Sharpe Ratio:
Estimation, Confidence Intervals, and Hypothesis Testing

Two Sigma Technical Report 2018-001

Matteo Riondato
Labs, Two Sigma Investments, LP
matteo@twosigma.com

June 14, 2018

Abstract

We survey and discuss methods proposed in the literature for *1.* estimating the Sharpe ratio; *2.* computing confidence intervals around a point estimation of the Sharpe ratio; and *3.* performing hypothesis testing on a single Sharpe ratio and on the difference between two Sharpe ratios.

1 Introduction

The *Sharpe ratio* [Sha65; Sha94] is a widely used measure of the performance of an investment strategy (i.e., of a *portfolio*). Informally, the Sharpe ratio is the risk-adjusted expected excess return of a portfolio w.r.t. a benchmark strategy. Formally, let p be a portfolio and let R_p be the return of p over a time interval t (e.g., R_p could be the daily return, so $t =$ one day).[1] We assume that R_p is a random variable with non-zero variance, i.e., $\mathrm{Var}[R_p] > 0$. Let also b be a benchmark investment strategy, and let R_b denote its return over t. The benchmark b may be riskless, hence R_b may be a fixed constant, or b may involve risk, in which case R_b is a random variable with non-zero variance. In either case, we assume that $\mathrm{Var}[R_p - R_b]$ exists and is non-zero.

Definition 1 (Sharpe ratio)**.** The *Sharpe ratio of* p *w.r.t.* b is the ratio between the expected excess return of p w.r.t. b over the standard deviation of this same excess return:

$$\zeta_{p,b} = \frac{\mathbb{E}[R_p - R_b]}{\sqrt{\mathrm{Var}[R_p - R_b]}} \ . \tag{1}$$

Unless otherwise specified, we consider the portfolio p and the benchmark b to be fixed, hence we almost always drop the subscripts from the notation of the Sharpe ratio, using ζ to denote $\zeta_{p,b}$.

Given a finite set D of observed returns from n time intervals, $D = \{(R_p^{(1)}, R_b^{(1)}), \ldots, (R_p^{(n)}, R_b^{(n)})\}$, it is not possible, in general, to obtain the *exact* Sharpe ratio ζ from D. For all practical purposes, a high-quality *estimation* $\hat{\zeta}$ of ζ, i.e., an estimation enjoying specific desirable statistical properties, can be used in place of the exact value ζ. The goal is then to compute the best possible estimation $\hat{\zeta}$

[1] We consider the case for absolute returns, i.e., the unit of measure for the values R_p is "dollars." Another possibility is to use *logarithmic relative returns* ("log-returns"), where R_p is the logarithm of the relative return w.r.t. the previous period $p - 1$. The logarithm is necessary to preserve additivity. Most of what we discuss in this work also applies to the case of log-returns. Whether to use absolute returns or log-relative returns must be a very deliberate and informed decision. Scenarios in which using log-relative return would probably be preferable are when the absolute returns are heteroskedastic.

of ζ from D. Once such a point estimation $\hat{\zeta}$ is available, it can be used to derive confidence intervals containing the true value ζ, or to test hypotheses about ζ or about the difference $\zeta_{p_1,b} - \zeta_{p_2,b}$ of the Sharpe ratios of two different portfolios (w.r.t. the same benchmark).

Purpose of this document. Our goal is to *survey existing methods* presented in the econometrics literature for estimating the Sharpe ratio, computing confidence intervals around a point estimation, and performing hypothesis testing involving the Sharpe ratio. We also aim to provide a few humble clarifications to some of the confusion in the existing literature.

The Sharpe ratio has received wide attention in the finance and economics literature, and it is heavily relied upon by practitioners. Not all the attention it has received has been in the form of praise, and many researchers have developed other measures or variants of the Sharpe ratio to measure the performance of investment strategies (see, e.g., [Sha94; Sch07; Isr05; Bac09, and references therein]). We do not comment on which measure should be used and in which cases one measure is better than another. Instead, we focus on the Sharpe ratio and how to estimate it correctly from a statistical point of view. A similar study would be warranted if we were to consider other performance measures.

Outline. This document is organized as follows. In Sect. 2 we present methods for point estimation of the Sharpe ratio. Sect. 3 contains procedures to obtain confidence intervals around a point estimation. We discuss hypothesis tests for a single Sharpe ratio and for the difference of two Sharpe ratios in Sect. 4. We conclude in Sect. 5, presenting some open questions and research directions. In Appendix A we discuss the double Sharpe ratio, an alternative measure of performance for investment strategies.

2 Estimation

In this section we present estimators for ζ and discuss their properties, including bias, variance, efficiency, and asymptotic distribution. We also touch on time aggregation (e.g., annualization), given its important role in practice.

Notation. In our general setting, each excess return $R_p - R_b$ is a random variable, which we denote with Y. Independently from the distribution of Y, the expectation of Y is denoted with $\mu = \mathbb{E}[Y]$ and the standard deviation of Y with $\sigma = \sqrt{\mathrm{Var}[Y]}$. We use $\hat{\mu}$ to denote the *sample mean* of a set $D = \{Y^{(1)}, \ldots, Y^{(n)}\}$ of n sample excess returns, and $\hat{\sigma}$ to denote the *sample standard deviation* computed from D. Formally:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y^{(i)} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y^{(i)} - \hat{\mu}\right)^2} \ .$$

Definition 2. The *basic estimator* $\hat{\zeta}_{\mathrm{basic}}$ for ζ is defined as the ratio between the sample mean $\hat{\mu}$ and the sample standard deviation $\hat{\sigma}$:

$$\hat{\zeta}_{\mathrm{basic}} = \frac{\hat{\mu}}{\hat{\sigma}} \ . \tag{2}$$

As observed by Miller and Gehr [MG78], the basic estimator is closely related to the Student's t-statistic for testing whether the expectation of a random variable is zero:

$$t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} = \sqrt{n}\hat{\zeta}_{\mathrm{basic}} \ . \tag{3}$$

Hence results that hold for this t-statistic also hold for the Sharpe ratio, up to the scaling factor, as we discuss in the following.

2.1 Distribution under normal i.i.d. excess returns

We now look at a first example of the close relationship between the t-statistic from (3) and $\hat{\zeta}_{\text{basic}}$.

First, we temporarily make the following key assumption, which simplifies the analysis but is extremely unrealistic, especially in finance settings [LM02]. We remove this assumption later.

Assumption 1 (Indendence and normality of the excess returns)**.** The excess returns $Y^{(i)}$, $1 \le i \le n$, are i.i.d. samples from a normal distribution $\mathcal{N}(\mu, \sigma^2)$.

Under Assumption 1, we have the following fact about the distribution of $\hat{\zeta}_{\text{basic}}$.

Fact 1 (Sect. 17.3.2 [SS10])**.** *Assume that Assumption 1 holds. Then the distribution of $\hat{\zeta}_{\text{basic}}$ is a rescaled non-central t-distribution with $n-1$ degrees of freedom. The non-centrality parameter is $\sqrt{n}\zeta$.*

2.2 Bias and variance

We now discuss the first and second moments of the basic estimator, i.e., its expectation and variance.

Bias. The bias of $\hat{\zeta}_{\text{basic}}$ is a first symptom of the limitations of this estimator. Under some assumption on the distribution of the excess returns, it is possible to quantify (or approximate) the bias of $\hat{\zeta}_{\text{basic}}$. Specifically:

- under Assumption 1, the *exact* expectation of $\hat{\zeta}_{\text{basic}}$ is [Pav15, Sect. 1.3]:

$$\mathbb{E}[\hat{\zeta}_{\text{basic}}] = \zeta \sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \quad (> \zeta) \ ,$$

  where $\Gamma(x) = (x-1)!$ is the gamma function. An unbiased estimator $\hat{\zeta}_{\text{unbiased}}$ for ζ can be obtained by dividing $\hat{\zeta}_{\text{basic}}$ by the bias factor on the r.h.s. of the above equation.

  The bias factor is greater than 1 and tends towards 1 from above quite rapidly as n grows. For example, it is ≈ 1.08 for $n = 12$, ≈ 1.02 for $n = 40$, and ≈ 1.01 for $n = 75$.

- under the assumption that the excess returns are i.i.d. (but not necessarily normal), the expectation of $\hat{\zeta}_{\text{basic}}$ can be rewritten as [Bao09, Sect. 1.2]:[2]

$$\begin{aligned}\mathbb{E}[\hat{\zeta}_{\text{basic}}] = \zeta + \frac{3}{4n}\zeta + \frac{49}{32n^2}\zeta - \gamma_1\left(\frac{1}{2n} + \frac{3}{8n^2}\right) + \gamma_2\zeta\left(\frac{3}{8n} - \frac{15}{32n^2}\right) + \frac{3}{8n^2}\gamma_3 - \frac{5}{16n^2}\gamma_4\zeta \\ - \frac{5}{4n^2}\gamma_1^2\zeta + \frac{105}{128n^2}\gamma_2^2\zeta - \frac{15}{16n^2}\gamma_1\gamma_2 + o(n^{-2}) \end{aligned} \tag{4}$$

  where

$$\gamma_i = \frac{\mathbb{E}[(Y-\mu)^{i+2}]}{\sigma^{i+2}}, i = 1, 2, 3, 4,$$

  i.e., γ_1 is the skewness, γ_2 is the kurtosis, and γ_3 and γ_4 are the third and fourth standardized moments of the distribution of the excess returns, according to the so-called Pearson's definition (e.g., a normal distribution has $\gamma_1 = 0$ and $\gamma_2 = 3$).

  Stopping at the first order term, the expectation can be written as [Bao09, Remark 2]:

$$\mathbb{E}[\hat{\zeta}_{\text{basic}}] = \zeta + \frac{3}{4n}\zeta - \frac{1}{2n}\gamma_1 - \frac{3}{8n}\gamma_2\zeta + o(n^{-1}) \ . \tag{5}$$

[2] Christie [Chr05] and Opdyke [Opd07] also give expressions for the expectation of $\hat{\zeta}_{\text{basic}}$ under the same assumptions, but Bao [Bao09] observes that such expressions are not correct, because they consider $\hat{\mu}$ to be a deterministic value.

The expressions on the r.h.s. of (4) and (5) are not proportional to ζ. Hence developing an (approximately) unbiased estimator becomes more problematic than under Assumption 1, given that ζ, γ_1, γ_2, γ_3, and γ_4 are unknown. In practice, it is possible to replace the unknown values with estimates in, e.g., (5), and obtain an *approximately unbiased* estimator:

$$\hat{\zeta}_{\text{apprunbiased}} = \hat{\zeta}_{\text{basic}} - \frac{3}{4n}\hat{\zeta}_{\text{basic}} + \frac{1}{2n}\hat{\gamma}_1 - \frac{3}{8n}\hat{\zeta}_{\text{basic}}\hat{\gamma}_2 \ . \tag{6}$$

Estimates of γ_1 and γ_2 can be obtained using Fisher's k statistics [Wei16].

Variance. A closed formula for the variance of $\hat{\zeta}_{\text{basic}}$ can be obtained under Assumption 1 [Pav15, Sect. 1.3].

Fact 2. *Under Assumption 1, we have*

$$\mathrm{Var}[\hat{\zeta}_{\text{basic}}] = \frac{(1+n\zeta^2)(n-1)}{n(n-3)} - \mathbb{E}[\hat{\zeta}_{\text{basic}}]^2 \ .$$

When the excess returns are i.i.d. but not necessarily normal, it is possible to obtain an approximation of the variance [Bao09, Sect. 1.2].

Fact 3. *Under the assumption that the excess returns are i.i.d. samples, we have*

$$\begin{aligned}\mathrm{Var}[\hat{\zeta}_{\text{basic}}] = &\frac{1}{n}\left(1+\frac{\zeta^2}{2}\right) + \frac{1}{n^2}\left(\frac{19}{8}\zeta^2+2\right) - \gamma_1\zeta\left(\frac{1}{n}+\frac{5}{2n^2}\right) + \gamma_2\zeta^2\left(\frac{1}{4n}+\frac{3}{8n^2}\right)\\ &+ \frac{5}{4n^2}\gamma_3\zeta - \frac{3}{8n^2}\gamma\zeta^2 + \gamma_1^2\left(\frac{7}{4n^2}-\frac{3}{2n}\zeta^2\right) + \frac{39}{32n^2}\gamma_2^2\zeta^2 - \frac{15}{4n^2}\gamma_1\gamma_2\zeta + o(n^{-2}) \ .\end{aligned}$$

When the assumption of normality and independence does not hold, one can estimate the variance of the estimator using the bootstrap techniques appropriate for non i.i.d. data [LW08; PR92]. In Sect. 3 we give more details about applications of the bootstrap in the context of confidence intervals estimation for the Sharpe ratio. We discuss the asymptotic variance of the basic estimator in Sect. 2.4.

2.3 Best scale invariant estimator and other estimators

Unhapipat, Chen, and Pal [UCP16] propose a different estimator for ζ. Specifically, they consider the parametric class of scale-invariant estimators

$$C = \left\{\hat{\zeta}_d = d\frac{\hat{\mu}}{\hat{\sigma}}, d \in \mathbb{R}\right\}$$

and identify the value d^* such that $\hat{\zeta}_{d^*}$ has the minimum Mean Squared Error (MSE) among all the members of C. The optimal d^* is

$$d^* = \frac{n-3}{\sqrt{2(n-1)}}\frac{\zeta^2}{\zeta^2+1/n}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

which involves the unknown exact Sharpe ratio ζ. Unhapipat, Chen, and Pal [UCP16] thus propose to use

$$d = \frac{n-3}{\sqrt{2(n-1)}}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \quad (\geq d^*) \ .$$

The resulting estimator is called the *best scale invariant estimator*,[3] denoted with $\hat{\zeta}_{\text{bsie}}$.

[3] The adjective "best" is slightly abused here, as it would be justified only when using d^*.

Other estimators. Challet [Cha15] presents a completely different approach to Sharpe ratio estimation, as his estimator does not use the sample mean and sample standard deviation. Instead, he argues that the total duration of the drawdowns and the drawups of a price time series is an efficient estimator of the Sharpe ratio. His derivation relies on the assumption that the returns are i.i.d.. Challet's estimator is much more efficient than $\hat{\zeta}_{\text{basic}}$ when the distribution of the returns has fat tails, but not otherwise. This estimator was introduced very recently, and no results are currently known on its properties.

Only for the sake of completeness, we mention the estimator proposed by Skrepnek and Sahai [SS13]. This estimator resembles that of Unhapipat, Chen, and Pal [UCP16], with an additional factor in the numerator, (claiming to be) taking into account an estimation of the coefficient of variation. Evaluating the correctness and the performances of this estimator is a strenuous task, due to lack of rigorous proofs in the paper. Similar issues also make it hard to evaluate the estimators presented in another work by the same authors [SS11].

2.4 Asymptotic distribution

We now present results on the asymptotic normality of the basic estimator of the Sharpe ratio, i.e., on the fact that its distribution tends toward a normal distribution as the sample size n grows. The most important aspect of this convergence is the *asymptotic variance*.[4] We start from a result requiring a most restrictive assumption on the distribution of the excess returns and then show how it can be relaxed using the Generalized Method of Moments (GMM) [Han82]. We present the results for $\hat{\zeta}_{\text{basic}}$, but they also hold for $\hat{\zeta}_{\text{unbiased}}$ and $\hat{\zeta}_{\text{bsie}}$, under the same assumptions [UCP16].

Assuming normal i.i.d. excess returns. The following result is a straightforward application of the Central Limit Theorem and of the delta method [Was03, Sect. 9.9].

Fact 4 ([Opd07; SS10][5]). *Under Assumption 1,*[6] *the basic estimator* $\hat{\zeta}_{\text{basic}}$ *is* asymptotically normal *in n, i.e.,*

$$\sqrt{n}(\hat{\zeta}_{\text{basic}} - \zeta) \xrightarrow{d} \mathcal{N}\left(0, 1 + \frac{\zeta^2}{2}\right) \quad . \tag{7}$$

Assuming a stationary distribution of excess returns. Mertens [Mer02] observes that wrongly assuming normality of the excess returns leads to estimates of the asymptotic variance that may be up to 70% off from their true values. The assumption of normality can be removed and the asymptotic variance can be computed under just the independence assumption. We focus here on the even more general case of that the process generating the excess returns is stationary and ergodic. In this case one can use the GMM [Han82] to prove the asymptotic normality of $\hat{\zeta}_{\text{basic}}$ and compute the asymptotic variance.

Fact 5 ([Lo02; LW08]). *Assume that the stochastic process from which the excess returns are sampled is stationary and ergodic.*[7] *Let $m_2 = \mathbb{E}[Y^2]$, and for any time interval t let*

$$\theta_t = \begin{bmatrix} Y_t - \mu \\ Y_t^2 - m_2 \end{bmatrix}$$

where Y_t is the excess return in the interval t. Let also

$$q = \begin{bmatrix} \frac{m_2}{(m_2-\mu^2)^{3/2}} \\ -\frac{\mu}{2(m_2-\mu^2)^{3/2}} \end{bmatrix} \quad .$$

[4] One could argue that the rate of convergence is at least as important an aspect. No results are currently available on the rate of convergence for the basic estimator.

[5] Lo [Lo02] gives the same result with the additional claim that it holds for the case of non-necessarily-normal i.i.d. excess returns but Bao [Bao09, Footnote 5] and Mertens [Mer02] observe that the claim is not justified.

[6] More precisely, under the assumption that the excess returns are i.i.d. samples from a distribution with the same constraints on the first four central moments as the normal [Mer02].

[7] Some additional technical assumptions are needed but are quite standard [Lo02; LW08].

Then

$$\sqrt{n}(\hat{\zeta}_{\text{basic}} - \zeta) \xrightarrow{d} \mathcal{N}\left(0, q^\intercal \Psi q\right) \text{ where } \Psi = \lim_{n\to\infty} \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \mathbb{E}[\theta_s \theta_t^\intercal] \ .^{8} \tag{8}$$

An estimation of the asymptotic variance in this case can be computed by first obtaining an Heteroskedastic and Autocorrelation Consistent (HAC) estimation of Ψ using the kernel methods by Andrews [And91], and then an estimation of d from the sample excess returns.

For completeness, we mention the work by Bao and Ullah [BU06], who discuss the asymptotic distribution of $\hat{\zeta}_{\text{basic}}$ when the excess returns are normal but not necessarily i.i.d.. There is significant evidence suggesting that real excess returns are not normally distributed, so we do not discuss the results by Bao and Ullah [BU06] any further. Pav [Pav15, Sect. 1.7.1] presents a discussion of a similar case, but assuming fixed autocorrelation. He also discusses adjustements to $\hat{\zeta}_{\text{basic}}$ under a limited case of heteroskedasticity [Pav15, Sect. 1.7.2].

2.5 Time aggregation (e.g., annualization)

We discuss now an important aspect of the Sharpe ratio that, despite not being a specific characteristic of the basic estimator nor strictly related to estimation, must be taken into consideration in practice: the Sharpe ratio is *not independent* from the time period t for which the excess returns $Y^{(i)}$ are considered [Sha94; Lo02]. In other words, a Sharpe ratio that considers returns over monthly periods cannot be compared directly to one that considers returns over years. The same holds, naturally, also for the basic estimator. It is possible to "transform" a Sharpe ratio using returns measured at higher frequency (e.g., monthly) into a Sharpe ratio using returns measured at lower frequency (e.g., yearly).[9]

Suppose that we want to transform a Sharpe ratio measured w.r.t. shorter time periods of length t (e.g., $t =$ one month) into a Sharpe ratio measured w.r.t. longer time periods of length T (e.g., $T =$ one year). Assume that t divides T evenly and let $q = T/t$. Given a sequence of q excess returns w.r.t. q consecutive time periods of length t, the excess return for the same period of length T is, ignoring the effects of compounding,

$$Y_T = \sum_{i=1}^{q} Y_t^{(i)} \ .$$

Under the assumption that the excess returns are i.i.d., the variance of Y_T is proportional to q. Hence the Sharpe ratio ζ_T w.r.t. periods of length T is

$$\zeta_T = \frac{\mathbb{E}[Y_T]}{\sqrt{\operatorname{Var}[Y_T]}} = \frac{q\mathbb{E}[Y_t]}{\sqrt{q\operatorname{Var}[Y_t]}} = \sqrt{q}\zeta_t, \tag{9}$$

where ζ_t is the Sharpe ratio w.r.t. periods of length t [Lo02].

[8] We would like to clarify some of the confusion in the econometrics literature about the correctness of this result. Lo [Lo02] uses a GMM approach to compute the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$, where $\theta = (\mu, \sigma^2)^\intercal$ and $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)^\intercal$. He then uses the delta method to obtain the result in (8) (in a slightly different form). Ledoit and Wolf [LW08] present essentially the same approach of Lo [Lo02], with a slightly different definition of the moment conditions (using m_2 instead of σ^2), which is what we follow.

Christie [Chr05] attacks the problem using a different GMM approach that does not require a successive application of the delta method. When the results are stated in the main body of the paper, the stated assumptions are that the process is stationary and ergodic. In the appendix, when the results are proved, it is mentioned that *normality* is *not* assumed, but it is not clearly stated whether *independence* is assumed. Bao [Bao09, Footnote 6] comments that it is "questionable" that Christie's "asymptotic variance expression holds under the very general setup of non-i.i.d. returns", but he does not point to specific mistakes, while Opdyke [Opd07] embraces Christie's result and claims that it holds for non-i.i.d. excess returns. A close look at Christie's derivation suggests that it may only hold for the i.i.d. case. Specifically, his definition of the variance-covariance matrix of the moment conditions [Chr05, Eq. 8] may only hold for the i.i.d. case. The definitions of the same matrix in [LW08, Sect. 3.1] and in [Lo02, Eq. A7] instead seem to hold for the general case..

[9] The opposite direction, i.e., transforming from lower to higher frequency, is mathematically possible, but not sensible from a statistical point of view.

A basic estimation for ζ_T is $\hat{\zeta}_T = \sqrt{q}\hat{\zeta}_t$. This is just the basic estimator and has the properties discussed in the previous paragraphs. For example, it is asymptotically normal, although with an appropriately scaled higher variance:

$$\sqrt{T}(\hat{\zeta}_T - \zeta_T) \xrightarrow{d} \mathcal{N}\left(0, q\left(1 + \frac{\zeta_t^2}{2}\right)\right) \ .$$

The simple scaling of ζ_t to obtain ζ_T cannot be used when the returns are not i.i.d., as the variance of Y_T is no longer the sum of the variances of the $Y_t^{(i)}$ [Lo02]. Instead, the scaling factor becomes a more complex expression of the autocorrelations of the excess returns. For ease of presentation, we only present the case of *stationary* excess returns. Lo [Lo02] discusses the general case for non-i.i.d. excess returns. Let ρ_k be the k-th order autocorrelation of Y_t:[10]

$$\rho_k = \frac{\mathrm{Cov}[Y_t, Y_{t-k}]}{\mathrm{Var}[Y_t]} \ .$$

Then the relationship between ζ_t and ζ_T is

$$\zeta_T = \frac{q}{\sqrt{q + 2\sum_{k=1}^{q-1}(q-k)\rho_k}}\zeta_t \ . \tag{10}$$

When estimating ζ_T with $\hat{\zeta}_T$ using $\hat{\zeta}_t$, the autocorrelations ρ_k must be replaced by appropriate estimations $\hat{\rho}_k$. The accuracy of these estimations may have an impact on the performances of $\hat{\zeta}_T$, including its asymptotic variance.

3 Confidence intervals

In this section we discuss how to compute confidence intervals for the Sharpe ratio. We start from methods that use the exact distribution of the basic estimator, and then we move to approaches relying on its asymptotic variance (see Sect. 2.4) and to bootstrap-based methods.

3.1 Confidence intervals with normal i.i.d. returns

From Fact 1, we have that under Assumption 1, we can obtain a confidence interval on ζ by inverting the cumulative distribution of the non-central t-distribution [Pav15]. The resulting confidence interval would have exactly the desired nominal coverage, but its computation may be too expensive, due to the necessity of inverting the CDF.

3.2 Confidence intervals with the asymptotic variance

Approximate confidence intervals, i.e., confidence intervals whose actual coverage is not guaranteed to be the nominal coverage, can be obtained by exploiting the fact that the basic estimator is asymptotically normal (Sect. 2.4) [UCP16; Lo02; Mer02].

The plugin approach. The derivation of the approximate confidence intervals may follow a well-known recipe [Was03, Thm. 6.16], known as the *plugin approach.* Let $\hat{\zeta}$ be a point estimate of ζ and let $\hat{\mathsf{se}}$ be the *estimated* standard error of $\hat{\zeta}$. Given $\alpha \in (0,1)$, let $z_{\alpha/2}$ be the $1-\alpha/2$-th quantile of the standard normal distribution, and define

$$C = (\hat{\zeta} - z_{\alpha/2}\hat{\mathsf{se}}, \hat{\zeta} + z_{\alpha/2}\hat{\mathsf{se}}) \ .$$

[10] On the basis of the efficient market hypothesis, one can assume the daily returns have n$\rho_k = 0$ for all $k \geq 2$. In this case, the annualization can be (approximately) obtained by just multiplying the daily Sharpe ratio by $\sqrt{q/1 + 2\rho_1}$.

Then C is an approximate confidence interval for ζ, with approximate coverage $1 - \alpha$:

$$\Pr(\zeta \in C) \to 1 - \alpha,$$

where the probability is taken w.r.t. the data-generating process.

Unhapipat, Chen, and Pal [UCP16] present experimental evidence that, in general, the confidence intervals obtained following this procedure fall short of the nominal coverage.

"Confidence interval – Hypothesis testing"–equivalency approach. Approximate confidence intervals can also be computed by inverting the expression

$$\frac{(\hat{\zeta} - \zeta)}{\mathsf{se}(\zeta)} \in [-z_{\alpha/2}, z_{\alpha/2}],$$

where $\mathsf{se}(\zeta)$ is the standard error of $\hat{\zeta}$, where we highlighted the dependency on ζ. Inverting this expression allows us to obtain inequalities for ζ. The square root of the estimated asymptotic variance can be used in place of the actual standard error.

3.3 Confidence intervals with the bootstrap

No result is currently available on the rate of convergence of Sharpe ratio point estimators to normality, and the confidence intervals obtained using the asymptotic variance may have less than nominal coverage. Improved confidence intervals can be obtained using the bootstrap [Was03, Ch. 8], especially the Studentized version [LW08, and references therein].[11] Specifically, let $\bar{\zeta}$ be the original point estimate, and let $\bar{\mathsf{se}}$ be the original estimation of the standard error. For any bootstrap resample, let $\hat{\zeta}$ be the point estimate obtained from that resample, and let $\widehat{\mathsf{se}}$ be the estimate of the standard error obtained from the resample. Define, for each resample,

$$T = \frac{\hat{\zeta} - \bar{\zeta}}{\widehat{\mathsf{se}}} \ .$$

An (approximate) $1 - \alpha$ confidence interval for ζ can be obtained as

$$C = \left(\bar{\zeta} - T_{\alpha/2}\bar{\mathsf{se}}, \bar{\zeta} - T_{1-\alpha/2}\bar{\mathsf{se}}\right),$$ [12]

where T_β is the $1 - \beta$ percentile of the values obtained from the bootstrap resamples.

Particular care must be taken when creating the bootstrap resamples when the original data is not a collection of i.i.d. samples. For example, Ledoit and Wolf [LW08] suggest the use of the circular block bootstrap [PR92]. The issue in using such a method in practice is that it involves fitting a semi-parametric model (e.g., VAR, GARCH) to the observed data, which could be computationally expensive and not theoretically motivated. Moreover, one may argue that the same model could be assumed in the derivation of the asymptotic variance, resulting in better confidence intervals in this case [SS10].

Unhapipat, Chen, and Pal [UCP16] report experimental results showing that, under the assumption of i.i.d. samples, the confidence intervals obtained with the simple (non-Studentized) bootstrap have coverage closer to the nominal than those obtained using the asymptotic normality.

4 Hypothesis testing

In this section we present statistical tests for the Sharpe ratio and for the difference of two Sharpe ratios.

[11] Vinod and Morey [VM99b] also present confidence intervals for the Sharpe ratio based on the bootstrap but Ledoit and Wolf [LW08] observes that their application of the Studentized bootstrap is incorrect.

[12] The "$-$" in the computation of the upper bound is *not* a typo: the value $T_{1-\alpha/2}$ is negative because it is the $\alpha/2$ percentile of the bootstrap distribution which is centered at zero.

4.1 Tests on a single Sharpe ratio

The task requires us to test

$$H_0 \,:\, \zeta = \zeta_0 \quad \text{versus} \quad H_a \,:\, \zeta \neq \zeta_0 \ .$$

It is easy to derive tests using the asymptotic normality of the point estimates [UCP16; Chr05; LW08]. Let $V(\zeta_0)$ be the asymptotic variance computed under H_0 (see (7) and (8)), then H_0 can be rejected with significance level α if

$$\sqrt{n}|\hat{\zeta} - \zeta_0| > z_{\alpha/2}\sqrt{V(\zeta_0)},$$

where $z_{\alpha/2}$ is the $1-\alpha/2$-th -quantile of the standard normal distribution. Not surprisingly, given the use of the asymptotic distribution, this test has low statistical power, i.e., it rejects a true null hypothesis in more than a $1-\alpha$ fraction of the cases [LW08].

When using the bootstrap to compute a confidence interval C, it is possible to reject H_0 with nominal level α if $\zeta_0 \notin C$.

4.2 Tests on the difference of two Sharpe ratios

A number of works [LW08; Chr05; SS10; VM99b; Opd07] study the problem of comparing the Sharpe ratios of two investment strategies (w.r.t. the same benchmark). Specifically, given two Sharpe ratios ζ_1 and ζ_2, the goal is testing the following hypotheses:

$$H_0 \,:\, \Delta = \zeta_1 - \zeta_2 = 0 \quad \text{versus} \quad H_a \,:\, \Delta \neq 0$$

The same approaches discussed for the estimation and testing of a single Sharpe ratio extend to this case,[13] including approaches using the GMM [Chr05; SS10], or the bootstrap [LW08; Opd07; VM99b].

Methodological [LW08, Remarks 3.1 and 3.3] and empirical [AS13] evidence suggests that the bootstrap-based approach by Ledoit and Wolf [LW08] is theoretically sound and outperforms other methods. On the other hand, as mentioned in Sect. 3.3, a correct application of the Studentized circular block bootstrap requires fitting a semi-parametric model, a non-straightforward and potentially computationally expensive operation.

5 Conclusions and research directions

We presented and discussed existing methods for estimation, computation of confidence intervals, and hypothesis testing of the Sharpe ratio. None of the approaches presented in the literature are entirely satisfying. For example, methods deriving or using the asymptotic normality of the estimator are known to perform worse (according to various metrics) than methods based on the bootstrap. On the other hand, a correct application of bootstrap techniques taking into account the time-series nature of the data requires impractical steps (fitting of a semi-parametric model), which may hinder the usefulness of the methods. Ignoring the correlation structure of the data simplifies the application of bootstrap-based methods, but doing so must be a deliberate choice.

Interesting directions for research include the derivation of more stringent confidence intervals (e.g., studying the problem using martingales), and the development of more descriptive expressions for the asymptotic variance when some information on the correlation structure of the data is known (Schmid and Schmidt [SS10] present some results in this direction). Additionally, it would be interesting to study the properties of Challet's estimator [Cha15], given its unconventional approach.

[13] More precisely, many of the approaches were originally motivated by the need of testing the difference of two Sharpe ratios, but are easy to extend (and the authors often do that or at least mention this possibility) to the case of a single Sharpe ratio.

References

[And91] Donald W. K. Andrews. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". In: *Econometrica* 59.3 (1991), pp. 817–858.

[AS13] Benjamin R. Auer and Frank Schuhmacher. "Performance hypothesis testing with the Sharpe ratio: The case of hedge funds". In: *Finance Research Letters* 10.4 (2013), pp. 196–208.

[Bac09] Carl Bacon. *How sharp is the Sharpe ratio? Risk-adjusted Performance Measures.* Technical Report. StatPro, 2009.

[Bao09] Yong Bao. "Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution". In: *Journal of Financial Econometrics* 7.2 (2009), pp. 152–173.

[BU06] Yong Bao and Aman Ullah. "Moments of the estimated Sharpe ratio when the observations are not IID". In: *Finance Research Letters* 3.1 (2006), pp. 49–56.

[Cha15] Damien Challet. "Moment-Free Sharpe Ratio Estimation from Total Drawdown Durations". 2015. URL: `http://ssrn.com/abstract=2603682`.

[Chr05] Steve Christie. *Is the Sharpe ratio useful in asset allocation?* MAFC Research Papers 31. Applied Finance Center, Macquarie University, 2005.

[Han82] Lars Peter Hansen. "Large Sample Properties of Generalized Method of Moments Estimators". In: *Econometrica* 50.4 (1982), pp. 1029–1054.

[Isr05] Craig Israelsen. "A refinement to the Sharpe ratio and information ratio". In: *Journal of Asset Management* 5.6 (2005), pp. 423–427.

[LM02] Andrew W. Lo and A. Craig MacKinlay. *A non-random walk down Wall Street.* Princeton University Press, 2002.

[Lo02] Andrew W. Lo. "The statistics of Sharpe ratios". In: *Financial Analysts Journal* 58.4 (2002), pp. 36–52.

[LW08] Oliver Ledoit and Michael Wolf. "Robust performance hypothesis testing with the Sharpe ratio". In: *Journal of Empirical Finance* 15.5 (2008), pp. 850–859.

[Mer02] Elmar Mertens. "Comments on the correct variance of estimated Sharpe Ratios in Lo (2002, FAJ) when returns are IID". 2002. URL: `http://www.elmarmertens.com/research/discussion/soprano01.pdf`.

[MG78] Robert E. Miller and Adam K. Gehr. "Sample size bias and Sharpe's performance measure: A note". In: *Journal of Financial and Quantitative Analysis* 13.05 (1978), pp. 943–946.

[Opd07] John Douglas Opdyke. "Comparing Sharpe ratios: so where are the p-values?" In: *Journal of Asset Management* 8.5 (2007), pp. 308–336.

[Pav15] Steven E. Pav. "Notes on the Sharpe ratio". 2015. URL: `http://cran.uni-muenster.de/web/packages/SharpeR/vignettes/SharpeRatio.pdf`.

[PR92] Dimitris N. Politis and Joseph P. Romano. "A circular block-resampling procedure for stationary data". In: *Exploring the limits of bootstrap.* Ed. by Raoul LePage and Lynne Billard. Wiley, 1992, pp. 263–270.

[Sch07] Hendrik Scholz. "Refinements to the Sharpe ratio: Comparing alternatives for bear markets". In: *Journal of Asset Management* 7.5 (2007), pp. 347–357.

[Sha65] William F. Sharpe. "Mutual fund performance". In: *Journal of Business* 39.1 (1965), pp. 119–138.

[Sha94] William F. Sharpe. "The Sharpe ratio". In: *The Journal of Portfolio Management* 21.1 (1994), pp. 49–58.

[SS10] Friedrich Schmid and Rafael Schmidt. "Statistical Inference for Sharpe Ratio". In: *Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds*. Ed. by Arjan B. Berkelaar, Joachim Coche, and Ken Nyholm. Palgrave Macmillan UK, 2010, pp. 337–357.

[SS11] Grant H. Skrepnek and Ashok Sahai. "An Estimation Error Corrected Sharpe Ratio Using Bootstrap Resampling". In: *Journal of Applied Finance and Banking* 1.2 (2011), pp. 189–206.

[SS13] Grant H. Skrepnek and Ashok Sahai. "Efficient Point Estimation of the Sharpe Ratio". In: *Journal of Statistical and Econometric Methods* 2.4 (2013), pp. 129–142.

[UCP16] Suntaree Unhapipat, Jun-Yu Chen, and Nabendu Pal. "Small Sample Inferences on the Sharpe Ratio". In: *American Journal of Mathematical and Management Sciences* 35.2 (2016), pp. 105–123.

[VM99a] Hrishikesh D. Vinod and Matthew R. Morey. "A double Sharpe ratio". 1999. URL: `http://ssrn.com/abstract=168748`.

[VM99b] Hrishikesh D. Vinod and Matthew R. Morey. "Confidence intervals and hypothesis testing for the Sharpe and Treynor performance measures: A bootstrap approach". In: *Computational Finance*. The MIT Press, 1999, pp. 25–39.

[Was03] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2003.

[Wei16] Eric W. Weisstein. "k-Statistic". from MathWorld – A Wolfram Web Resource. 2016. URL: `http://mathworld.wolfram.com/k-Statistic.html`.

Appendix A The double Sharpe ratio

Vinod and Morey [VM99a] introduce the *double Sharpe ratio*, a modified version of the basic estimator that takes into account the standard deviation of $\hat{\zeta}_{\text{basic}}$. The double Sharpe ratio is *not* actually an estimator for the Sharpe ratio ζ but rather (an estimator for) an alternative measure of the performance of an investment strategy. We present it here mainly because it attracted our interest.

Definition 3 (Double Sharpe ratio)**.** The *double Sharpe ratio* is defined as

$$\hat{\zeta}_{\text{double}} = \frac{\hat{\zeta}_{\text{basic}}}{\sqrt{\operatorname{Var}[\hat{\zeta}_{\text{basic}}]}} \ . \tag{11}$$

The standard deviation of the basic estimator appearing in the definition of $\hat{\zeta}_{\text{double}}$ is unknown, and in practice it is estimated using the bootstrap procedure [Was03, Ch. 8].

Vinod and Morey [VM99a] do not clarify whether the numerator in (11) should actually be $\hat{\zeta}_{\text{basic}}$ or the *mean of the estimations of the Sharpe ratio created during the bootstrap* (or even the median). The choice may have a practical impact: in the experimental results reported by Vinod and Morey [VM99a], the latter quantity was always (slightly) *larger* than the former. The authors argue that this fact can be explained by the non-normality and skewness properties of the bootstrap sampling distribution of the basic estimator, although they do not explain why such a distribution should be expected, and the argument is not entirely convincing. We hypothesize that this overestimation is due to the convexity of the Sharpe ratio, and would follow from Jensen's inequality. Our hypothesis is inspired by an observation by Christie [Chr05, App. C].

Bao [Bao09] introduces a minor variant of the double Sharpe ratio, which takes the bias of $\hat{\zeta}_{\text{basic}}$ into account in the numerator of (11).