Performance Measurement of Portfolio Management with the Generalized Sharpe Ratio and Economic Performance Measure: Can we find improved measures?

Richard Lu ***(Corresponding author***
Department of Risk Management and Insurance,
Feng Chia University, Taichung, Taiwan

Adrian (Wai-Kong) Cheung
Faculty of Finance, City University of Macau, Macao, SAR, China.

Sardar M. N. Islam
Institute for Sustainable Industries & Liveable Cities (ISILC),
Victoria University, City Flinders Campus, Australia

Abstract

The objective of this paper is to examine the relative performances of the Economic Performance Measure (EPM) and the Generalized Sharpe Ratio (GSR) for measuring portfolio performance. In addition, which estimation method (parametric vs nonparametric) provides us with better results? It is found that both GSR and EPM provide rankings consistent with high-order risk preference, but the SR does not. The parametric GSR and EPM do not always approximate well to the nonparametric ones. The nonparametric GSR and EPM can solve the problems of the Sharpe ratio paradox and the manipulated Sharpe ratio. Furthermore, a comparison between a Buy-and-Hold (BH) strategy and a Constant Proportion Portfolio Insurance (CPPI) strategy reveals that BH performs better in terms of SR. At the same time, CPPI is better in terms of EPM(NP) and GSR(NP), suggesting that the nonparametric GSR and EPM solve the CPPI problem satisfactorily. Thus, GSR and EPM are preferred over the SR for measuring portfolio performance when the data is non-normal or investors have higher-order risk preferences. For estimation purposes, the nonparametric approach is recommended. Therefore, this study has significant academic and practical implications for designing portfolios and investment planning.

Keywords: Sharpe Ratio, Generalized Sharpe Ratio, Economic Performance Measure
JEL codes: D81 G11

1. Introduction

The Sharpe Ratio (SR), defined as the ratio of the excess return to the standard deviation of the return, is a popularly used performance measure of portfolio management. However, the SR has some limitations and shortcomings, which have been extensively discussed in the literature. It is derived from the mean-variance model with the strict assumption of either quadratic preferences or normally distributed returns. When the returns' distribution deviates from normality, it may lead to unreasonable results. Indeed, the challenge of modeling and optimizing portfolios with non-elliptical, asymmetric, and leptokurtic returns is a significant area of modern financial research (Paolella, 2014). For instance, Hodges (1998) demonstrates the Sharpe ratio paradox, which points out that the Sharpe ratio may provide a ranking inconsistent with that of First Order Stochastic Dominance. Furthermore, the Sharpe ratio can be manipulated, so its reliability as a performance measure is doubtful (Goetzmann et al., 2002; Ingersoll et al., 2007; Leland, 1999; Spurgin, 2001). The limitations of the static SR have prompted further research into dynamic and risk-aware extensions. For instance, some models seek to optimize a stochastic Sharpe Ratio under specific constraints, such as portfolio drawdown, to better manage tail risk in a dynamic environment (Biswas et al., 2020). The Sharpe ratio is built up under the mean-variance framework. The mean-VaR model and the mean-CVaR model are also developed. (Guo et al., 2019). Other extensions of the measures include Leung and Wong (2008), Wong et al. (2012), Bai et al. (2011a, 2011b, 2012, 2015), Leung et al. (2012), Ng et al. (2017), Niu et al. (2017, 2018), and Chan et al. (2020).

To go beyond the mean and variance framework and mitigate some of the shortcomings, Hodges (1998) and Zakamouline and Koekebakker (2009) proposed to extend the Sharpe ratio to a new performance measure called the Generalized Sharpe Ratio (GSR), while Homm and Pigorsch (2012) developed another new performance measure called the Economic Performance Measure (EPM), which has been extended by Niu et al. (2018). The former is based on the expected utility theory of portfolio choice. At the same time, the latter replaces the standard deviation of the Sharpe ratio with a risk measure proposed by Aumann and Serrano (2008). Both performance measures are a kind of generalized Sharpe ratio in the following sense: the GSR reduces to the Sharpe ratio, and the EPM ranking is equivalent to the Sharpe ratio ranking, when the return distribution converges to a normal distribution.

To compute the performance measures, both parametric and nonparametric estimation methods are available. By assuming that the return follows a Normal Inverse Gaussian (NIG) distribution, Zakamouline and Koekebakker (2009) and Homm and Pigorsch (2012) each derived a different closed-form formula for the parametric estimators. Another closed-form formula for a parametric estimator has also been derived by Alexander (2008), using the Taylor approximation. All three estimators take into account the first four moments of the distribution in computing the performance measures. The nonparametric estimation is based on numerical methods and takes into account all the moments of the distribution. The parametric methods are easier to compute, but they are only valid to the extent that the parametric assumptions can capture or approximate the key features of the distribution. Little is known about the performance of these estimation methods.

The literature suggests four interesting cases of return distributions, where the Sharpe ratio is known to have failed to give sufficiently accurate measures of performance. The first is the distributions with which the Sharpe ratio paradox was first shown in Hodges (1998). The second is the distributions constructed in Zakamouline and Koekebakker (2009) to show how the Sharpe ratio can be manipulated. The third one is the return distributions generated by the Constant Proportion Portfolio Insurance (CPPI) strategy (Dichtl & Drobetz, 2011).[1] This strategy can effectively cut off the lower tail of the return's distributions, resulting in a lower standard deviation and thus a lower Sharpe ratio. Finally, high-order risk preference, especially prudence and temperance, is also considered because the Sharpe ratio only considers the first two moments of return distributions, but higher-order moments may affect preference for prudence and/or temperance. It is interesting to understand whether GSR and EPM can perform well in these four cases.

This paper aims to examine the relative performances of the EPM and the GSR for measuring portfolio performance. In particular, we study the performance of these two measures in the four interesting cases where the Sharpe ratio may not deliver. In

[1] Note that Lu and Hsiung (2018) examine the performance of CPPI strategy using the EPM. We also use the GSR.

addition, we also check for the performance of the parametric and nonparametric estimation methods.

When it comes to the choice of an estimation method, the evidence always supports the use of a nonparametric method because the parametric estimation methods do not approximate well. For example, in the paradox case, the parametric methods give different rankings from those of the nonparametric method. The parametric estimator of the GSR by Alexander (2008) approximates well in the first case but not in the second. Thus, although the parametric estimation methods are easy to compute, we have to be cautious about their poor approximations.

Both GSR and EPM provide rankings consistent with high-order risk preference, but the SR does not. In addition, the (nonparametric) GSR and EPM perform well in resolving the Sharpe paradox, the manipulated Sharpe ratio problem, and the CPPI problem.

The remainder of this paper is organized as follows: Section 2 reviews the two methodologies for extending the Sharpe ratio. In Section 3, the new performance measures are applied in the four cases. Section 4 provides an empirical application. The final section is the conclusion.

2. The Generalized Sharpe Ratio and Economic Performance Measure

Two methodologies that extend the Sharpe ratio are discussed here. One methodology is based on the expected utility theory of portfolio choice, while the other uses the reward-to-risk framework.

2.1 The GSR

Markowitz's portfolio theory, which is based on a mean-variance model, indicates that investors always choose the optimal risky portfolio, with the highest Sharpe ratio, within a feasible set. Thus, the Sharpe ratio is a natural definition of a performance measure. Also, the mean-variance model can be set up by assuming that the investor has negative exponential utility and that the risky asset returns are normally distributed. Hodges (1998) points out that in the mean-variance model with one free asset and one risky portfolio, an investor's expected optimal utility is given by:

$$\mathrm{E}(U^*) = -e^{-\frac{\mathrm{SR}^2}{2}}, \tag{1}$$

where SR is the Sharpe ratio of the risky portfolio. Thus, the higher the Sharpe ratio of the risky portfolio is, the higher the level of the expected optimal utility that an investor can receive.

To go beyond the mean-variance model and to generalize the Sharpe ratio, Hodges (1998) conjectured that the relationship between an investor's expected optimal utility and the Generalized Sharpe Ratio (GSR) is:

$$\mathrm{E}(U^*) = -e^{-\frac{\mathrm{GSR}^2}{2}}, \tag{2}$$

for any return distribution of the risky portfolio. Rearranging the equation for GSR, the GSR can be computed using:

$$\mathrm{GSR} = \sqrt{-2\log\left(-\mathrm{E}(U^*)\right)}, \tag{3}$$

by first finding out the expected optimal utility.

Nonparametric Estimation of GSR (GSR(NP))

To determine the expected optimal utility for any return distribution, Zakamouline and Koekebakker (2009) suggested a numerical method of solving the following maximization problem:

$$\max_{a}\ \mathrm{E}\left(-e^{-\lambda a(r_p - r_f)}\right), \tag{4}$$

where "a" is the decision variable, λ is the parameter of the negative exponential utility function, r_p is the rate of return of the risky portfolio, and r_f is the risk-free rate. This method of finding the expected optimal utility and calculating the GSR is referred to as the nonparametric estimation (i.e., GSR(NP)).

Parametric Estimation of GSR (GSR(P))

To apply the GSR without using the numerical method, Zakamouline and Koekebakker (2009) derived a parametric formula by assuming that the portfolio return follows a Normal Inverse Gaussian (NIG) distribution:

$$f(r_p;\ \alpha,\beta,\eta,\delta) = \frac{\alpha\delta e^{\delta\varphi+\beta(r_p-\eta)}}{\pi\sqrt{\delta^2+(r_p-\eta)^2}} K_1(\alpha\sqrt{\delta^2+(r_p-\eta)^2}, \tag{5}$$

where

$$\varphi = \sqrt{\alpha^2-\beta^2}; \tag{6}$$

$$K_1(x) = \frac{1}{2}\int_0^{\infty} e^{-\left(\frac{1}{2}\right)x(z+z^{-1})} dz\,, \tag{7}$$

and $\alpha,\ \beta,\ \eta,$ and δ are parameters. These parameters are linked to the mean, variance, skewness and kurtosis of the NIG distribution as follows:

$$\mathrm{E}(r_p) = \mu = \eta + \delta\frac{\beta}{\varphi},\ \mathrm{Var}(r_p) = \sigma^2 = \delta\frac{\alpha^2}{\varphi^3}, \tag{8}$$

$$\mathrm{Skew}(r_p) = \chi = 3\frac{\beta}{\alpha\sqrt{\delta\varphi}}, \mathrm{Kurt}(r_p) = \kappa = 3 + \frac{3}{\delta\varphi}(1+4(\frac{\beta}{\alpha})^2). \tag{9}$$

From these equations, we can derive:

$$\alpha = \frac{3\sqrt{3\kappa-4\chi^2-9}}{\sigma^2(3\kappa-5\chi^2-9)}, \beta = \frac{3\chi}{\sigma(3\kappa-5\chi^2-9)}, \eta = \mu - \frac{3\chi\sigma}{(3\kappa-4\chi^2-9)}, \text{ and } \delta = \frac{3\sigma^2\sqrt{3\kappa-5\chi^2-9}}{(3\kappa-4\chi^2-9)}. \tag{10}$$

However, to get meaningful parameters α and δ, the following condition must be satisfied:

$$\kappa > 3 + \frac{5}{3}\chi^2. \tag{11}$$

The formula of the parametric estimation of GSR is given by:

$$\sqrt{2\left(\lambda a^*(\eta - r_f) - \delta\left(\varphi - \sqrt{\alpha^2-(\beta-\lambda a^*)^2}\right)\right)}, \tag{12}$$

where

$$a^* = \frac{1}{\lambda}\left(\beta + \frac{\alpha(\eta - r_f)}{\sqrt{\delta^2+(\eta-r_f)^2}}\right). \tag{13}$$

Thus, to implement the parametric estimation, we simply estimate the first four moments of the return's distribution and use them to compute the four parameters of the NIG's distribution and the GSR. This estimation method is called GSR(P-Z).

Another parametric method to estimate the GSR has been developed by Alexander (2008), and is called GSR(P-A). She uses Taylor's expansion of the expected utility to get the certainty equivalent of the portfolio investment and obtain the maximum expected utility by using Pézier's (2008) approximating method. The formula is shown as follows:

$$\sqrt{\left(\frac{\mu - r_f}{\sigma}\right)^2 + \frac{\chi\left(\frac{\mu - r_f}{\sigma}\right)^3}{3} - \frac{(\kappa - 3)\left(\frac{\mu - r_f}{\sigma}\right)^4}{12}}. \tag{14}$$

This formula collapses to the Sharpe ratio when the skewness equals 0 and the kurtosis equals 3. It also indicates that negative skewness and/or high kurtosis have the effect of decreasing the GSR.

2.2 The Aumann and Serrano (2008) index of riskiness and the EPM

The Sharpe ratio is just a reward-to-risk performance measure. However, it is derived from the mean-variance framework. There, the standard deviation is a legitimate risk measure. However, the standard deviation is not a good measure under general conditions. Homm and Pigorsch (2012) proposed a new performance measure by replacing the standard deviation with an index of riskiness, proposed by Aumann and Serrano (2008) (hereafter called the AS index). This performance measure is referred to as the Economic Performance Measure (EPM).

The AS index is derived from being based on two key axioms: Duality and positive homogeneity. The duality requires a risk index that reflects the way less risk-averse individuals accept riskier assets. Thus, it satisfies monotonicity with respect to the Second-order Stochastic Dominance (SSD). This is an important property for risk measurement. If portfolio A dominates portfolio B in terms of its SSD, and we know that all the risk-averse investors prefer A to B, a risk measure with monotonicity will indicate that portfolio A is less risky than portfolio B. As risk measures, standard deviation, semi-standard deviation, value at risk, and expected shortfall all violate the monotonicity property.

Aumann and Serrano (2008) defined the economic index of riskiness for a risky asset as the reciprocal of the positive risk aversion parameter of an individual with Constant Absolute Risk Aversion (CARA) indifferent about taking or not taking the risky asset. Under their setup, the AS index must satisfy the following equation:

$$\mathrm{E}U(W + S_t - S_0) = U(W), \tag{15}$$

where U is the utility function, W is the initial wealth, S_t is the price of the risky asset at time *t*. Assuming no cash dividend, $S_t - S_0$ is the absolute return from holding the asset for the time interval. Aumann and Serrano (2008) constructed the index of riskiness by using an exponential utility function. Thus, the AS index of the risky asset, $\mathrm{AS}(S_t)$ is defined implicitly as follows:

$$\mathrm{E}e^{-(S_t - S_0)/\mathrm{AS}(S_t)} = 1. \tag{16}$$

They proved that the $\mathrm{AS}(S_t)$ is a unique positive number, and any index satisfying the two axioms will be a positive multiple of the $\mathrm{AS}(S_t)$ if some of the absolute returns are negative, and the mean of the absolute return is positive.

Under this setup, the investment risk is a kind of additive risk. However, if the individual places her initial wealth in the risky asset, the risk becomes multiplicative. For a multiplicative risk, similar to Aumann and Serrano's approach (2008), Schreiber (2014) defined an economic index of relative riskiness for a risky asset as the reciprocal of the positive risk aversion parameter of an individual with Constant Relative Risk Aversion (CRRA) who is indifferent about taking or not taking the risky asset. Under this setup, the index of relative riskiness must satisfy the following equation:

$$\mathrm{E}U(W(S_t/S_0)) = U(W). \tag{17}$$

Schreiber (2014) adopted a power utility and derived the index of relative riskiness which, in fact, equals the AS index applied to the log return instead of the absolute return. That is, the index of relative riskiness, $\mathrm{RS}(S_t)$ is defined implicitly as follows:

$$\mathrm{E}e^{-(\ln S_t - \ln S_0)/\mathrm{RS}(S_t)} = 1. \tag{18}$$

The index of relative riskiness also satisfies positive homogeneity in the sense of $\mathrm{RS}(S_1^t) = t\mathrm{RS}(S_1)$, which states t-year investment risk equals the one-year investment risk repeated t times. In addition, if an investor wants to take out the time value involved, $\mathrm{RS}(S_t)$ is defined implicitly as:

$$\mathrm{E}e^{-(\ln S_t - \ln S_0 - r_f t)/\mathrm{RS}(S_t)} = 1. \quad (19)$$

Thus, to measure the relative riskiness of a risky asset, the excess log return formula should be displayed. This idea has been applied to measure the risk effect of time diversification in Lu et al. (2018). Similarly, Lu and Hsieh (2019) employed the EPM to correctly evaluate the performance of dollar-cost averaging strategies, which are known to generate non-normal return distributions with negative skewness and excess kurtosis that would be misevaluated by traditional measures. Furthermore, Lu et al. (2022) applied the Sharpe ratio and EPM to examine the corporate governance and stock performance relationships in the Taiwan stock market from a portfolio performance perspective, finding that well-governed portfolios (such as the TWSE Corporate Governance 100 Index) outperform benchmarks with higher returns, lower volatility, and superior downside risk protection, highlighting the practical value of these measures in investment planning under non-normal distributions.

The EPM is defined as:

$$\mathrm{EPM} = \frac{\mathrm{E}(\tilde{r})}{\mathrm{AS}(\tilde{r})}, \quad (20)$$

where $\mathrm{E}(\tilde{r})$ is the expected excess return of an investment portfolio, and $\mathrm{AS}(\tilde{r})$ is the AS index of the random excess return. Homm and Pigorsch (2012) demonstrated that the EPM has the positive property of monotonicity with respect to the first-order and second-order stochastic dominance.

The EPM can also be regarded as a generalized Sharpe measure because the EPM equals two times the squared Sharpe ratio under normality. Thus, the EPM is equivalent to the Sharpe measure ranking under normality.

Nonparametric Estimation of EPM (EPM(NP))

In applying the EPM, the subtle element is to calculate the AS index. According to Aumann and Serrano (2008), the index can be obtained by solving $\mathrm{AS}(\tilde{r})$ in the equation:

$$\mathrm{E}\left(e^{-\frac{\tilde{r}}{\mathrm{AS}(\tilde{r})}}\right) = 1. \quad (21)$$

Given any distribution function, this equation can be applied and solved for the AS index. This is a nonparametric estimation method. The EPM estimated by this method is referred to as the EPM(NP).

Parametric Estimation of EPM (EPM(P))

By also assuming a normal inverse Gaussian distribution, Homm and Pigorsch (2012) derived the AS index and the EPM as follows:

$$\mathrm{AS} = \frac{3\kappa(\mu - r_f) - 4(\mu - r_f)\chi^2 - 6\chi\sigma + \frac{9\sigma^2}{(\mu - r_f)}}{18}, \tag{22}$$

and

$$\mathrm{EPM} = \frac{18(\mu - r_f)}{3\kappa(\mu - r_f) - 4(\mu - r_f)\chi^2 - 6\chi\sigma + \frac{9\sigma^2}{(\mu - r_f)}}. \tag{23}$$

This is a parametric estimation method and is called EPM(P). The formula clearly indicates that positive skewness increases the performance measure, but high kurtosis decreases it.

3. Performance evaluation using the GSR and EPM

Both the parametric and the nonparametric methods have advantages and disadvantages. For example, the parametric methods are easier to compute than the nonparametric ones because the latter use numerical estimation methods. However, the parametric methods provide good approximations only to the extent that the parametric assumptions are valid. In particular, the three parametric estimation methods only account for the first four moments of the distribution and ignore information from the higher moments, which the nonparametric methods can account for. In this section, both the parametric and the nonparametric estimation methods are applied to situations where the Sharpe ratio has failed to give a good performance measure.

3.1 The Sharpe ratio paradox case

To demonstrate the shortcomings of the Sharpe ratio, Hodges (1998) used an example of two assets, A and B, with excess return distributions as shown in Table 1.

Table 1. Probability distributions of assets A and B

Probability	0.01	0.04	0.25	0.4	0.25	0.04	0.01
A's return	-25%	-15%	-5%	5%	15%	25%	35%
B's return	-25%	-15%	-5%	5%	15%	25%	45%

Here, apart from asset A having an excess return of 35% with a 1% chance, while asset B has an excess return of 45% with the same chance, both distributions are identical. Therefore, the return distribution of asset B's first-order stochastically dominates asset A. Thus, everyone with an increasing utility function prefers asset B to asset A. However, asset A has a Sharpe ratio of 0.5, which is higher than the Sharpe ratio of asset B, at 0.493. According to the Sharpe ratio, asset A should outperform asset B. This is the so-called Sharpe ratio paradox.

Table 2. Performance measures of the two assets in the Sharpe ratio paradox

Asset	**SR**	**GSR(P-A)**	**GSR(NP)**	**EPM(P)**	**EPM(NP)**
A	0.500	0.498	0.498	0.300	0.455
B	0.493	0.500	0.499	0.150	0.495

As shown by Zakamouline and Koekebakker (2009), using GSR(P-Z) and GSR(NP) can resolve the paradox. Here, GSR(P-A), EPM(P), and EPM(NP) are verified to see if they can also resolve this paradox.[2] Table 2 shows the estimated results where GSR(NP) is also included for comparison. According to the GSR(P-A) or EPM(NP), asset B is higher than asset A. Thus, they can also resolve the paradox. Also, the value of GSR(P-A) approximately equals the GSR(NP). However, the EPM(P) does not approximate well to the EPM(NP). The EPM(P) gives the ranking as being opposite to that of the EPM(NP). Thus, EPM(P) fails to resolve the paradox.

3.2 The manipulated Sharpe ratio case

The Sharpe ratio can be manipulated; it can be increased by accompanying it with large negative skewness and/or positive kurtosis. For example, Leland (1999) and Spurgin (2001) showed that it can be increased by selling off the upper tail of the return's distribution; Goetzmann et al. (2002) demonstrated that the Sharpe ratio of a stock

[2]As discussed in Section 4, GSR(P-Z) is not examined because of its poor empirical performance.

portfolio can be increased by selling out-of-the-money call and/or put options on the portfolio.

Based on Goetzmann et al. (2002), Zakamouline and Koekebakker (2009) conducted a case study on the manipulated Sharpe ratio. They assumed that the price of a stock follows the geometric Brownian motion, with the drift and volatility parameters both equal to 15%. Their manipulation strategy consisted of holding the stock, plus selling 2.58 puts when the strike price is 12% lower than the current stock price, and selling 0.77 calls when the strike price is 12% higher than the current stock price. Under the Black-Scholes model, and a risk-free rate of 5%, they simulated one million one-year stock returns as the returns of the benchmark strategy and the returns of the manipulation strategy. They show that both GSR(P-Z) and GSR(NP) can solve the manipulation problem for these performance measures, that is, the benchmark strategy outperforms the manipulation strategy.

By the same simulation design, the returns of the two strategies are simulated again. The returns' statistics and the performance measures are listed in Table 3.[3]

Table 3. Returns' statistics and performances of the Benchmark (B) and Manipulated (M) strategies

Strategy	Mean	Std	Skew	Kurt	SR	GSR(P-A)	GSR(NP)	EPM(P)	EPM(NP)
B	0.162	0.175	0.441	3.240	0.639	0.665	0.672	0.764	1.004
M	0.141	0.120	-2.048	10.25	0.764	0.271	0.622	0.071	0.551

The statistics' results are similar to those of Zakamouline and Koekebakker (2009). The manipulation strategy has a lower mean and standard deviation but large negative skewness and positive kurtosis. By the Sharpe ratio, the manipulation strategy is better than the benchmark strategy. However, according to the GSR and the EPM, the benchmark strategy outperforms the manipulation strategy. Both parametric and nonparametric estimation methods produce the same ranking, but both the parametric methods, the GSR(P-A) and the EPM(P), do not approximate well to the nonparametric methods in the case of the manipulation strategy.

[3]We exclude GSR(P-Z) from the analysis because of its poor empirical performance as discussed in Section 4.

3.3 The CPPI case

Portfolio insurance is a strategy to cut off the lower tail of the return's distribution. It can be done by shorting a put option. Thus, this strategy may reduce the mean and standard deviations in such a way that the Sharpe ratio is lower than the benchmark strategy, i.e., a Buy-and-Hold (BH) strategy (Dichtl & Drobetz, 2011). Thus, according to the Sharpe ratio, portfolio insurance performs worse than the benchmark strategy. However, one might get different results with the GSR or EPM because portfolio insurance also generates non-normalities.

Following the same simulation design in Dichtl and Drobetz (2011), both strategies' one-year log return distributions are generated under the normal-return and normal-volatility market scenarios. To implement the portfolio insurance strategy, a Constant Proportion Portfolio Insurance (CPPI) strategy is adopted.[4] The multiplier is set to five, and the floor is 100% of the initial amount invested. With the round-trip transaction costs at 0.1%, the asset allocation is reset when the stock price moves (up or down) by over 2%. The stock price also follows the geometric Brownian motion, with the drift being 11.5% and the volatility 20%. The risk-free rate is 4.5%. All these numbers are the same as those from Dichtl and Drobetz (2011).

Table 4. Returns statistics and performances of the BH and CPPI strategies

Strategy	Mean	Std	Skew	Kurt	SR	GSR(P-A)	GSR(NP)	EPM(P)	EPM(NP)
BH	0.095	0.201	-0.026	2.965	0.250	0.250	0.250	0.126	0.125
CPPI	0.060	0.064	2.797	13.74	0.226	0.244	0.256	0.008	0.171

Table 4 reports the returns' statistics and the performance measures. As expected, the skewness and kurtosis of the BH are very close to a normal distribution, but the CPPI has a large positive skewness and kurtosis. According to the Sharpe ratio, the BH is preferred because it is higher than the CPPI. The parametric measures, such as GSR(P-A) and EPM(P), also indicate that the BH is higher. Both the parametric methods approximate well in the BH case, but not in the CPPI case. The application of the parametric GSR(P-Z) does not lead to any result because the condition for the

[4] See Zieling et al. (2014) for other portfolio insurance strategies.

meaningful parameters (i.e., $\kappa > 3 + \frac{5}{3}\chi^2$) is violated, and therefore no meaningful estimate is available for it. As the parametric methods do not approximate well, both of the nonparametric measures, GSR(NP) and EPM(NP), which do not require any assumption on the distribution function, indicate the opposite results, i.e., the CPPI is higher.

3.4 Prudence and Temperance cases

In this subsection, we evaluate the performance of the lottery pairs designed by Deck and Schlesinger (2010) to know if the GSR and EPM are consistent with the preferences of prudence and temperance.

Prudence and temperance are two high-order risk preferences against negative skewness and high kurtosis, respectively. Eeckhoudt and Schlesinger (2006) define both concepts as preferences over simple 50-50 lottery pairs. They consider prudence and temperance as preferences for disaggregating the harms. Thus, a prudent decision maker prefers a zero-mean lottery attached to a higher wealth state instead of a lower wealth state, while a temperate decision maker prefers two zero-mean lotteries being separated in two different states instead of putting them together in the same state.

For the prudence case, we use the third task in Deck and Schlesinger (2010)'s version 1 experiment. In this task, a subject was endowed with \$12.5, and then she could choose to receive an additional \$1 and a zero-mean lottery of receiving +/-\$5 either in the same state or in two different states. Suppose she chooses the two items in the same state. In that case, she will end up with a portfolio of \$12.5 and the two items with possible returns of 0%, 48% and -36% and associated probabilities of 50%, 25% and 25%, respectively. Alternatively, if she chooses to put the two items in two different states, she will get possible returns of 8%, 40% and -40% and associated probabilities of 50%, 25% and 25%, respectively. In other words, each of these two strategies ends up with a different probability distribution.

Table 5. Returns Statistics and Performances of the Prudence case

Return Type	Mean	Std	Skew	Kurt	SR	GSR(P-A)	GSR(NP)	EPM(P)	EPM(NP)
Preferred	0.040	0.286	0.412	2.038	0.140	0.142	0.142	0.049	0.041
Not preferred	0.040	0.286	-0.412	2.038	0.140	0.139	0.139	0.045	0.038

Table 5 reports the returns summary statistics and the performance measures for the prudence case. As far as the first four moments are concerned, the only difference between the two distributions is in the skewness. The distribution preferred by the prudent decision maker exhibits a positive skewness, while the one not preferred displays a negative skewness. However, as shown in Table 5, the SR cannot detect any difference between the two distributions because it uses information only from the first two moments, while the GSR and EPM can tell that the preferred distribution outperforms.

For the temperance case, we use the second task in the version 1 experiment of Deck and Schlesinger (2010) with some adjustments. For this task, a subject was endowed with \$15, and then she could choose to receive two zero-mean lotteries of receiving +/- \$5 either in the same state or in two different states. To ensure that the mean return is positive, we add one more dollar to each state. Thus, if she likes to put the two items in the different states, she will end up with a portfolio of \$15 and an uncertain outcome with 40% and -26.67% returns of equal probability. Alternatively, if she chooses to put the two items in the same state, she will get possible returns of 6.67%, 73.33% and 60% with associated probabilities of 75%, 12.5% and 12.5%, respectively.

Table 6. Returns Statistics and Performances of the Temperance case

Return Type	Mean	Std	Skew	Kurt	SR	GSR(P-A)	GSR(NP)	EPM(P)	EPM(NP)
Preferred	0.067	0.333	0.000	1.000	0.200	0.201	0.201	0.133	0.089
Not preferred	0.067	0.333	0.000	4.000	0.200	0.200	0.200	0.067	0.076

Table 6 reports the returns summary statistics and the performance measures for the temperance case. The first four moments of the two distributions are exactly the same except for kurtosis. The distribution preferred by the prudence has a lower kurtosis and skewness. Again, the SR cannot tell any difference between the two distributions, while the GSR and EPM reveal that the preferred distribution outperforms.

4. An empirical application: Performance evaluation of hedge funds

In this section, we use the SR, GSR, and EPM to evaluate the performance of hedge funds. Both the parametric and the nonparametric methods of the GSR and the EPM are employed. The return data of hedge funds is similar to the data used in Zakamouline and Koekebakker (2009). In particular, we use the monthly returns of 14 Credit Suisse (CS) hedge fund indexes (including the main index), which track almost every major management style of hedge fund. One index is excluded because of its negative mean return. The sample period is from April 1994 to February 2016. The three-month US T-bill rate is used as the proxy of the risk-free rate. All performance measures calculated here are based on the excess monthly returns[5].

Table 7. Summary statistics of hedge fund indexes (Source: Credit Suisse hedge fund indexes)

Hedge fund index	Mean	Std	Skewness	Kurtosis
All Hedge Index	0.0047	0.0199	-0.2283	6.0692
Convertible Arbitrage	0.0055	0.0188	-2.5988	19.7375
Emerging Markets	0.0061	0.0394	-0.8218	9.2178
Equity Market Neutral	0.0042	0.0276	-12.3065	181.7022
Event Driven	0.0067	0.0178	-2.0298	12.2631
Event Driven Distressed	0.0076	0.0181	-2.0520	13.7812
Event Driven Multi-Strategy	0.0063	0.0194	-1.5891	9.3891
Event Driven Risk Arbitrage	0.0047	0.0116	-0.8812	7.2374
Fixed Income Arbitrage	0.0044	0.0152	-4.7099	38.8022
Global Macro	0.0090	0.0258	0.1629	7.6701
Long/Short Equity	0.0077	0.0268	0.0027	6.7302
Managed Futures	0.0051	0.0336	0.0174	2.8427
Multi-Strategy	0.0063	0.0145	-1.7001	9.5005

[5]All data are from DataStream and Credit Suisse Hedge Index LLC.

Table 7 reports the summary statistics of the hedge funds' index returns. As we can see from the table, 10 out of 13 indices exhibit negative skewness, and only one index does not have excess kurtosis. Thus, the sample resembles reality, as it can capture some stylized facts of the hedge funds' returns as reported in the literature.

Table 8. Performance measures of hedge fund indexes

`	SR	EPM(P)	EPM(NP)	GSR(NP)	GSR(P-A)	GSR(P-Z)
All Hedge Index	0.2384	0.0085	0.1050	0.2352	0.2345	2.2028
Convertible Arbitrage	0.2953	0.0020	0.1084	0.2619	0.2330	5.2101
Emerging Markets	0.1539	0.0052	0.0423	0.1503	0.1497	1.5412
Equity Market Neutral	0.1525	0.0001	0.0222	0.1233	0.0254	NA
Event Driven	0.3745	0.0042	0.1745	0.3328	0.2992	8.2874
Event Driven Distressed	0.4187	0.0042	0.2024	0.3652	0.3122	6.6764
Event Driven Multi-Strategy	0.3229	0.0057	0.1478	0.2972	0.2840	6.9945
Event Driven Risk Arbitrage	0.4091	0.0066	0.2428	0.3814	0.3707	6.3614
Fixed Income Arbitrage	0.2872	0.0007	0.0858	0.2404	0.1581	NA
Global Macro	0.3468	0.0110	0.2131	0.3427	0.3419	1.9138
Long/Short Equity	0.2867	0.0115	0.1515	0.2839	0.2831	1.7267
Managed Futures	0.1509	0.0598	0.0458	0.1513	0.1510	NA
Multi-Strategy	0.4377	0.0057	0.2449	0.3904	0.3524	10.9093

Notes: SR stands for the Sharpe Ratio. GSR(NP) and GSR(P) are the nonparametric and parametric estimates of the Generalized Sharpe Ratio (GSR), respectively. EPM(P) and EPM(NP) are the parametric and nonparametric estimates of the economic performance measure, respectively.

Table 8 shows different estimates for the performance measures of the hedge fund indices. It is obvious that, compared to other performance measures, the GSR(P-Z) tends to produce relatively large numbers. In three indexes, it does not lead to any result because the conditions for meaningful parameters are violated. Thus, we exclude the GSR(P-Z) from our further analysis. The EPM(P) does not approximate well to the EPM(NP) either. Except for one index, the EPM(P) gives relatively smaller numbers than those of the EPM(NP).

Table 9. Rankings of hedge fund indexes by different measures and methods

Hedge fund index	SR	EPM(NP)	GSR(NP)	GSR(P-A)	EPM(P)
All Hedge Index	10	9	10	8	4
Convertible Arbitrage	7	8	8	9	11
Emerging Markets	11	12	12	12	8
Equity Market Neutral	12	13	13	13	13
Event Driven	4	5	5	5	9
Event Driven Distressed	2	4	3	4	10
Event Driven Multi-Strategy	6	7	6	6	7
Event Driven Risk Arbitrage	3	2	2	1	5
Fixed Income Arbitrage	8	10	9	10	12
Global Macro	5	3	4	3	3
Long/Short Equity	9	6	7	7	2
Managed Futures	13	11	11	11	1
Multi-Strategy	1	1	1	2	6

Notes: SR stands for the Sharpe Ratio. GSR(P-A) is the parametric estimate of the generalized Sharpe ratio (GSR), proposed by Alexander (2008). GSR(NP) is the nonparametric estimate of GSR. EPM(P) and EPM(NP) are the parametric and nonparametric estimates of the economic performance measure, respectively.

Table 9 reports the rankings based on the SR, EPM(NP), GSR(NP), GSR(P-A), and EPM(P), respectively. Obviously, the EPM(P) produces very different rankings from those of the others because of its poor approximation. Only the GSR(P-A) approximates the GSR(NP) well for most cases. Thus, we suggest that, for performance evaluation, the GSR(P-Z) and EPM(P) should be used with caution.

Table 10. Rank correlation for the rankings of hedge fund indexes

	SR	EPM(NP)	GSR(NP)	GSR(P-A)	EPM(P)
SR	1				
EPM(NP)	0.7700	1			
GSR(NP)	0.8500	0.9200	1		
GSR(P-A)	0.7400	0.9200	0.9000	1	
EPM(P)	-0.0300	0.21	0.13	0.23	1

Notes: SR stands for the Sharpe Ratio. GSR(P-A) is the parametric estimate of the Generalized Sharpe Ratio (GSR), proposed by Alexander (2008). GSR(NP) is the nonparametric estimate of GSR. EPM(P) and EPM(NP) are the parametric and nonparametric estimates of the economic performance measure, respectively.

Finally, we list the rank correlation (Kendall's τ) for the rankings in Table 10. Table 10 shows that there is a relatively high correlation (> 0.9) among the EPM(NP), GSR(NP), and GSR(P-A). However, the SR is a little bit less correlated with the above three, as the rank correlation coefficients are 0.77, 0.85, and 0.74, respectively. The EPM(P) has the lowest rank correlation with the other measures.

5. Conclusion

This paper studies the performance measurement of two new performance measures, the GSR and EPM, which extend the Sharpe ratio. These two measures go beyond the mean-variance framework. Thus, they are particularly useful for evaluating the return distributions that deviate from normal distributions. There are parametric and nonparametric estimation methods for computing the GSR and EPM. The results show that both measures can mitigate the shortcomings of the Sharpe ratio by using the nonparametric methods. However, although the parametric methods, which approximate the nonparametric estimations by using the first four moments of the distributions, have the advantage of easy computation, they sometimes do not approximate well and might produce different performance rankings.

We find that both GSR and EPM provide rankings consistent with high-order risk preference, but the SR does not. The parametric GSR and EPM do not always approximate well to the nonparametric ones. The nonparametric GSR and EPM can solve the problems of the Sharpe ratio paradox and manipulate the Sharpe ratio. Furthermore, a comparison between a Buy-and-Hold (BH) strategy and a Constant Proportion Portfolio Insurance (CPPI) strategy reveals that BH performs better in terms of SR, while CPPI is better in terms of EPM(NP) and GSR(NP), suggesting that the nonparametric GSR and EPM solve the CPPI problem satisfactorily.

The findings imply that both the GSR and EPM are better performance measures than the SR because they provide rankings consistent with high-order risk preference, and resolve the Sharpe ratio paradox, the manipulated Sharpe ratio problem, and the CPPI problem. This suggests that the standard deviation is not a good risk measure for portfolio insurance (PI) because PI creates a return distribution that is far from a normal distribution. Finally, for computing both GSR and EPM, the nonparametric methods are recommended because the parametric estimation methods sometimes do not approximate well to the nonparametric methods.

The findings suggest that the Sharpe ratio can give the investors a misleading ranking if the data is not normal or the investor has a utility function whose arguments go beyond expectation and variance of returns. Hence, it can be said that whenever the return data is skewed or shows excess kurtosis or investors exhibit high-order risk preferences, it is not advisable to use the Sharpe ratio for evaluating portfolio performance. Alternative performance measures like GSR and EPM that allow for higher moments and/or higher-order risk preferences are recommended in this situation. In addition, the findings also show that the estimation method matters; in general, nonparametric methods are recommended, probably due to their robustness property. The findings provide some valuable insights into the controversy about SR and the relative merits of SR, GSR, and EPM for evaluating and planning portfolio management.

Portfolio performance is admittedly a multi-dimensional concept; GSR and EPM account for two of the dimensions only, i.e., higher moments and higher-order risk preferences.[6] These dimensions can extend even further, encompassing the underlying drivers of firm-level performance, such as corporate governance and ownership structure, which are themselves critical areas of study within financial decision sciences (Chan & Chu, 2022). Future research may cover more performance measures and evaluate their performance along other dimensions.

The contributions of this paper are that the findings of the paper can be generalized for further work in this area of finance by others, in an important area of finance. The results of the study provide some useful insights into the controversy about SR and the relative merits of SR, GSR, and EPM for evaluating and planning portfolio management. Academics and practitioners can apply the theory developed in this paper to many areas, including stock markets (Wong et al., 2005), ADR (Aquino et al., 2005) capital markets (Shrestha et al., 2007), futures markets (Lean et al., 2010), international markets (Lam et al., 2008), and many others.

---

[6] See Cogneau and Hübner (2009a, 2009b) for other dimensions of portfolio performance and performance measures..

References

Alexander, C. (2008). *Quantitative methods in finance*. John Wiley & Sons.

Aquino, K. P., Poshakwale, S., & Wong, W. K. (2005). Is it still worth diversifying internationally with ADRs? *International Journal of Finance, 17*(3), 3622–3643.

Aumann, R. J., & Serrano, R. (2008). An economic index of riskiness. *Journal of Political Economy, 116*(5), 810–835. https://doi.org/10.1086/591947

Bai, Z., Hui, Y., Wong, W. K., & Zitikis, R. (2012). Prospect performance evaluation: Making a case for a non-asymptotic UMPU test. *Journal of Financial Econometrics, 10*(4), 703–732. https://doi.org/10.1093/jjfinec/nbs008

Bai, Z., Li, H., Liu, H., & Wong, W. K. (2011a). Test statistics for prospect and Markowitz stochastic dominances with applications. *The Econometrics Journal, 14*(2), 278–303. https://doi.org/10.1111/j.1368-423X.2010.00337.x

Bai, Z., Li, H., McAleer, M., & Wong, W. K. (2015). Stochastic dominance statistics for risk averters and risk seekers: An analysis of stock preferences for USA and China. *Quantitative Finance, 15*(5), 889–900. https://doi.org/10.1080/14697688.2014.943273

Bai, Z., Wang, K., & Wong, W. K. (2011b). The mean–variance ratio test—A complement to the coefficient of variation test and the Sharpe ratio test. *Statistics & Probability Letters, 81*(8), 1078–1085. https://doi.org/10.1016/j.spl.2011.03.002

Biswas, S., Jawaid, S., & Mukherjee, D. (2020). Multi-asset portfolio optimization with stochastic Sharpe ratio under drawdown constraint. *Annals of Financial Economics, 15*(1), Article 2080001. https://doi.org/10.1142/S201049522080001X

Chan, R. H., Clark, E., Guo, X., & Wong, W. K. (2020). New development on the third-order stochastic dominance for risk-averse and risk-seeking investors with application in risk management. *Risk Management, 22*(2), 108–132. https://doi.org/10.1057/s41283-019-00057-9

Chan, W. K., & Chu, E. Y. (2022). The impacts of corporate governance mechanisms and ownership structure on firm performance: A case study of Chinese dual-listed companies. *Advances in Decision Sciences, 26*(4), 98–126. https://doi.org/10.47654/v26y2022i4p98-126

Cogneau, P., & Hübner, G. (2009a). The (more than) 100 ways to measure portfolio performance. Part 1: Standardized risk-adjusted measures. *Journal of Performance Measurement, 13*(4), 56–71.

Cogneau, P., & Hübner, G. (2009b). The (more than) 100 ways to measure portfolio performance. Part 2: Special measures and comparison. *Journal of Performance Measurement, 14*(1), 56–69.

Deck, C., & Schlesinger, H. (2010). Exploring higher order risk effects. *The Review of Economic Studies*, *77*(4), 1403-1420.

Dichtl, H., & Drobetz, W. (2011). Portfolio insurance and prospect theory investors: Popularity and optimal design of capital protected financial products. *Journal of Banking & Finance, 35*(7), 1683–1697.

https://doi.org/10.1016/j.jbankfin.2010.11.012

Eeckhoudt, L., & Schlesinger, H. (2006). Putting risk in its proper place. *American Economic Review*, *96*(1), 280-289.

Goetzmann, W. N., Ingersoll, J. E., Jr., Spiegel, M. I., & Welch, I. (2002). *Sharpening Sharpe ratios* (Working Paper No. 9116). National Bureau of Economic Research. https://www.nber.org/papers/w9116

Guo, X., Chan, R. H., Wong, W. K., & Zhu, L. (2019). Mean–variance, mean–VaR, and mean–CVaR models for portfolio selection with background risk. *Risk Management, 21*(2), 73–98. https://doi.org/10.1057/s41283-018-00044-6

Hodges, S. (1998). *A generalization of the Sharpe ratio and its applications to valuation bounds and risk measures* (Working Paper). Financial Options Research Centre, University of Warwick.

Homm, U., & Pigorsch, C. (2012). Beyond Sharpe ratio: An application of the Aumann-Serrano index to performance measurement. *Journal of Banking & Finance, 36*(8), 2274–2284. https://doi.org/10.1016/j.jbankfin.2012.04.005

Ingersoll, J. E., Jr., Spiegel, M. I., Goetzmann, W. N., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies, 20*(5), 1503–1546. https://doi.org/10.1093/rfs/hhm025

Lam, K., Liu, T., & Wong, W. K. (2008). The magnitude effect in the over-and-underreaction in international markets. *International Journal of Finance, 20*(3), 4833–4862.

Lean, H. H., Lam, K., & Wong, W. K. (2010). Stochastic dominance and investors' behavior towards risk: The Hong Kong stocks and futures markets. *International Journal of Finance, 28*(2), 113–135.

Leland, H. E. (1999). Beyond mean-variance: Risk and performance measurement in a nonsymmetrical world. *Financial Analysts Journal, 55*(1), 27–36. https://doi.org/10.2469/faj.v55.n1.2240

Leung, P. L., Ng, H. Y., & Wong, W. K. (2012). An improved estimation to make Markowitz's portfolio optimization theory users friendly and estimation accurate with application on the US stock market investment. *European Journal of Operational Research, 222*(1), 85–95. https://doi.org/10.1016/j.ejor.2012.04.023

Leung, P. L., & Wong, W. K. (2008). On testing the equality of the multiple Sharpe ratios, with application on the evaluation of iShares. *Journal of Risk, 10*(3), 1–16.

Lu, R., & Hsieh, M. S. (2019). A performance analysis of dollar-cost averaging and self-annuitization. *Annals of Financial Economics, 14*(4), Article 1950017. https://doi.org/10.1142/S2010495219500172

Lu, R., & Hsiung, L.-Y. (2018). A performance evaluation of constant proportion portfolio insurance: An application of the economic index of riskiness. *Academia Economic Papers, 46*(3), 429–462.

Lu, R., Yang, C. C., & Islam, S. M. N. (2022). Studying the Corporate Governance and

Stock Performance Relationships: A Portfolio Performance Approach. *The International Journal of Finance*, forthcoming.
Lu, R., Yang, C. C., & Wong, W. K. (2018). Time diversification: Perspectives from the economic index of riskiness. *Annals of Financial Economics, 13*(3), Article 1850011. https://doi.org/10.1142/S2010495218500116
Ng, P., Wong, W. K., & Xiao, Z. (2017). Stochastic dominance via quantile regression with applications to investigate arbitrage opportunity and market efficiency. *European Journal of Operational Research, 261*(2), 666–678. https://doi.org/10.1016/j.ejor.2017.02.047
Niu, C., Guo, X., McAleer, M., & Wong, W. K. (2018). Theory and application of an economic performance measure of risk. *International Review of Economics & Finance, 56*, 383–396. https://doi.org/10.1016/j.iref.2017.11.007
Niu, C., Wong, W. K., & Xu, Q. (2017). Kappa ratios and (higher-order) stochastic dominance. *Risk Management, 19*(3), 245–253. https://doi.org/10.1057/s41283-017-0020-1
Paolella, M. S. (2014). Fast methods for large-scale non-elliptical portfolio optimization. *Annals of Financial Economics, 9*(2), Article 1440001. https://doi.org/10.1142/S2010495214400018
Pézier, J. (2008). *Maximum certain equivalent returns and equivalent performance criteria* (ICMA Centre Discussion Papers in Finance). University of Reading.
Schreiber, A. (2014). Economic indices of absolute and relative riskiness. *Economic Theory, 56*(2), 309–331. https://doi.org/10.1007/s00199-013-0782-0
Shrestha, K., Thompson, H. E., & Wong, W. K. (2007). Are the mortgage and capital markets fully integrated? A fractional heteroscedastic cointegration analysis. *International Journal of Finance, 19*(3), 4495–4513.
Spurgin, R. B. (2001). How to game your Sharpe ratio. *Journal of Alternative Investments, 4*(3), 38–46. https://doi.org/10.3905/jai.2001.318973
Wong, W. K., Leung, P. L., & Xu, J. (2005). The GARCH effects on the volume of China stock markets. *International Journal of Finance, 17*(1), 3290–3329.
Wong, W. K., Wright, J. A., Yam, S. C. P., & Yung, S. P. (2012). A mixed Sharpe ratio. *Risk and Decision Analysis, 3*(1-2), 37–65. https://doi.org/10.3233/RDA-2011-0055
Zakamouline, V., & Koekebakker, S. (2009). Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. *Journal of Banking & Finance, 33*(7), 1242–1254. https://doi.org/10.1016/j.jbankfin.2009.01.002
Zieling, D., Mahayni, A., & Balder, S. (2014). Performance evaluation of optimized portfolio insurance strategies. *Journal of Banking & Finance, 43*, 212–225. https://doi.org/10.1016/j.jbankfin.2014.03.006