1999	0.1624	0.2605	0.2552	0.1227
Q2-1999	0.3420	0.4592	0.4533	0.1301
Q3-1999	0.0950	0.1648	0.1605	0.1128
Q4-1999	0.3853	0.5202	0.5141	0.1321
Q1-2000	0.0897	0.2000	0.1959	0.1138
Q2-2000	0.0655	0.2188	0.2144	0.1142
Q3-2000	-0.0715	0.0907	0.0853	0.1018
Q4-2000	0.1005	0.2138	0.2087	0.1120
Q1-2001	0.0436	0.1374	0.1326	0.1085
Q2-2001	0.1918	0.2883	0.2824	0.1197
Q3-2001	0.0032	0.0975	0.0937	0.1071
Q4-2001	0.3028	0.4268	0.4208	0.1288
Q1-2002	0.2816	0.3864	0.3817	0.1273
Total observations	24	24	24	24
Average over 24 quarters	0.1739	0.2908	0.2863	0.1204

allocations obtained at this stage were designated as the “true” Sharpe ratios and allocations, respectively. Each of these quantities is based on approximately 60 observations (daily returns over a quarter).

In the simulations, 10,000 random realizations of the return vectors were generated for each quarter. The parameters (means and covariance matrices) of the generated returns were set equal to the observed means and variances for that quarter. It was assumed, for a first set of simulations, that returns follow a Gaussian distribution and, for a second set, that they follow a  $t$ -distribution with four degrees of freedom. Taking two distributions in this way means that we encompass the usual “baseline” multivariate normal setup and also allows a comparison with

**TABLE I** (continued) Comparison of “true” and simulated maximum Sharpe ratios – normal distribution

	Sharpe ratio of true maximum Sharpe ratio portfolio	Mean of 10,000 simulated max Sharpe ratios	Median of 10,000 simulated max Sharpe ratios	Std dev of 10,000 simulated max Sharpe ratios
<b>B: Annual formation periods – daily returns</b>				
1996*	0.1410	0.2111	0.2091	0.0680
1997	0.0883	0.1375	0.1372	0.0575
1998	0.0669	0.1112	0.1102	0.0580
1999	0.1491	0.1987	0.1979	0.0600
2000	-0.0062	0.0500	0.0487	0.0516
2001	0.0142	0.0679	0.0660	0.0515
Average over 6 years	0.0755	0.1294	0.1282	0.0578
<b>C: Entire period – daily returns</b>				
Q2-1996 to Q1-2001	0.0165	0.0344	0.0344	0.0242
<b>D: Entire period – monthly returns</b>				
April 1996 to March 2002	0.1447	0.1868	0.1833	0.1139
<b>E: Entire period – quarterly returns</b>				
Q2-1996 to Q1-2001	0.2269	0.2998	0.2878	0.2170

\*Consists of only 3 quarters of data from the 1996 period.

results obtained when returns follow a very heavy-tailed distribution.

From each of the 10,000 realizations, in each quarter and for each assumed distribution, the portfolio with the greatest Sharpe ratio was determined, and these 10,000 maximum Sharpe ratios and the accompanying allocations formed the basis of the analysis. In contrast to Jorion (1992), we evaluated the Sharpe ratios of the generated portfolios with respect to the generated (random) parameters, whereas Jorion evaluated the Sharpe ratios of the generated portfolios with respect to the original (true) parameters (with respect to which they are always suboptimal, of course). This distinction is because we wish to assess bias in estimation of the maximum Sharpe ratios, rather than the extent to which a particular portfolio (eg, the world index, or a US bond index) is suboptimal by comparison with the portfolio of maximum Sharpe ratio.

Following this extensive simulation of quarterly data, we also grouped the daily returns annually for portfolio formation, and, in a coarser categorization, took the entire set of daily data as a unit for portfolio formation. This allowed us to examine the effect of increasing sample size from approximately 60 to 250, then

to 1,500. Then, further, we repeated the entire-period analyses using monthly and quarterly rather than daily returns; the monthly and quarterly returns are progressively less variable than the daily, but there are (progressively) fewer of them.

To demonstrate the extent of the bias, Table 1 shows the value of the true maximum Sharpe ratio compared with the mean and median of the generated maximum Sharpe ratios, for each of the sub-periods (quarterly, annual, overall). In Table 1, panel A reports quarterly simulations calculated using daily data for portfolio formation, panel B gives annual simulations using the daily data (although, for 1996, we have only three quarters of data) and panel C reports the simulation for the entire sample period, again using the daily data for portfolio formation. Panels D and E show the results for the entire sample period, using monthly and quarterly returns for portfolio formation.

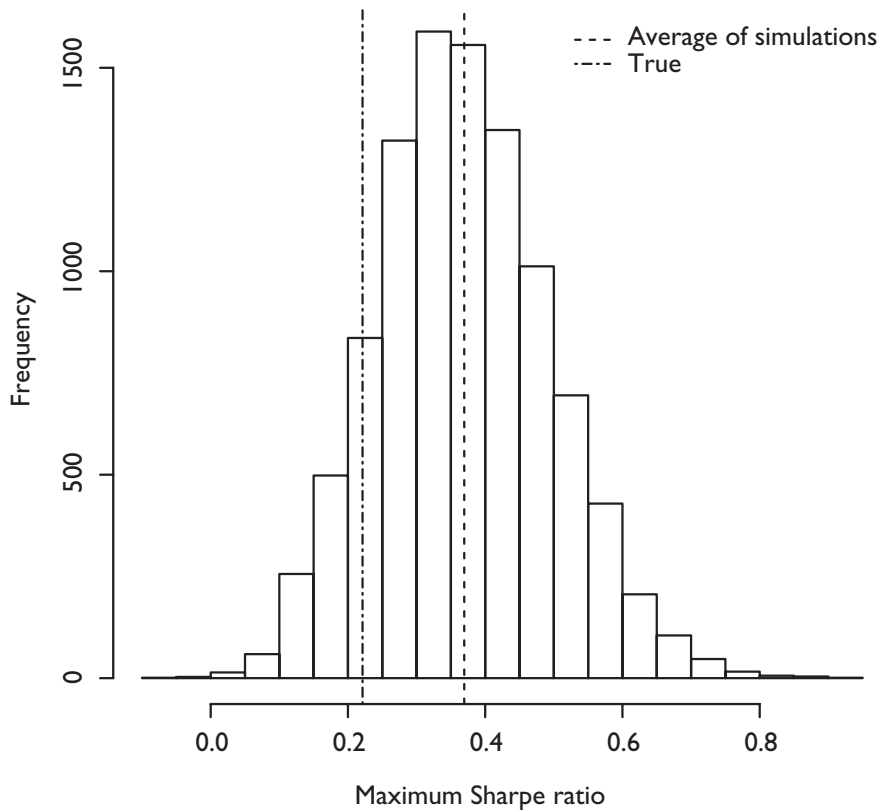
The simulations clearly illustrate the effect predicted by Proposition 1. In no instance is the true Sharpe ratio greater than the average Sharpe ratio. Some skewness can be detected in the distributions of the estimated Sharpe ratios, but there is no instance of a true maximum Sharpe ratio exceeding the median of the 10,000 simulations. To illustrate, we reproduce the histogram (Figure 1) of simulated Sharpe ratios from the first quarter of the study (Quarter 2, 1996) (the other quarters gave qualitatively very similar results):

Table 1 also provides evidence consistent with the Corollary to Proposition 2. As we increase the sample size from quarterly to annual and then to the entire study period, we find that the bias falls. The average difference between the true and mean Sharpe ratios obtained from quarterly simulations based on portfolios formed from daily returns is 0.1169, whereas it is 0.0539 when obtained from annual simulations and 0.0179 from simulations over the entire period. Nevertheless, these differences are large and remain significant by comparison with the standard deviations of the simulated maximum Sharpe ratios. The comparable figures when portfolios were formed from monthly or quarterly returns were 0.0421 and 0.0729 (for simulations over the entire period).

The upper bound in Equation (4) for the expected bias takes the value  $\sqrt{18/1500} = 0.11$  when evaluated relevant to the “entire period” analysis reported in Table 1. This is very conservative by comparison with the bias of  $0.0344 - 0.0165 = 0.0179$  estimated from the simulations. The additional bound in Equation (5) takes the value 0.0004 in this situation, so (4) is by far the dominant term. We expect this to be the case fairly generally, as Equation (5) is of the order of magnitude of the Sharpe ratio itself, divided by  $\sqrt{T}$ , whereas Equation (4) is of the order of the square root of  $N$ , the number of stocks in the portfolio, divided by  $\sqrt{T}$ , a much larger quantity in general.

Simulations based on a  $t$ -distribution with four degrees of freedom produced results consistent with those reported in Table 1. In each instance, the mean and median simulated maximum Sharpe ratios were greater than the true maximum Sharpe ratios. Consistent with the Corollary to Proposition 2, the average difference between the true and mean Sharpe ratios decreased from 0.1146 using quarterly simulations, to 0.0649 for annual simulations and to 0.0189 for the entire

**FIGURE 1** Histogram of 10,000 maximum Sharpe ratios simulated under normally distributed returns.



period (results based on daily returns). Since these results were qualitatively very similar to those for the normal distribution, we omit further details of them here (results are available on request from the second author).

We conclude this section with some related comments. In general, the sample portfolio allocations (the  $\hat{x}_T$ , in our notation) are biased estimators of the true optimal allocations with high variability. This was apparent in our simulations both for the normal and the  $t$ -distributions, and is consistent with literature demonstrating that portfolio composition is highly sensitive to the inputs (Black and Litterman, 1992), so we omit further discussion.

Finally, it is natural to ask about the out-of-sample performances of the portfolio generating the true maximum Sharpe ratio by comparison with those of the simulated portfolios. We carried out the following analyses. For each simulated optimal portfolio formed in quarter  $t$  (as determined by its allocations  $\hat{x}_T$ ), we examined its performance in quarter  $t+1$  (and similarly, for the annual simulations, for each portfolio formed in year  $t$ , we examined its performance in year  $t+1$ ) by calculating average returns and their standard deviations over the 10,000

simulations. Again we will omit details, but it was clear that the performances of neither the “true”, nor other portfolios selected according to various criteria, were significantly different from the average performance of all the random portfolios.

#### 4 Discussion

Our results accord with, and extend, those of Miller and Gehr (1978) and Jobson and Korkie (1981), who investigated the Sharpe ratio of an individual asset or portfolio of assets; in other words, the ratio of a single mean excess return (asset or portfolio) to its standard deviation. (Thus, no optimization was involved in their estimations.) They also demonstrated a theoretical upward bias in this measure, when estimated as the ratio of the usual unbiased sample mean and standard deviation estimators, compared with its true value. The Sharpe ratio of an individual portfolio is a much easier statistic to analyze, however, than the maximum Sharpe ratio achievable from a portfolio of assets combined in an optimal manner, which is the object of our attention, because, in general, there is no explicit expression for the latter, which has to be obtained by some numerical procedure. Nevertheless, as we showed in Proposition 1 and the simulations, it is possible to prove the existence of the bias both theoretically and in a practical example.

#### 5 Conclusion

The paper has presented two propositions relating to the bias and consistency of the maximum Sharpe ratio. Proposition 1 implies that a sample Sharpe ratio can be expected to be greater than the true Sharpe ratio, on average. The Corollary to Proposition 2 suggests that this bias disappears asymptotically. Empirical evidence, using simulations based on Spiders and iShares, instruments that may be readily incorporated in an internationally diversified portfolio, shows, however, that bias persists in very large, finite, samples.

Somewhat surprisingly, our results for the  $t$ -distribution did not differ greatly from those for the normal. This may be attributed perhaps to the “self-normalizing” property of the Sharpe ratio (familiar already in regard to the usual  $t$ -statistic), whereby extreme fluctuations in the sample tend to cancel as a result of dividing sample mean by sample standard deviation. This robustness property of the portfolio allocation procedure deserves further exploration.

We have noted that the bias in the maximum Sharpe ratio has important ramifications both for investors and for financial econometricians. Correcting the bias is important, but it requires further research.

### Appendix A

Let  $C$  be any subset of  $\mathbb{R}^N$ . Then argue as follows:

$$\begin{aligned} E\left(\widehat{SR}_T\right) &= E \sup_{x \in C} \left( \frac{x' \hat{\mu}_T}{\sqrt{x' \hat{\Sigma}_T x}} \right) \geq \sup_{x \in C} E \left( \frac{x' \hat{\mu}_T}{\sqrt{x' \hat{\Sigma}_T x}} \right) \\ &= \sup_{x \in C} E(x' \hat{\mu}_T) E \left( \frac{1}{\sqrt{x' \hat{\Sigma}_T x}} \right) \quad (\text{by independence}) \\ &= \sup_{x \in C} E \left( \frac{x' \mu}{\sqrt{x' \hat{\Sigma}_T x}} \right) \quad (\text{since } \hat{\mu}_T \text{ is unbiased for } \mu) \end{aligned}$$

Since we assumed  $SR > 0$ , there is an  $x$  in  $C$  for which  $x' \mu > 0$ , and  $C' = C \cap \{x : x' \mu > 0\}$  is non-empty. Then

$$\begin{aligned} E\left(\widehat{SR}_T\right) &\geq \sup_{x \in C'} \left( \frac{x' \mu}{E \sqrt{x' \hat{\Sigma}_T x}} \right) \quad (\text{by Jensen's inequality}) \\ &\geq \sup_{x \in C'} E \left( \frac{x' \mu}{\sqrt{x' \Sigma x}} \right) \quad (\text{Jensen's inequality again}) \\ &= SR \end{aligned}$$

Jensen's inequality for convex/concave functions was used twice in the above chain of inequalities: to deduce  $E(1/X) \geq 1/EX$ , and to deduce  $E(\sqrt{X}) \leq \sqrt{EX}$ , for a positive random variable  $X$ . The inequalities are strict as long as  $\hat{\mu}_T$  and  $\hat{\Sigma}_T$  are not degenerate at the constant values  $\mu$  and  $\Sigma$ , as we implicitly assume. This proves Proposition 1.

### Appendix B

Let  $x \in \mathbb{R}^N$  and write

$$\begin{aligned} \frac{x' \hat{\mu}_T}{\sqrt{x' \hat{\Sigma}_T x}} &= \frac{x' \mu}{\sqrt{x' \hat{\Sigma}_T x}} + \frac{x' (\hat{\mu}_T - \mu)}{\sqrt{x' \hat{\Sigma}_T x}} \\ &= \frac{x' \mu}{\sqrt{x' \Sigma x}} + \frac{x' \mu (\sqrt{x' \Sigma x} - \sqrt{x' \hat{\Sigma}_T x})}{\sqrt{x' \Sigma x} \sqrt{x' \hat{\Sigma}_T x}} + \frac{x' (\hat{\mu}_T - \mu)}{\sqrt{x' \hat{\Sigma}_T x}} \\ &= \frac{x' \mu}{\sqrt{x' \Sigma x}} + A_T(x) + B_T(x), \text{ say} \end{aligned} \tag{A1}$$

Now, for each  $x \in \mathbb{R}^N$ ,

$$A_T(x) = \frac{(x'\mu)x'(\Sigma - \hat{\Sigma}_T)x}{\sqrt{x'\Sigma x}\sqrt{x'\hat{\Sigma}_T x}\left(\sqrt{x'\Sigma x} + \sqrt{x'\hat{\Sigma}_T x}\right)} \tag{A2}$$

Some routine calculations show that

$$A_T^2(x) = \frac{\left(u'\Sigma^{-\frac{1}{2}}\mu\right)^2\left(u'(I - D_T)u\right)^2}{u'D_T u\left(1 + u'D_T u + 2\sqrt{u'D_T u}\right)} \tag{A3}$$

where  $u = \Sigma^{1/2}x/\sqrt{x'\Sigma x}$  is a unit vector, and  $D_T = \Sigma^{-1/2}\hat{\Sigma}_T\Sigma^{-1/2}$ . Since  $\sup_{|u|=1}(u'\Sigma^{-1/2}\mu)^2 = \mu'\Sigma^{-1}\mu$ , and

$$\sup_{|u|=1} \frac{\left(u'(I - D_T)u\right)^2}{u'D_T u} = \sup_{|z|=1} \left(z'(D_T^{-1} - I)z\right)^2 = \|D_T^{-1} - I\|^2$$

we see that  $A_T(x)$  is no larger in modulus than

$$\sqrt{\mu'\Sigma^{-1}\mu} \left\| \Sigma^{\frac{1}{2}}\hat{\Sigma}_T^{-1}\Sigma^{\frac{1}{2}} - I \right\| = A_T, \text{ say} \tag{A4}$$

uniformly in  $x \in \mathbb{R}^N$ . Similarly,  $B_T(x)$  is no larger in modulus than

$$\sqrt{(\hat{\mu}_T - \mu)'\hat{\Sigma}_T^{-1}(\hat{\mu}_T - \mu)}$$

uniformly in  $x \in \mathbb{R}^N$ . Then (3) follows from

$$\left| \widehat{SR}_T - SR \right| \leq A_T + B_T$$

REMARKS: (i) From (3),  $\widehat{SR}_T - SR \leq A_T + B_T$ , which together with Proposition 1 gives

$$0 \leq E\left(\widehat{SR}_T\right) - SR \leq E\left(A_T\right) + E\left(B_T\right) \tag{A5}$$

provided that the expectations on the right exist. In principle, (A5) can be used to give bounds on the bias of  $\widehat{SR}_T$ . We did this in Equation (4) for  $E(B_T)$ , but  $E(A_T)$  seems difficult to evaluate except in simple cases, so we gave the bound (5) for the median of  $A_T$ .

(ii) In going from (A3) to (A4) we neglected the term in brackets in the denominator of (A3), which, since  $D_T \xrightarrow{P} I$ , is of the order of 4 in large samples. Thus (taking the square root) we can expect our overestimate to be too big perhaps by a factor of 2, from this source at least.

(iii) If the consistency assumption is strengthened to  $\hat{\mu}_T = \mu + O_p(1/\sqrt{T})$  and  $\hat{\Sigma}_T = \Sigma + O_p(1/\sqrt{T})$ , as commonly hold, then we can deduce from (3) that  $\sqrt{T}(\widehat{SR}_T - SR) = O_p(1)$ , as  $T \rightarrow \infty$ , giving a  $\sqrt{T}$ -rate of convergence of  $\widehat{SR}_T$  to  $SR$ .

(iv) Given our consistency result, and the preceding remark, it is natural to ask also if there is a central limit theorem for the empirical maximum Sharpe ratio; is it approximately normally distributed, in large samples? This question is of course important for hypothesis testing and other purposes. See, for example, Whitelaw (1997) for some empirical evidence in this respect. We can show in fact that the empirical maximum Sharpe ratio is asymptotically normal in some situations but not in others. We defer details of this analysis to another time.

## REFERENCES

- Black, F., and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal* **48**(5), 28–43.
- Elton, E. J., Gruber, M. J., Comer, G., and Li, K. (2002). Spiders: where are the bugs? *Journal of Business* **75**, 453–72.
- Emmer, S., Klüppelberg, C., and Korn, R. (2001). Optimal portfolios with bounded capital-at-risk. *Mathematical Finance* **11**, 365–84.
- Jobson, J. D., and Korkie, B. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance* **36**, 889–908.
- Jorion, P. (1992). Portfolio optimization in practice. *Financial Analysts Journal* **48**, 68–74.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance* **7**, 77–91.
- Miller, R. E., and Gehr, A. K. (1978). Sample size bias and Sharpe's performance measure: A note. *Journal of Financial and Quantitative Analysis* **12**, 943–6.
- Olienyk, J., Schwebach, R., and Zumwalt, J. K. (1999). WEBS, SPDRs, and country funds: An analysis of international cointegration. *Journal of Multinational Financial Management* **9**, 217–32.
- Olienyk, J., Schwebach, R., and Zumwalt, J. K. (2001). The impact of financial crises on international diversification. *Global Finance Journal* **13**, 147–161.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19**, 425–42.
- Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business* **39**, 119–38.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, Fall, 49–58.
- Tobin, J. (1958). Liquidity preference as behavior toward risk. *Review of Economic Studies* **25**, 65–86.
- Whitelaw, R. (1997). Time-varying Sharpe ratios and market timing. Working paper, available at <http://pages.stern.nyu.edu/rwhitela/research.html>
- Wu, G., and Xiao, Z. (2002). An analysis of risk measures. *Journal of Risk* **4**, 53–75.

