Financial Economics Letters 2025 4(1) 31-36

ANSER PRESS

Financial Economics Letters

Homepage: https://www.anserpress.org/journal/fel

Sector-Level Out-of-Sample Performance of the Naive and Sharpe Portfolios Using a Covid-Correction Breakpoint

M. Ryan Haley [a,*]

[a] *Department of Economics, University of Wisconsin Oshkosh, Oshkosh, WI, US*

ABSTRACT

Recent research has demonstrated that many mean-variance and shortfall-based optimal portfolio selection fail to out-perform the Naive (1/n) Portfolio in out-of-sample testing. This paper revisits this line of inquiry by applying the Naive and Sharpe Portfolios to 1100 sector-specific S&P 500 re-sampled data sets from the 2007-2021 time frame. Using April 2020 as the baseline train-test split break point, the Naive Portfolio delivers statistically significantly superior Sharpe Ratios in the test data in ten of the eleven sectors. However, the Sharpe Portfolio delivers statistically significantly superior shortfall values in all eleven sectors in the test data. Using March 2020 and May 2020 as alternative breakpoints gave similar results to the baseline analysis. Interestingly, when the data set was truncated at February 2020 (i.e., before the Covid correction) the Sharpe Portfolio returned statistically significantly better Sharpe Ratios than the Naïve Portfolio in the test data in all but the Energy sector; as in the baseline analysis, the Sharpe Portfolio returned statistically significantly superior shortfall values for all eleven sectors. Thus, the Sharpe Portfolio can deliver acceptable out-of-sample performance, but the conditions for success appear to vary by sector and test data erraticism.

KEYWORDS

Optimal Portfolio Selection; Naive Diversification; Portfolio Choice

---

* Corresponding author: M. Ryan Haley
E-mail address: haley@uwosh.edu

ISSN 2972-3426
doi: 10.58567/fel04010003

Received 9 November 2024; Accepted 21 March 2025; Available online 21 May 2025; Version of Record 15 March 2025

1. Introduction and Motivation

DeMiguel et al. (2009), and later Haley (2016, 2017) among others, investigate how an array of mean-variance and shortfall-based optimal portfolio selection rules perform relative to the Naive (1/n) Portfolio, wherein an equal share of investable wealth is allocated to each candidate asset. Extensive out-of-sample testing indicates that optimal portfolio selection methods often struggle to retain their defining performance attributes (e.g., Sharpe Ratio maximization or shortfall minimization) in test data. In fact, the Naive Portfolio often delivers superior out-of-sample performance. Prevailing explanations for this observed phenomenon include estimation error and high transaction costs (Kirby and Ostdiek, 2012), though other explanations have also been advanced (e.g., Hwang et al., 2018; Platanakis et al., 2021). Regardless, these findings naturally raise concerns about the practical efficacy of optimal portfolio selection methods in investment settings.

The purpose of this paper is to revisit these findings with large amounts of more recent stock return data and to assess the corresponding out-of-sample performance. However, unlike prior studies, the focus here is sector-specific performance before and after Covid. Effecting this inquiry entails an extensive series of empirical simulations based on monthly return data from the S&P 500 during the 2007-2021 time frame. The empirical investigations are conducted on a sector-by-sector basis, as if the investor were creating sector-specific portfolios. The overall findings are intriguing: for most sectors, using April 2020 as the train-test split break point, the Naive Portfolio dominates in out-of-sample testing (in terms of Sharpe Ratios) in nearly every sector. Also intriguing is how the Sharpe Portfolio delivers a remarkably stable shortfall advantage over the Naïve Portfolio, indicating that the latter method may be more exposed to more frequent and potentially larger negative monthly returns.

The remainder of the paper is organized into four sections. The next section reviews the Sharpe Portfolio. Section three describes the data and the cross-validation methodology. Cross-validation results, robustness checks, limitations, and discussions appear in section four. The final section briefly summarizes the contribution and outlines several prospects for future research.

2. The Sharpe Portfolio

The Sharpe Portfolio (e.g., Sharpe, 1994) directs investors to hold assets in a way that solves the following optimization problem:

$$max_{<W_p>} W_p^T \mu / \left(W_p^T \Sigma W_p\right)^{1/2}$$

where w_p is a vector that denotes the proportion of investable wealth allocated to each of the *n* candidate assets; μ is a vector containing the average historical return for each of the *n* candidate assets; and Σ is the covariance matrix of historical (excess) returns. The baseline case in the analysis herein uses a zero risk-free rate for simplicity. The optimal weights (i.e., the weight values that maximize the Sharpe Ratio) are as follows:

$$W_{sr} = \Sigma^{-1}\mu / \delta^T \Sigma^{-1}\mu$$

where δ is a conformable vector of ones. The w_{sr} values maximize the average portfolio (excess) return per unit standard deviation of portfolio returns, philosophically consistent with the reward-risk paradigm to which the Sharpe Portfolio belongs. Note that shorting is permitted, which implies that w_{sr} must sum to one, but individual weights are not constrained to the [0,1] interval.

3. Data and Methodology

The primary data source was S&P 500 monthly returns from the 2007-2021 time frame. Monthly stock returns were calculated using the monthlyReturn() function from the quantmod package in the R programming language, for a total of 180 monthly returns per stock.

The following eleven sectors were identified and studied:

- Communication Services
- Consumer Discretionary
- Consumer Staples
- Energy
- Financials
- Health Care
- Industrials
- Information Technologies
- Materials
- Real Estate
- Utilities

For each sector, 100 samples containing n_{sector} stocks were drawn at random from the primary S&P 500 data source noted above. The sample size, n_{sector}, was adjusted to be 50% of the stock count in each sector; this produced a stratified sample wherein each stock sector sample was equally representative of the total stock count per sector, which varied from 21 to 74. In total, 1100 data sets (100 from each sector) were passed through holdout cross-validation wherein 90% of the data served as the training set while the remaining 10% of the data (i.e., the most recent data) served as the test set. For each train-test split, the training data was used to find the Sharpe Portfolio weights, which were then applied to the test data. The out-of-sample performance of the Sharpe Portfolio in the test data was the primary focus, but in-sample performance was also reported to add context for the out-of-sample findings. Each sector was assessed individually.

Two primary portfolio performance metrics were recorded for each data set and then averaged within each sector: the shortfall proportion (see Roy, 1952; Stutzer 2000; Haley and Whiteman, 2008, or Haley, 2016) and the Sharpe Ratio. The former is the proportion of monthly portfolio returns below the risk-free rate, which was set to zero in the baseline analysis; large shortfall values indicate a high chance a (monthly) portfolio return will be negative. The latter is the usual Sharpe Ratio, defined as the average excess return over the standard deviation of returns.

Two additional details warrant disclosure. First, any outliers that occurred during the simulations, albeit extremely rare, were detected using the Mahalanobis distance and/or the reciprocal condition number of the covariance matrix (for the Sharpe Portfolio). However, these filters proved mostly unnecessary given the shrinkage estimator (from R's corpcor library) used for the covariance matrix, but were still included to avoid any possible outlier effects. Second, any stocks that did not have complete monthly return data for the 2007-2021 time frame were excluded from the analysis.

4. Results and Discussion

The results from the baseline cross-validation process outlined above appear in Table 1, which specifically contains the aforementioned metrics for the Sharpe and Naive Portfolios in the training and test data. The three focal points of the estimation were as follows:

1) The first point of interest was to assess the out-of-sample performance of the Naïve vs. Sharpe Portfolio.

2) Two sensitivity analyses were conducted wherein the Covid breakpoint was perturbed to March and May 2020, respectively.

3) A secondary analysis mirrored the primary analysis above except excluded data from March 2020 and forward. This analysis permitted an assessment using pre-Covid-only data. A 90%-10% split was used for this analysis as well.

Table 1. Cross Validation Summary Results with Covid Break Point*.

		Training Data		Test Data	
Sector	Method	Shortfall	Sharpe Ratio	Shortfall	Sharpe Ratio
Communication	Naive	0.376	0.204	0.327	0.322
(n=14)	Sharpe	0.021	0.414	0.022	0.184
Consumer-D	Naive	0.399	0.173	0.283	0.611
(n=32)	Sharpe	0.020	0.586	0.025	0.099
Consumer-S	Naive	0.369	0.215	0.340	0.341
(n=16)	Sharpe	0.022	0.379	0.020	0.364
Energy	Naive	0.439	0.048	0.380	0.350
(n=11)	Sharpe	0.021	0.248	0.033	-0.185
Financials	Naive	0.396	0.093	0.291	0.601
(n=33)	Sharpe	0.017	0.522	0.026	0.150
Health Care	Naive	0.340	0.279	0.275	0.404
(n=32)	Sharpe	0.018	0.524	0.020	0.360
Industrials	Naive	0.355	0.187	0.275	0.521
(n=37)	Sharpe	0.018	0.580	0.023	0.134
Information-T	Naive	0.373	0.229	0.286	0.591
(n=37)	Sharpe	0.015	0.562	0.022	0.281
Materials	Naive	0.374	0.150	0.367	0.657
(n=14)	Sharpe	0.013	0.340	0.028	0.081
Real Estate	Naive	0.410	0.091	0.360	0.528
(n=15)	Sharpe	0.016	0.345	0.020	0.174
Utilities	Naive	0.370	0.113	0.352	0.242
(n=14)	Sharpe	0.027	0.370	0.031	-0.076

*Note: *The figures presented are averaged over 100 sector-specific data sets. 90%-10% train-test split cross validations of each data set were used; the train-test split point was April 2020. Each data sets was created by randomly selecting n_{sector} stocks from the corresponding sector. Shortfall is the proportion of negative portfolio returns; larger values are worse, indicating higher potential for negative monthly portfolio returns. The Naive Portfolio delivered superior out-of-sample Sharpe Ratio performance – the primary focus – in all sectors except Consumer Staples; this performance was statistically significantly better based on a paired-t assessment (|t| = 6.708). However, the Sharpe Portfolio delivered remarkably stable shortfall performance in out-of-sample testing and was superior in all sectors to the Naive Portfolio's shortfall performance (|t| = 12.25).*

The in-sample results in Table 1 are intuitively appealing. The Sharpe Portfolio dominates the Naive Portfolio in all sectors, delivering superior Sharpe Ratios and superior shortfall values. The out-of-sample comparisons, however, contrast sharply to the in-sample performance, in most cases affirming the prior research findings of Naive Portfolio dominance in out-of-sample settings (i.e., larger Sharpe Ratios). However, the Sharpe Portfolio's out-of-sample shortfall values are superior in every sector, mirroring its shortfall dominance in the training data. The results presented in Table 1 were similar when using March 2020 or May 2020 as the breakpoint instead of April 2020.

To more fully explore the findings in Table 1, the data from March 2020 to the end of 2021 were excluded in a second analysis. The goal was to ascertain if a holdout (test) set taken from a more stable market era void of extreme variations like the Covid correction would allow the sector-specific Sharpe Portfolio's optimal properties to present in the test data. The results support this possibility. In this secondary analysis, the Naive Portfolio only beat the Sharpe Portfolio (in terms of out-of-sample Sharpe Ratio) in the Energy sector; the results were statistically significant based on a paired t-test (|t| = 4.76). As in the baseline results from Table 1, the Sharpe Portfolio delivered statistically significantly better shortfall values in all eleven sectors (|t| = 23.69).

As with any study, it is important to list caveats, limitations, and shortcomings that may affect the results. First, the data come from a particular time period (2007-2021), which may not fully reflect the larger data-generating processes. Second, the sector-specific results presume that the investor is intent on building a sector-specific portfolio instead of a fully diversified portfolio that involves all sectors simultaneously. Third, while many simulation cycles were considered, any given data set is quite small and suffers from estimation error to some degree.

5. Summary, Discussion, and Conclusions

This paper revisits the findings of DeMiguel et al. (2009) using an abundance of recent stock return data. 1100 data sets selected from the eleven sectors in the S&P 500 formed the basis of this inquiry. The Naive Portfolio delivered better out-of-sample performance (in terms of Sharpe Ratio) than the Sharpe Portfolio in most sectors in the baseline analysis wherein the training and test data were divided at the Covid breakpoint (April 2020). However, the Sharpe Portfolio outperformed the Naïve Portfolio in all eleven sectors in terms of shortfall. Several alternative Covid breakpoints were tried, giving similar results.

A secondary study was done wherein the post-Covid data was excluded; these results were very different, indicating that the Sharpe Portfolio outperformed the Naïve Portfolio in terms of out-of-sample Sharpe Ratio and shortfall values. These findings vividly illustrate how the choice of test set can influence the results. The primary and secondary results, taken together, suggest that optimization-based portfolio selection rules like the Sharpe Portfolio are more affected by marked differences in the training vs. test data, at least in terms of Sharpe Ratios. However, the results may also be driven, at least in part, by the sector-specific nature of the portfolios studied herein.

One general direction for future research would be developing simple modifications of the Sharpe Portfolio (and other basic portfolio selection rules) that can improve out-of-sample performance without adding extensive complexity (e.g., Kirby and Ostdiek, 2012). Another research option would be comparing the out-of-sample efficacy of non-standard portfolio selection rules (e.g., Haley, 2008; Haley, 2018), which don't rely on traditional moments like mean and standard deviation, to the Naive Portfolio on a sector-by-sector basis.

Funding Statement

This research received no external funding.

Acknowledgments

Acknowledgments to anonymous referees' comments and editor's effort.

Conflict of interest

The author claims that the manuscript is completely original. The author also declares no conflict of interest.

References

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial studies* 22, 1915-1953. https://doi.org/10.1093/rfs/hhm075

Haley, M. (2008). A Simple Nonparametric Approach to Low-Dimension, Shortfall-Based Portfolio Selection. *Finance Research Letters* 5, 183--190. https://doi.org/10.1016/j.frl.2008.04.002

Haley, M. (2016). Shortfall Minimization and the Naive (1/N) Portfolio: An Out-of-Sample Comparison. *Applied Economics Letters* 23, 926–929. https://doi.org/10.1080/13504851.2015.1119788

Haley, M. (2017). K-Fold Cross Validation Performance Comparisons of Six Naive Portfolio Selection Rules: How Naive Can You Be and Still Have Successful Out-of-Sample Portfolio Performance?. *Annals of Finance* 13, 341--358. https://doi.org/10.1007/s10436-017-0301-4

Haley, M. R. (2018). A Moment-Free Nonparametric Quantity-of-Quality Approach to Optimal Portfolio Selection: A Role for Endogenous Shortfall and Windfall Boundaries?. *Journal of the Operational Research Society* 69, 1678-1687.

Haley, M., and Whiteman, C. (2008). Generalized Safety First and a New Twist on Portfolio Performance. *Econometric Reviews* 27, 457--485. https://doi.org/10.1080/07474930801960360

Hwang, I., Xu, S., and In, F. (2018). Naive Versus Optimal Diversification: Tail risk and Performance. *European Journal of Operational Research* 265, 372-388. https://doi.org/10.1016/j.ejor.2017.07.066

Kirby, C., and Ostdiek, B. (2012). It's all in the timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification. *Journal of Financial and Quantitative Analysis* 47, 437-467. https://doi.org/10.1017/S0022109012000117

Platanakis, E., Sutcliffe, C., and Ye, X. (2021). Horses for courses: Mean-variance for asset allocation and 1/N for stock selection. *European Journal of Operational Research*, 288, 302-317.

Roy, A. (1952). Safety First and the Holding of Assets. *Econometrica*, 20, 431--449.

Sharpe, W. (1994). The Sharpe Ratio. *Journal of portfolio management*, 21, 49-58.

Stutzer, M. (2000). A Portfolio Performance Index. *Financial Analysts Journal*, May/June, 52--61.