J Econ Finan (2013) 37:150–157
DOI 10.1007/s12197-011-9213-8

# A note on the evaluation of long-run investment decisions using the sharpe ratio

**Ken Johnston · John Hatem · Elton Scott**

Published online: 28 October 2011


**Abstract** This paper reexamines the use of the Sharpe ratio to measure the performance of large and small company stocks along with corporate bonds over different holding periods. It builds on previous research which cites the effects of serial correlation and non-normality in the creation of estimation error in the calculation of the Sharpe ratio. It finds that higher order moments such as skewness and kurtosis are a further source of error that must be accounted for when making inferences about asset performance.

**Keywords** Sharpe Ratio · Investment Horizon · Serial Correlation

## 1 Introduction

The Sharpe ratio measures excess return per unit of risk where risk is measured by the standard deviation of the excess returns (Sharpe 1994). The ratio itself is a simple calculation but as pointed out by Lo (2002) the accuracy of the ratio is dependent on the time series properties of the return series. Mean reversion, serial correlation, and aggregation methodology all have an effect on the calculation. Bao and Ullah (2006) examine the bias that results from autocorrelation in the return series, while Bao

K. Johnston (✉)
Campbell School of Business, Department of Accounting and Finance, Berry College, 2277 Martha Berry Hwy NW, Mt Berry, GA 30149, USA
e-mail: kjohnston@berry.edu

J. Hatem
College of Business, Department of Finance and Quantitative Analysis, Georgia Southern University, Statesboro, GA 30460, USA
e-mail: jhatem@georgiasouthern.edu

E. Scott
Flagler College—Tallahassee, 444 Appleyard Drive, Tallahassee, FL 32304, USA
e-mail: scott.elton@gmail.com

(2009) relaxes the normality assumption. The above cited works suggest that the Sharpe ratio should be adjusted for the estimation error in order to make correct inferences about asset return series.

To further highlight this point, the work of Best et al. (2007) (hereafter referred to as BHY (2007)) is reexamined. It is shown that absence the presence of serial correlation, other properties of the return series, such as higher order moments, are an additional source of estimation error that must be accounted for before inferences can be made with the Sharpe ratio.

## 2 Literature review

BHY (2007) discuss how the presence of serial correlated returns will result in different Sharpe ratios than those implied by the multi-period independent return based Sharpe ratio (see: Levy (1972)). This is due to the effect serial correlation has on volatility over time. Mean reversion (negative serial correlation) reduces return volatility. Mean aversion (positive serial correlation) increases volatility. Fama and French (1988) found mean reversion in long horizon stock returns as did Poterba and Summers (1988) and Lewellen (2001). Strong and Taylor (2001) and Siegel (2002) found mean aversion in fixed-income securities.

BHY (2007) argue that their simulation results, using annual data from the Ibbotson Associates yearbook, are consistent with mean reversion in equity returns and mean aversion in fixed income returns.[1] However, they fail to mention the possibility of estimation bias due to autocorrelation which is discussed by Lo (2002) and Bao and Ullah (2006). They also don't examine the effect of their sampling methodology on the correlation of the return series.

## 3 Methodology and results

Annual data for large and small company stocks, corporate bond, and treasury bills from 1926 through 2006 (80 data points) are collected using the Ibbotson Associates yearbook (Morningstar 2007).[2] Sample return distributions for holding periods of one to 25 years are generated.

Previous empirical evidence cited by BHY (2007) that finds serial correlation typically uses overlapping monthly observations. For example, Fama and French (1988) when estimating annual autocorrelations, have data whose adjacent observations overlap by 11 months. This is done since autocorrelations are biased downward in finite samples when the effective sample sizes in long-horizon regressions are small. For instance, with a 75 year sample and 5-year returns, there are 15 independent observations. With this small a sample the bias toward finding mean reversion could be significant. To remedy this problem most researchers

[1] Originally published by Ibbotson, the Stocks, Bonds, Bills, and Inflation® yearbooks are now available from Morningstar. See the yearbook for a detailed description of the portfolios.

[2] Results of this paper do not change if sample data is restricted to the 1926–2000, the sample time period used by Best et al. (2007).

**Table 1** Annual returns first order serial correlation 1926–2006 (80 observations) $R_t = \alpha_t + B_1R_{t-1} + \varepsilon_t$

| | α | B1 | R2 | | α | B1 | R2 |
|---|---|---|---|---|---|---|---|
| LCS | 0.1199 | 0.0291 | 0.0008 | CB | 0.057 | 0.0777 | 0.006 |
| | (t=4.50) | (t=0.26) | | | (t=4.81) | (t=0.69) | |
| SCS | 0.1657 | 0.0576 | 0.0033 | T-Bills | 0.0035 | 0.9124 | 0.8316 |
| | (t=3.96) | (t=0.51) | | | (t=1.54) | (t=19.63) | |

$R_t$=is the annual return for period t. LCS=Large Company Stocks, SCS=Small Company Stocks, CB=Corporate Bonds, and T-bills=Treasury Bills

estimate these autocorrelations using overlapping monthly observations to increase the power of the tests. This procedure induces strong correlations among the residuals and methods to adjust for the correlation are valid only in large samples.

BHY (2007) have 75 independent annual returns (1926–2000). Their methodology preserves the auto-correlation of the annual return series rather than the n-period historical non-overlapping serial correlation. For example, when simulating 5-year holding period returns, they are sampling with replacement 5-years of consecutive return data. While this preserves in each trial of the simulation, the 5-year annual serial correlation, it does not preserve the inherent historical serial correlation of 5-year holding period returns based on non-overlapping periods. This is not an issue if you assume that the investor also has an n-period investment horizon. If one makes this assumption it is the auto-correlation of the annual returns series that is of interest. Since it the serial correlation of the annual data that is of interest, it needs to be determined if the annual Ibbotson Associates data set has significant serial correlation.

Table 1 shows the annual first order serial correlation in the data. This measures the degree to which returns in one time period are directly correlated with returns in the next time period.[3] Large company stocks (LCS), Small company stocks (SCS) and corporate bonds (CB) all exhibit very low levels of mean aversion (positive first order serial correlation). Excluding T-bills, the regression $R^2$s and t-statistics indicate no significant serial correlation is found in the annual Ibbotson Associates data.

An additional issue is related to the need for simulations that attempt to preserve the historical correlations in returns across time. To create this data, holding period returns are based on consecutive years. For a given holding period, a year is randomly selected from the total number of years minus (n-1) years.[4] The compounded n-year holding period return, for every portfolio, is computed using each portfolio's return from the randomly selected year and the returns for the next n-1 consecutive years.[5] This procedure is repeated creating sample holding-period return distributions for each portfolio. When sampling with replacement and with a large enough simulation all data points in each n-period return distribution will be represented in the overall simulation in approximately the same proportion as in the

[3] Higher order serial correlation results are similar.

[4] The elimination of the last n-1 years is necessary to guarantee that the n-year holding-period return can be computed.

[5] Lin and Chou (2003) examining equity returns also use this procedure to preserve serial correlation.

initial n-period historical overlapping distribution. When calculating mean returns and standard deviations and hence Sharpe ratios, the order of returns does not matter. Therefore the simulated n-period mean returns and standard deviations will approximately equal the historical overlapping n-period historical distribution mean return and standard deviation. Therefore there is no need for the simulation.

In Table 2 Panel A, the historical 5-year overlapping return distributions' average return and standard deviation for the different asset classes is shown. Comparing these average returns and standard deviations with the simulated results in Panel B, overall as the number of trials in the simulation increase, the means and standard deviations approach that of the historical distributions. The results demonstrate that, for the stocks and bonds, the simulation size (250) used in BHY (2007) is too small.

Our results indicate that the findings of BHY (2007) are not driven by serial correlation since no significant annual serial correlation exists in the data. The serial correlations are too small to significantly affect the n-period volatility and hence the Sharpe ratio. What then is driving their findings that using the auto-correlated returns procedure with the original Sharpe ratio are so much different from the results using the multi-period Sharpe ratio which assumes returns are independently and identically distributed (i.i.d.) across time?

BHY (2007), find that when using auto-correlated returns the Sharpe ratios for large and small company stocks are larger than the Sharpe ratio for corporate bonds for all holding periods. When using independent returns the Sharpe ratio for corporate bonds is larger than the Sharpe ratios for large and small company stocks, when the holding period is greater than 16 years. Therefore to examine what is driving their results, in Table 3, the historical 20-year serial correlated overlapping returns and the 20-year simulated independently and identically distributed returns are shown.[6,7]

Hodges et al. (1997) compared the original Sharpe ratio with the multi-period Sharpe ratio. Both simulation methods assumed that securities are i.i.d.. They find that the behavior of the multi-period Sharpe ratios is close to that of the single-period Sharpe ratios where bonds outperformed stocks in the long run, therefore for simplicity the original Sharpe ratio is used instead of the multi-period Sharpe ratio of Levy (1972).

What is driving their results is that sampling a sequence of returns reduces the likelihood of getting data points from the tails of the historical return distributions in consecutive years. Therefore, BHY's (2007) serial correlated methodology results in a distributional shift when compared to the i.i.d. simulated distribution (brings in the tails of the distribution, higher order moments (skewness, kurtosis) are also significantly changed). These distributional shifts affect the standard deviation and hence the Sharpe ratio.

[6] Only 30,000 trials are run in this simulation due to software constraints of Insight from AnalyCorp. Although it can run up to 1,000,000 trials (gives you the mean and standard deviation), it will only let you view a maximum of 30,000 trials. The individual trial data is required to calculate the higher order moments.

[7] With a large sample size, a simulated n-period serial correlated return distribution will be nearly identical to the historical n-period overlapping distribution. Therefore it is appropriate to use the historical distribution in place of the simulated.

**Table 2** 5 year holding returns summary statistics (1926–2006)

| Historical serial correlated returns (overlapping data) | | | | | Simulated serial correlated returns (overlapping data) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$=77 | LCS | SCS | CB | T-bills | Simulation size 250 | LCS | SCS | CB | T-bills |
| Average Return | 0.7358 | 1.1969 | 0.3626 | 0.2138 | Average Return | 0.7345 | 1.1905 | 0.3662 | 0.2009 |
| Standard Deviation | 0.6175 | 1.3420 | 0.3149 | 0.1776 | Standard Deviation | 0.6789 | 1.3708 | 0.3004 | 0.1677 |
| | | | | | Simulation Size 500,000 | | | | |
| | | | | | Average Return | 0.7377 | 1.1981 | 0.3622 | 0.2137 |
| | | | | | Standard Deviation | 0.6141 | 1.3347 | 0.3123 | 0.1763 |

LCS=Large Company Stocks, SCS=Small Company Stocks, CB=Corporate Bonds, and T-bills= Treasury Bills

In Table 3, for large and small company stocks, as the sampling method is changed from using serial correlated returns to sampling i.i.d., the distributions become much more leptokurtic (positive kurtosis). That is, this distribution has a higher probability than a normally distributed variable of values near the mean and a higher probability than a normal distributed variable of extreme values. These distributions also become more positively skewed. For example, the maximum positive values for large and small company stocks increase from 25.8160 (2,582%) to 145.1572 (14,516%) and 45.2324 (4,523%) to 2,031.8428 (203,184%) respectively. The higher probability of extreme observations and the large increase in magnitude of these extreme observations significantly increases the standard deviations. For large company stocks the standard deviation increase from 5.8403 (584%) to 9.3353 (934%), while for small company stocks the standard deviation increases from 9.7200 (972%) to 46.7183 (4,672%). These increases in standard deviations are not accompanied by similar increases in returns therefore you see significant decreases in the Sharp ratios.

Comparing the 20-year historical serial correlated (overlapping) returns to 20-year simulated i.i.d. returns, the Sharpe ratios decline for large and small company stocks from 1.3267 and 1.5625 to 0.8684 and 0.4771 respectively.[8]

For corporate bonds as the sampling method changes from the serial correlated returns to sampling i.i.d., the distributions changes from platykurtic (negative kurtosis) to leptokurtic. A platykurtic distribution has a smaller peak around the mean and thin tails, which indicates a lower probability than a normal distribution of values near the mean and extreme values. It also has a higher probability of mid-range values. Furthermore, as the sampling method changes from the serial correlated returns to sampling i.i.d., the increase in size of extreme values is not near the magnitude of that found with the large and small company stocks. The maximum positive value for corporate bonds increases from 8.8674 to 13.4426, a

[8] The Sharpe ratios in this study are calculated using the differential return. Each trial's n-period compounded risk free rate is subtracted from the n-period compounded asset return. This array of differences is used to calculate the average difference and standard deviation of differences (See: Sharpe (1994)).

**Table 3** 20 year serial correlated and independently and identically distributed returns (1926–2006)

| 20 year historical serial correlated returns (Overlapping) | | | | | 20 year independent and identically distributed returns (30,000 Trials) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | LCS | SCS | CB | T-bills | | LCS | SCS | CB | T-bills |
| Average Return | 9.2737 | 16.7305 | 2.6000 | 1.4907 | Average Return | 9.2085 | 23.3904 | 2.3390 | 1.0946 |
| Standard Deviation | 5.8403 | 9.7200 | 2.4328 | 1.2094 | Standard Deviation | 9.3353 | 46.7183 | 1.2196 | 0.2771 |
| Min | 0.8442 | 2.0533 | 0.3061 | 0.0878 | Min | −0.8309 | −0.9698 | −0.1438 | 0.3253 |
| Max | 25.8160 | 45.2324 | 8.8674 | 3.4233 | Max | 145.1572 | 2031.8428 | 13.4426 | 2.6403 |
| Skewness | 0.8026 | 0.9610 | 0.9548 | 0.3312 | Skewness | 2.6696 | 11.4520 | 1.4476 | 0.5888 |
| Kurtosis | 0.4301 | 0.4594 | −0.5291 | −1.4846 | Kurtosis | 12.6547 | 300.6316 | 3.9351 | 0.6017 |
| Sharpe Ratio | 1.3267 | 1.5625 | 0.6131 | | Sharpe Ratio | 0.8684 | 0.4771 | 1.0389 | |

LCS=Large Company Stocks, SCS=Small Company Stocks, CB=Corporate Bonds, and T-bills=Treasury Bills

52% increase. Compare that to a 462% increase for large stocks and a 4,392% increase for small stocks. Thus for corporate bonds the mid-range values have a much bigger impact on the standard deviation than they do for stocks. Given their respective platykurtic and leptokurtic distributions, the historical 20-year serial correlated distribution (overlapping) of corporate bonds has a higher standard deviation than the simulated 20-year i.i.d. distribution. Since there is not much different in the returns, the Sharpe ratio declines when 20-year historical serial correlated (overlapping) returns are used instead of 20-year i.i.d. sample.

## 4 Conclusion

This study demonstrates that simulation methods can induce bias in the Sharpe ratio that lead to a misinterpretation of the results generated from portfolios of large company stocks, small company stocks, and corporate bonds for holding periods of one to 25 years.

There are two main issues related to this study. First, researchers need to show that serial correlation is indeed present in any return series rather than a by product of the simulation method. The second issue has to do with the need for simulations, related to the auto-correlated returns, in the first place. When sampling with replacement and conducting a large enough simulation all data points in each n-period return distribution will be represented in the overall simulation in approximately the same proportion as in the initial n-period historical overlapping distribution. When calculating returns and standard deviations and hence Sharpe ratios, the order of returns does not matter. Therefore the simulated n-period mean returns and standard deviations will approximately equal the historical overlapping n-period historical distribution mean return and standard deviation. Therefore there is no need for the simulations.

Results indicate that the use of auto-correlated n-period returns brings in the tails of the distribution and significantly changes the higher order moments (skewness, kurtosis) when compared to an n-period i.i.d. simulated distribution. These distributional shifts effect the standard deviation and hence the Sharpe ratio.

## References

Bao Y (2009) Estimation risk-adjusted Sharpe ratio and fund performance ranking under a general return distribution. J Financ Econometrics 7(2):152–173

Bao Y, Ullah A (2006) Moments of the estimated sharpe ratio when the observations are not IID. Finance Res Lett 3:49–56

Best R, Hodges C, Yoder J (2007) The sharpe ratio and long-run investment decisions. J Invest 70–76

Fama E, French K (1988) Permanent and temporary components of stock prices. Scot J Polit Econ 246–273

Hodges C, Walton R, Yoder J (1997) Stocks, bonds, the sharpe ratio, and the investment horizon. Financ Analysts J 74–80

Levy H (1972) Portfolio performance and the investment horizon. Manag Sci 645–653
Lewellen J (2001) Temporary movements in stock prices. Working paper, MIT
Lin MC, Chou PC (2003) The pitfall of using the sharpe ratio. Finance Lett 1:84–89
Lo AW (2002) The statistics of Sharpe ratios. Financ Analysts J 58:36–52
Morningstar (2007) Stocks, bonds, bills, and inflation. 2007 yearbook. Morningstar, Inc, Chicago
Poterba J, Summers L (1988) Mean reversion in stock prices: evidence and implications. J Financ Econ 27–59
Sharpe W (1994) The Sharpe ratio. J Portf Manag 49–58
Siegel JJ (2002) Stocks for the long run. McGraw-Hill, New York
Strong N, Taylor N (2001) Time diversification: empirical tests. J Bus Finance Account 28(3–4):263–302