BLACKTHORNE
CAPITAL MANAGEMENT
www.blackthorne.com
1221 Innovation Dr.
Whitewater, WI 53190

BLACKTHORNE WINS ABOVE REPLACEMENT PORTFOLIO (BWARP)

William L. Dougan, PhD, Professor of Management, University of Wisconsin, Whitewater
Head of Research, Blackthorne Capital Management
Joe Tenpenny, Vice President of Operations, Blackthorne Capital Management

Introduction

Since its initial appearance, the Sharpe ratio has been a dominant and ubiquitous tool in portfolio construction and performance evaluation. The nominal form of the ratio (Sharpe, 1956, 1994), expresses the risk-adjusted return of an investment as function of the difference between the annualized rate of return on the investment and a benchmark rate (typically the return rate on US Treasury securities) adjusted by return volatility, in a manner consistent with Markowitz (1952).

$$S_a = \frac{E[R_a - R_f]}{\sigma_a} = \frac{E[R_a - R_f]}{\sqrt{\sigma^2[R_a - R_f]}}$$

Where $S_a = \textit{the Sharpe ratio for portfolio a}$
$R_a = the\ annualized\ average\ daily\ return\ on\ portfolio\ a$
$R_f = the\ reference\ (risk - free)\ rate\ of\ return\ (usually\ US\ Treasuries)$
$\sigma_a = the\ annualized\ volatility\ of\ portfolio\ a$

The Sharpe ratio builds on the work of Roy (1952), who suggested that a minimal acceptable level of return, or “disaster level” of return be used as a basis for comparison, and that it also be adjusted by return volatility to account for the additional risk. Clark, de Silva and Thorley (2015) suggest that any risky benchmark rate can be used as a basis for comparison, and they refer to the resulting metric as the information ratio.

However, considerable subsequent work has shown that use of the Sharpe ratio or similar metrics for evaluating and assembling portfolios may introduce bias due to a number of characteristics of return distributions that do not meet assumptions of normality and/or asymptotic continuity (cf., Mandelbrot, 1963; Lo, 2002; Scholz, 2007; Bailey and Lopez de Prado, 2012, 2013). Numerous suggestions for revision of the measure have been proposed to adapt the measure to the observed, higher-moments of distributions of returns. For example, Sortino and Price (1994), identify the issue of asymmetry in returns distributions, and note the need to derive a measure that rewards upside volatility while penalizing downside volatility. Young (1991), suggests that annualized average return over a lookback period should be adjusted by the average percentage of drawdown during that period rather than volatility as a means of addressing the ability of a strategy to generate returns in relation to risk. Young’s measure was named after his employer at the time, California Managed Accounts Reports (CALMAR). Sometimes this measure is identified as the Return to Maximum Drawdown.

One recent work by Artemis Capital (2022), provides an insightful critique of some of the dangers of overreliance on Sharpe ratios in portfolio construction, and suggests the CWARP™, or Cole Wins above Replacement Portfolio as an alternative method for assembling portfolios. This method is intended to address issues of portfolio risk due to long-term regime changes in markets, and/or the attendant volatility that can occur under unusual but foreseeable market conditions. The method depends on both the correlation of Sortino metrics and the Calmar ratios of portfolio strategies, and is expressed as a comparison between an existing portfolio and a portfolio which includes an additional strategy that is under consideration for inclusion in that portfolio. It evokes the approach suggested by Woolner (2002) from baseball, which considers the value of a player in comparison with the value of the average player that would replace that player on a team. The CWARP™ measure is defined as follows:

$$\text{CWARP}^{\text{TM}} \equiv \left[\sqrt{\left(\frac{S_n}{S_p}\right) * \left(\frac{RMDD_n}{RMDD_p}\right)} - 1 \right] * 100$$

Where $CWARP^{TM} = Cole\ Wins\ above\ Replacement\ Portfolio\ metric$

$S_n = Sortino\ ratio\ of\ the\ new\ portfolio$

$S_p = Sortino\ ratio\ of\ the\ replacement\ portfolio$

$RMDD_n = Return\ to\ Max\ Drawdown\ \ of\ the\ new\ portfolio$

$RMDD_p = Return\ to\ Max\ Drawdown\ of\ the\ replacement\ portfolio$

This method is touted as an improvement on the Sharpe ratio, because of its ability to identify the benefits of diversification on a portfolio, particularly in the timing and magnitude of drawdowns. A portfolio component generates a higher CWARP™ value when it produces returns that are uncorrelated with other return streams in the portfolio, and when it demonstrates smaller average drawdowns in comparison to other component strategies in the portfolio. The creators of the method suggest that examination of metrics over a very long term is extremely important in creation of a successful portfolio. They then present some evidence of the effectiveness of the method in portfolio management.

The CWARP™ method is also not without its issues, however. Under conditions where both of the Sortino measures are negative, the CWARP™ measure can become inverted, and can result in a positive CWARP™ measure, which would prompt an interpretation that is opposite of the initial intent of the measure. Also, if the Sortino of the replacement portfolio is close to zero, the value of the resulting CWARP™ measure may become inflated in a manner that would misrepresent the positive value of the alternative under consideration. Representative examples of these scenarios are given in Table 1, and these reflect quite realistic circumstances for those evaluating portfolio performance.

The authors of the CWARP™ measure may argue that these cases are not likely to represent viable choices in portfolio construction, but they do make uninformed use of the measure problematic. In this paper, we conduct a simulation analysis to determine the frequency of occurrence of these problems using a sample of returns streams from the financial markets. We sample from the holdings listed in sectors of the Standard & Poor's Depositary Receipts Exchange Traded Funds (SPDR ETFs), implying that they are typical returns streams that an investor might encounter in the process of creating a portfolio. We suggest a modification of the CWARP™ method that will eliminate problems, and will provide some advantages for a useful approach for portfolio management.

Table1: Representative Problem Scenarios for CWARP™ Measure

Scenario	S_p	S_n	$RMDD_p$	$RMDD_n$	*Appropriate Signal*	*CWARP Measure*	*Problem*
Neg to Pos	-0.47	0.12	-0.04	0.13	Positive	-8.91	CWARP Inverted
Pos to Neg	0.31	-0.90	0.23	-0.67	Negative	190.81	CWARP Inverted
Neg to Neg	-1.11	-0.55	-0.62	-0.45	Positive	-40.03	CWARP Inverted
Small Number	0.03	0.58	0.02	0.31	Positive	1,768.99	CWARP Inflated

The Blackthorne Wins above Replacement Portfolio Metric

This research introduces an alternative method, the Blackthorne Wins above Replacement Portfolio (BWARP), to further address some of the shortcomings of the Sharpe ratio in providing a comprehensive assessment of a portfolio's risk and return profile. We first define the measure, then describe how it behaves in a manner that is different from the Sharpe and CWARP™, and how it can be used in portfolio construction. We argue that the BWARP method identifies portfolio constituents that produce attractive performance while reducing portfolio volatility. We also argue that the calculation time and computing requirements of portfolio construction using BWARP methods are acceptable for most portfolios and that BWARP methods eliminate the likelihood of problems in calculation and interpretation that exist with other methods.

In the following section, we define the BWARP metric. We then report the results of test of its efficacy, and then discuss the implications of its use in practice.

Definition of BWARP

The Blackthorne Wins Above Replacement Portfolio metric can be calculated using the following formula:

$$BWARP = \left[\frac{tanH(S_n - S_p) + tanH(RMDD_n - RMDD_p)}{2}\right] x\ 100$$

Where $BWARP = Blackthorne\ Wins\ above\ Replacement\ Portfolio\ metric$

$S_n = Sortino\ ratio\ of\ the\ new\ portfolio$

$S_p = Sortino\ ratio\ of\ the\ replacement\ portfolio$

$RMDD_n = Return\ to\ Max\ Drawdown\ of\ the\ new\ portfolio$

$RMDD_p = Return\ to\ Max\ Drawdown\ of\ the\ replacement\ portfolio$

The BWARP metric expresses the relative change in a portfolio by evaluating a prospective component as a function of two significant performance statistics, the Sortino ratio, which measures excess return over the downside deviation of a portfolio, and the Return to Maximum Drawdown metric, which captures the ratio of return in excess of the reference (risk-free) rate of return over a period of time in relation to the Maximum Drawdown (MDD) of a portfolio expressed as a percentage of the maximum over the same period of time. The use of the differences between the performance measures (as opposed to the ratio of the measures, as is used in CWARP™) eliminates the problem in the CWARP™ of excessively inflating the CWARP™ measures when the S_p or $RMDD_p$ measures are close to zero. The BWARP metric also eliminates the issue of the CWARP™ value being inverted when both Sortino measures are negative.

The hyperbolic tangent is introduced into the BWARP as a method of reducing the values of each of the elements of the measure to values between one and negative one. This eliminates the problem with the CWARP™ measure as described previously, while retaining the abilities to incorporate the information from both Sortino and Return to Maximum Drawdown, which are laudable considerations. The hyperbolic tangent function is asymptotically strictly monotonic (see Figure 1), and becomes nearly constant at input values greater than 2 or less than -2, which is consistent with the expectations about the likelihood of sustained performance of input parameters from financial markets. The shape of the function can act to dampen the effects of overreaction (Debondt and Thaler, 1985; 1987) by reducing the impact of extreme values of performance on the subsequent portfolio decisions. This represents a middle ground between performance-weighted portfolio theory (Black and Litterman, 1992) and models asserting the superiority of naïve, equal weighting (DeMiguel, Garlappi and Uppal (2007).

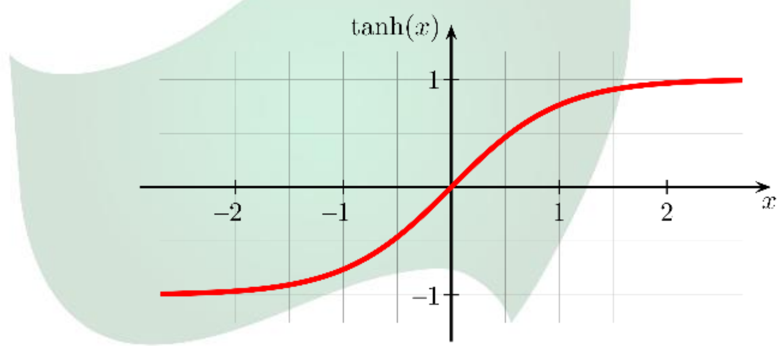


Figure 1: Hyperbolic tangent function

The equation produces a value theoretically centered on zero. A positive value of BWARP means the component under consideration for addition to the portfolio would increase the portfolio's downside diversification. A negative value of BWARP means the component under consideration for addition to the portfolio would reduce the downside diversification of the portfolio. The following section describes a simulation test of the CWARP™ method to demonstrate its relative performance under a variety of conditions. The first section describes a test of the relative occurrence of failure when calculating the CWARP^{TM} measure. The next section performs a robustness test on both measures by examining the performance of both measures from a random selection of market periods.

Stress Testing the CWARP^{TM} Metric

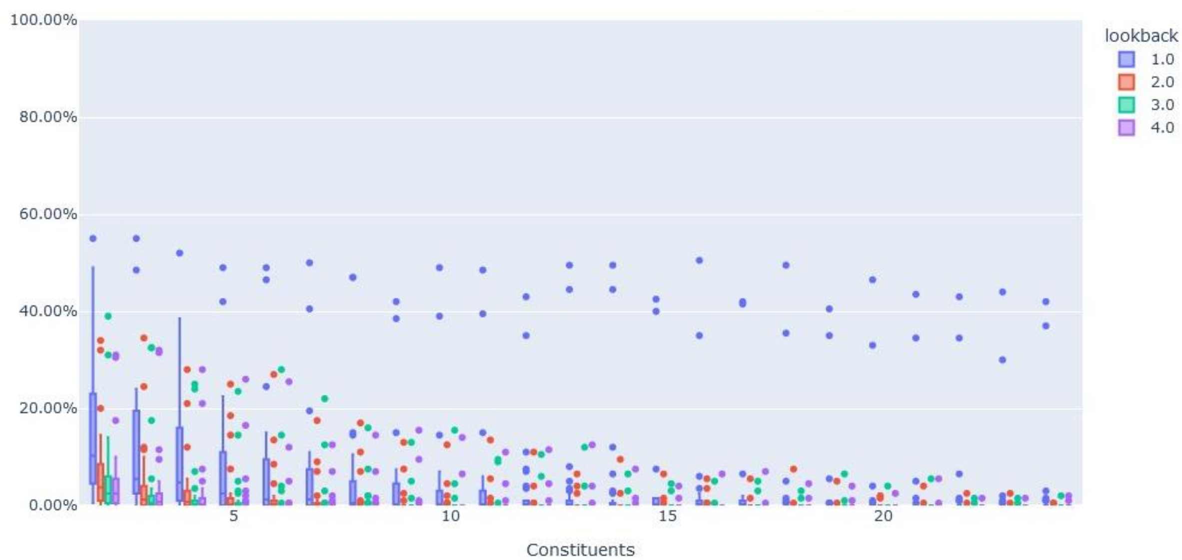
To assess the viability and behavior of the CWARP^{TM} metric, a simulation analysis drawing from a sample of actual market returns was used. A systematic approach was employed to evaluate the likelihood of the previously-described inversion problem with real market returns. First, a sampling pool was created by selecting daily returns streams of the top eight to ten equities from following SPDR ETFs: XLK, XLY, XLP, XLU, XLE, XLB, XLV, XLRE, and XLC. This created a pool of eighty equity streams from which random portfolios could be generated.

Portfolios of varying sizes, ranging from 2 to 24 constituents, were generated, to determine the incidence of calculation problems as a function of the portfolio size. The lookbacks for the measures were also varied, to examine the effect of longer samples of daily returns on the measures. Presumably, the incidence of inversion problems would be reduced with longer lookbacks, because of a reduction in volatility of the estimates of the parameters. Thirty random dates during the period from 01-01-2010 to 01-01-2023 were selected as time points for calculating the measures to eliminate the problem of idiosyncrasy of results from a single sampling period.

Starting with an existing portfolio of a given size, a new market was selected from the pool of equities, and the values of the Sortino and RMDD metrics from each resulting returns streams were calculated. The values of CWARP^{TM} were calculated, and any inversions of the CWARP^{TM} calculations (both measures of Sortino or RMDD negative) were noted. The values of CWARP^{TM} were calculated to determine whether the calculation of the measure was problematic (i.e., that the produced results were inverted). The comparison market was then replaced in the pool and the corresponding values of the Sortino and RMDD metrics were calculated for other combinations of portfolios and candidate equities. Incidence rates were calculated for portfolios of one year, two year, three years and four years in length. A total of six thousand portfolios (200 portfolios for each of the 30 sampling periods) were generated for each of the portfolio size and lookback combinations (23x4 = 92). This yielded a total of 552,000 data points. The number of times that the resulting calculations failed due to the nature of the measures was noted. This resulted in a series of inversion rates for the CWARP^{TM} metric. These results are shown in Figure 2. They are displayed as box plots as a function of portfolio size and in colors corresponding to the different lookbacks.

The results reveal that the incidence of inversions can be greater than forty percent, even when using lookbacks of 4 years with portfolios of 24 included constituents. Those rates increase considerably when the sampling periods are shorter and the number of constituent strategies in a portfolio are smaller. Because the BWARP measure produces interpretable results under all of those conditions, we argue that it represents a preferred method of determining suitability of a candidate strategy. We also note that the measures were calculated using real returns streams from equities that were presumably chosen for inclusion in the SPDR ETFs because of their reliability and appropriateness for particular sectors of the economy. This obviates the possibility that the problematic cases were due to the unrealistic nature of the returns streams that might have occurred with returns streams generated by Monte Carlo methods.

Figure 2: Box Plots of Incidence Rates of Problem Measures of CWARP™ Measure by Lookback and Portfolio Size



The CWARP™ measure does show a pattern whereby the percentage of problems is reduced when the number of constituents in a portfolio is larger. Thus the CWARP™ measure might be more effective for larger portfolios over longer sampling periods. However, it seems evident that the absence of any problems for the BWARP metric would suggest use of that method. The fact that the boxplots were created using data from multiple samples suggests that the incidence of problems would likely be higher for individual investors. Also, the fact that problems with the CWARP™ persist, even in more favorable conditions suggests a systemic sensitivity in the metric, most likely to periods of sustained market downturn. This may not be as likely in equity markets, which retain an upward drift, but may be problematic in futures markets, where such circumstances are more common.

Distributional Characteristics of the CWARP™ and BWARP Metrics

In order to further demonstrate the behavior of the BWARP and CWARP™ measures, we calculated the maxima, minima and first three moments of the distributions of measures from the sample of CWARP™ measures which were not eliminated because they were problematic and from the BWARP measures. Those values are given below in Tables 2 and 3, panels a and b.

Table 2: Distribution Characteristics of CWARP and BWARP Measures by Length of Lookback over the Sample Period 1/1/2010-1/01/2023

Panel a

Lookback (Years)	CWARP Mean	CWARP Median	CWARP Std Dev	CWARP Skew	CWARP Min	CWARP Max
1	14.37	0.84	498.56	121.17	(99.95)	83,486.92
2	7.75	0.78	384.37	165.35	(99.88)	84,294.10
3	7.44	0.80	651.82	321.21	(99.87)	228,802.84
4	4.85	0.93	137.13	148.92	(99.82)	32,220.10

Panel b

Lookback (Years)	BWARP Mean	BWARP Median	BWARP Std Dev	BWARP Skew	BWARP Min	BWARP Max
1	1.18	1.09	23.01	0.11	(99.91)	100.00
2	0.70	0.86	15.31	(0.04)	(99.84)	99.85
3	0.66	0.83	13.82	(0.27)	(99.99)	99.99
4	0.60	0.95	13.18	(0.32)	(99.83)	100.00

Table 3: Distribution Characteristics of CWARP and BWARP Measures by Number of Portfolio Components over the Sample Period 1/1/2010-1/01/2023

Panel a

Constituents	CWARP Mean	CWARP Median	CWARP Std Dev	CWARP Skew	CWARP Min	CWARP Max
2	60.43	7.83	1,667.12	119.85	(99.95)	228,802.84
3	34.11	4.91	566.52	59.08	(99.82)	49,084.50
4	23.71	3.69	585.62	100.39	(99.90)	75,718.50
5	15.57	2.66	499.68	97.58	(99.61)	53,982.46
6	14.13	2.05	619.60	117.28	(98.11)	84,294.10
7	7.20	1.73	185.32	120.19	(99.88)	26,015.80
8	8.88	1.55	553.65	146.34	(98.76)	83,486.92
9	4.39	1.21	72.64	55.24	(99.02)	6,097.27
10	2.83	0.99	38.70	35.47	(99.80)	2,674.48
11	3.46	0.97	88.81	78.64	(99.68)	10,239.77
12	9.85	0.91	724.76	103.73	(99.44)	81,024.56
13	2.07	0.76	44.34	69.72	(98.76)	4,882.92
14	1.70	0.63	39.14	75.58	(96.12)	4,390.46
15	1.70	0.63	30.03	31.02	(94.45)	1,901.94
16	1.25	0.54	29.94	53.82	(98.08)	2,762.09
17	1.70	0.59	58.95	111.53	(97.85)	8,054.04
18	1.24	0.52	51.56	118.36	(98.55)	7,211.78
19	1.38	0.42	51.27	97.48	(99.81)	6,566.65
20	0.95	0.48	34.87	99.36	(99.43)	4,527.48
21	2.99	0.39	184.91	92.01	(95.87)	20,454.42
22	0.94	0.45	45.36	139.58	(94.54)	6,765.36
23	0.67	0.42	22.72	71.82	(96.32)	2,482.73
24	0.93	0.38	41.99	97.94	(92.86)	5,363.50

Table 3: Distribution Characteristics of CWARP and BWARP Measures by Number of Portfolio Components over the Sample Period 1/1/2010-1/01/2023
Panel b

Constituents	BWARP Mean	BWARP Median	BWARP Std Dev	BWARP Skew	BWARP Min	BWARP Max
2	4.89	7.37	35.94	(0.26)	(99.99)	100.00
3	3.45	4.81	30.14	(0.17)	(99.82)	99.99
4	2.31	3.61	26.38	(0.19)	(99.74)	99.90
5	1.65	2.67	23.23	(0.08)	(98.23)	99.85
6	1.13	2.04	20.96	(0.17)	(99.35)	99.38
7	0.87	1.77	19.23	(0.14)	(94.55)	98.71
8	1.01	1.69	17.52	(0.03)	(94.33)	98.10
9	0.54	1.30	16.00	(0.05)	(95.20)	95.95
10	0.28	1.04	15.04	(0.23)	(94.70)	96.06
11	0.29	1.05	13.79	(0.15)	(93.20)	94.75
12	0.48	0.98	13.17	0.00	(97.22)	95.36
13	0.18	0.82	12.28	(0.06)	(93.66)	93.89
14	0.17	0.67	11.68	0.05	(81.67)	95.10
15	0.15	0.69	11.29	0.07	(87.09)	91.69
16	0.13	0.59	10.58	0.10	(81.10)	88.01
17	0.28	0.64	10.02	0.29	(70.19)	88.11
18	0.12	0.56	9.71	0.18	(83.93)	85.81
19	(0.03)	0.48	9.09	(0.09)	(71.33)	81.75
20	0.12	0.51	8.62	0.03	(77.99)	82.24
21	0.08	0.43	8.35	(0.04)	(66.94)	80.88
22	0.14	0.49	7.99	0.09	(80.69)	79.31
23	0.12	0.47	7.86	0.40	(66.99)	77.77
24	0.06	0.43	7.50	0.01	(79.57)	74.91

The tables demonstrate the significant differences in the behavior of the measures, even under conditions where the inverted measures are removed. The contrast between the measures in the higher moments is notable, particularly in the skew. This is attributable to the value inflation problem noted earlier.

Though the distributions for both measures are not centered on zero (these are, after all, returns streams associated with established firms that appear in SPDR ETFs because they are reliably positive), they reflect a trend toward zero as the estimates are affected by more observation periods and returns streams.

Potential for Further Analysis

The next task for evaluating BWARP measure is to determine its ability to improve the performance of the portfolio construction and allocation process. This will be addressed in our next paper. Another important task is to investigate the inclusion of other measures of performance into the BWARP approach. Because the use of the Hyperbolic tangent enables the standardization of measures of portfolio performance, these measures can be incorporated into the BWARP. Though the measures of Sortino and Return to Maximum Drawdown are equally weighted in the formulation provided in this paper, it is possible to consider a weighted combination of measures that have been standardized by the hyperbolic tangent transformation. Performance measures like dollar profitability can be added and incorporated into the measure. This would also open the measure to the use of approaches that employ weighted linear optimization as a basis for multi-objective decision making. We hope to formulate and test versions of these models in the future.

Practical Implications for Portfolio Managers

1) **The BWARP metric represents a viable approach to portfolio optimization.** The Blackthorne BWARP model introduces a novel metric for portfolio evaluation and construction, offering a robust alternative to traditional metrics like the Sharpe ratio. The research findings highlight the absence of problems in using BWARP, showcasing its versatility across different market signals, portfolio sizes, and starting points. Understanding these implications can enhance decision-making processes in real-world investment scenarios. This advances the arguments made in Artemis Capital (2022) by providing another approach to addressing the needs expressed by those authors.

2) **Variance of the metric increase with small portfolios.** It is crucial to recognize that the variance of the BWARP metric tends to increase when applying the method to small portfolios. For example, the transition from a single asset portfolio to a two-asset portfolio can result in extreme values of the BWARP measure. Consequently, careful consideration should be given when interpreting the BWARP metric for smaller portfolios.

3) **The BWARP metric, because it is robust to conditions that affect the CWARP, can be extended to other criteria that are important to portfolio construction**. Additional measures like dollar profitability, percentage of wins, conditional value at risk, etc., might be added and tested as components in a linear model. This furthers the wins above replacement approach to portfolio construction.

4) **Application of the BWARP metric may be computation intensive**. The methods for determination of wins above replacement involve a significant increase in the number of possibilities that must be considered in order to reach an optimum allocation. This may require methods for approximation of optimum as portfolio sizes increase.

References

Artemis Capital, (2022). *Moneyball for Modern Portfolio Theory: The Sharpe Ratio Problem and Cole Wins Above Replacement Portfolio Solution*. https://www.artemiscm.com/cwarp

Bailey, D. and López de Prado, M. (2012): The Sharpe Ratio Efficient Frontier, *Journal of Risk*, 15(2):3-44. https://ssrn.com/abstract=1821643

Bailey, D. and M. Lopez de Prado (2013): The strategy approval decision: A Sharpe ratio indifference curve approach, *Algorithmic Finance* 2(1):99-109. https://ssrn.com/abstract=2003638

Black, F. and Litterman, R. (1992) "Global portfolio optimization". *Financial Analysts Journal*, 48(5):28-43.

Clarke, R.G.; de Silva, H.; and Thorley, S. (2015), *Analysis of Active Portfolio Management, CFA Institute*

De Bondt, W. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3):793-805.

De Bondt, W. and Thaler, R. (1987). Further evidence on investor overreaction and stock market seasonality, *The Journal of Finance*, 42(3):557-581.
DeMiguel, V., Garlappi, L. and Uppal, R., (2009). Optimal versus Naïve diversification How Inefficient is the 1/N Portfolio Strategy? The Review of Financial Studies, 22(5): 1915-1953.

Lo, A.W. (2002): The statistics of Sharpe ratios. *Financial Analysts Journal* 58(4):36-52. https://doi.org/10.2469/faj.v58.n4.2453
Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394-419.

Markowitz, H. (1952) Portfolio selection. *The Journal of Finance,* 7(1):77-94.

Roy, A.D. (1952). Safety first and the holding of assets". Econometrica. ***20****(3):431–450. doi:10.2307/1907413. JSTOR 1907413.*

Scholz, Hendrik (2007). Refinements to the Sharpe ratio: Comparing alternatives for bear markets. *Journal of Asset Management,* ***7****(5):347–357. doi:10.1057/palgrave.jam.2250040. S2CID 154908707.*

Sharpe, W. F. (1966). "Mutual Fund Performance". Journal of Business. ***39****(S1):119–138. doi:10.1086/294846.*

Sharpe, W. F. (1994). "The Sharpe Ratio". The Journal of Portfolio Management. ***21****(1):49–58. doi:10.3905/jpm.1994.409501. S2CID 55394403*

Sortino, F. -A., and Price, L.N. (1994). Performance measurement in a downside risk framework. The Journal of Investing, 3(3):59-64.

Woolner, K. (2002). Understanding and measuring replacement level," in Joe Sheehan, Ed., *Baseball Prospectus* (Brassey's Inc: Dulles, VA,): pp. 455-466

Young, T.W. (1991). "Calmar Ratio: A Smoother Tool", Futures