A Note on Ambiguity in Portfolio Performance Measures
Author(s): David Peterson and Michael L. Rice
Source: *The Journal of Finance*, Vol. 35, No. 5 (Dec., 1980), pp. 1251-1256
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2327098
Accessed: 28/11/2009 08:22

THE JOURNAL OF FINANCE • VOL. XXXV, NO. 5 • DECEMBER 1980

A Note on Ambiguity in Portfolio Performance Measures

DAVID PETERSON and MICHAEL L. RICE*

IN A RECENT ARTICLE in the *Journal of Finance* [4], Richard Roll argues that the Security Market Line (SML) criterion gives ambiguous performance signals for portfolio evaluation. This note reports results of an empirical test of robustness of the SML criterion applied to different indices over two time periods. The evidence suggests the ambiguity issue may be moot.

Roll argues that ambiguity results from the inability to decide upon a unique index to be used in the estimation of portfolio betas. When a mean-variance inefficient index is used, portfolios can be ranked on the basis of vertical deviations from an empirically fit SML. However, different inefficient indices will yield different SML's, and thus different rankings. When a mean-variance efficient index is used, all observations will plot on the SML and it will be impossible to assign rankings to the portfolios. As Roll demonstrates, it is possible that rankings can be reversed from one mean-variance inefficient index to another.

Mayers and Rice [3] take issue with the findings of Roll. They do not question Roll's proof of the ambiguity of the SML criterion, but instead argue that the SML criterion can be used to identify superior performance. By setting up a model with an informed investor and uninformed investors, they show that the SML criterion will designate inferior portfolio managers correctly, but may incorrectly designate superior portfolio managers. Mayers and Rice claim that their conclusions do not warrant rejection of the SML criterion as a ranking device.

As Roll [5] points out in his reply, Mayers and Rice are missing the point because they assume that everyone agrees upon which market index is to be used. An inefficient index is required to obtain rankings of portfolios, but there is no guarantee that the ranking reflects actual preference orderings of investors. Roll goes on to argue that the SML criterion should be abandoned because of the ambiguity issue. As an alternative criterion, he suggests measuring portfolio performance against the efficient frontier in mean-variance space. Since an index does not have to be chosen for this measurement, a degree of ambiguity is removed. Roll suggests that if the SML criterion is to be saved, it needs to be empirically demonstrated that commonly used indices do not rank portfolios very differently.

In the section that follows we examine the ambiguity issue empirically. For

* Respectively, Doctoral Candidate, Graduate School of Business Administration, University of North Carolina, and Associate Professor, Babcock Graduate School of Management, Wake Forest University.

purposes of comparison, we have also examined the evaluative robustness of two widely used performance measures developed by Treynor[1] and Sharpe.[2]

I. Empirical Evidence

In order to test whether different indices rank portfolios ambiguously, quarterly total returns on fifteen mutual funds were collected over two five-year periods, 1967–71 and 1972–76. Quarterly closing net asset values were obtained from the *Wall Street Journal* and dividend and capital gains distributions were obtained from *Standard and Poor's Dividend Record.* The fifteen funds were randomly selected from those funds for which data was available for the entire ten years. For the same two five-year periods, quarterly total returns were calculated for the Dow Jones Industrial Average and the Standard and Poor's 500 Stock Index.[3] In addition, total quarterly returns were calculated for a value weighted index and for an equally weighted index, where each index was created from the monthly CRSP tape of all New York Stock Exchange issues. The value weighted index should closely *approximate the* N.Y.S.E. Composite Index, adjusted for dividends.

Mutual fund betas were estimated from the time series regression

$$R_{jt} - R_{Ft} = \alpha_j + \beta_j(R_{Mt} - R_{Ft}) + \epsilon_{jt}, \tag{1}$$

where R_{jt} is the total return of the jth mutual fund for the tth quarter, R_{Mt} is the total return on the market index for quarter t, and R_{Ft} is the yield on three month Treasury Bills[4] issued at the beginning of the quarter. ϵ_{jt} is the disturbance term with assumed standard properties. Equation (1) was estimated for each mutual fund against each of the four market indices, for each of the two twenty-quarter time periods.

Next, a SML was fit for the two twenty-quarter time periods for each of the four sets of market index β's using the cross sectional regression

$$\bar{R}_j = \gamma_o + \gamma_1 \beta_j + \mu_j, \tag{2}$$

where $\bar{R}_j$ is the mean total return of mutual fund j for the appropriate twenty quarter time period, β_j is the appropriate mutual fund beta, and μ_j is the disturbance term. An empirical line was also fit by substituting each fund's standard deviation of returns for beta in Equation (2).

Table I ranks the mutual funds' performance in each five year period using the SML criteria for each of the four indices: the Dow Jones Industrial Average, the Standard and Poor's 500 Stock Index, the equally weighted CRSP index, and the value weighted CRSP index. An SML ranking based on standard deviation is also included. These rankings are based on vertical distances from empirically fit

[1] The Treynor performance measure is excess returns divided by beta. See [8].

[2] The Sharpe performance measure is excess returns divided by standard deviation of returns. See [6].

[3] From Standard & Poor's *Trade and Security Statistics*, Standard & Poor's *Daily Stock Price Record*, and *The Dow Jones Investors Handbook*.

[4] T-bill yields were taken from the *Federal Reserve Bulletin*.

Table I

Mutual Fund Performance Rankings Using the Security Market Line Criterion with Risk Measured by β, on Four Different Market Indices, and Standard Deviation

	1967–71					1972–76				
	Dow Jones Industrials	S&P 500	Equally Weighted CRSP	Value Weighted CRSP	Standard Deviation	Dow Jones Industrials	S&P 500	Equally Weighted CRSP	Value Weighted CRSP	Standard Deviation
Bulloch	4*	4*	4*	4*	4*	6*	6*	6*	5*	6*
Delaware	10	10	10	10	9	8	8	7	8	8
Dreyfus	13	13	12	13	12	9	9	9	9	9
First Investors Growth	8	9	8	9	10	15	15	15	15	15
Fundamental Investors	14	14	14	14	13	13	13	13	13	13
National Security Income	6	6	6	6	6	5*	7	5*	7*	5*
New Horizon	1*	1*	1*	1*	1*	14	12	14	12	14
Pioneer	3*	3*	3*	3*	3*	2*	2*	2*	2*	2*
Putnam Growth	5	5	5	5	5*	7*	5*	8	6*	7
Putnam Income	12	12	11	11	11*	12*	14	10	14	10
Templeton Growth	2*	2*	2*	2*	2	1*	1*	1*	1*	1*
United Income	7	7	7	7	7	10	10	12	10	12
Value Line Income	9	8	9	8	8	4*	4*	4*	4*	4*
Value Line Special Situations	11	11	13	12	15	3*	3*	3*	3*	3*
Wellington	15	15	15	15	14	11	11	11	11	11
			W = .9857					W = .9734		

Table II

Mutual Fund Performance Rankings Using the Treynor Performance Measure, on Four Different Market Indices, and the Sharpe Performance Measure

	1967–71 TREYNOR					1972–76 TREYNOR				
	Dow Jones Industrials	S&P 500	Equally Weighted CRSP	Value Weighted CRSP	SHARPE	Dow Jones Industrials	S&P 500	Equally Weighted CRSP	Value Weighted CRSP	SHARPE
Bulloch	4	4	4	4	4	6	6	6	7	6
Delaware	11	11	11	10	10	8	8	8	8	8
Dreyfus	12	12	12	12	12	10	10	10	11	10
First Investors Growth	7†	8†	8	7†	9	15	15	15	15	15
Fundamental Investors	14	14	14	14	14	13	13	13	13	13
National Security Income	6	6	6	6	6	5	4	4	5	4
New Horizon	2	2	2	2	2	14	14	14	14	14
Pioneer	3	3	3	3	3	2	2	2	2	2
Putnam Growth	5	5	5	5	5	7	7	7	6	7
Templeton Growth	13	13	13	13	13	4	5	5	3†	5
United Income	1	1	1	1	1	1	1	1	1	1
Value Line Income	7†	7	7	7	7	12	12	12	12	12
Value Line Special Situations	9	8†	9	9†	8	3	3	3	3†	3
Wellington	10	10	10	11	11	11	11	11	10	11
	15	15	15	15	15	9	9	9	9	9
			W = .9952					W = .9949		

† Indicates Tie

SML's. As an example, in the period 1967–71, with the Dow Jones Industrial Average as the index, the New Horizon Fund was ranked in position number one as the top performer, the Value Line Income Fund was ranked in position number nine, and the Wellington Fund was ranked in position number fifteen as the worst performing fund. Asterisks mark the rankings of funds judged "winners" because they plotted above the fit SML; funds not marked were judged "losers" because they plotted below the fit SML.

Table II ranks the funds in each five year period by their Treynor performance measure for each of the four market indices. Rankings are also provided for the Sharpe performance measure. For both time periods it can be seen that all of the performance measures ranked the funds nearly identically.[5]

The results indicate that for the fifteen mutual funds examined over the two five-year periods, little change in ranking occurred when the market index was varied. In addition, when standard deviation was substituted for beta, little change in ranking occurred. For each time period, for both the SML criterion and for the Treynor and Sharpe performance measures, all rankings were highly correlated. The largest change in ranking for any fund, for either time period, and for either evaluation technique (SML or performance measures) was four positions. The rankings for the Treynor and Sharpe performance measures exhibited less ambiguity then did the SML criterion.

As a statistical test of the hypothesis that the rankings under different indices are not significantly different, the Kendall Coefficient of Concordance (W) was calculated for each set of rankings. The Coefficient of Concordance is an appropriate non-parametric measure of interest reliability where W expresses the degree of association among a number of sets of rankings.[6] When the number of items ranked is greater than seven, then $k(N-1)W$ is approximately distributed as chi-square with $N-1$ degrees of freedom. Using this test statistic on each of the four sets of rankings, in all cases it is possible at the .001 level of significance to reject the null hypothesis that the rankings are unrelated.

Despite their theoretical short-comings, the portfolio evaluation tools tested are widely used in paractice. We have shown in this note that little serious injustice is committed in the process. Of course, until more evidence is available, evaluation techniques that rely on market indices, such as the SML criterion and Treynor performance measure, should be used with caution and awareness of the ambiguity potential.

[5] Richard Roll, in response to an earlier draft, pointed out that ranking portfolios by mean return ($\bar{R}$) provides a ranking not greatly different from those of the other measures tested. Though we appreciate his point, the similarity of ranking by $\bar{R}$ is due to the relative flatness of slope of the Security Market Line (SML) over the periods examined. We doubt whether the rankings between $\bar{R}$ and the risk-adjusted measures would be so similar during periods of more steeply sloped SML's since the SML ranking device is based on vertical distances from the line.

[6] $W = \dfrac{s}{1/12k^2(N^3 - N)}$ where s is the sum of squares of observed deviations from the mean of the sums of the ranks, k is the number of sets of rankings, N is the number of items ranked, and the denominator represents the maximum possible sum (s) which could occur with perfect agreement among k rankings. $0 \leq W \leq 1$; 0 represents no agreement in rankings and 1 indicates perfect agreement in rankings. See (7).

REFERENCES

1. B. Fielitz, and D. Thompson. "Composite Risk Measures in Performance Evaluation." Unpublished manuscript.
2. M. Jensen. "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios." *Journal of Business*, Vol. 42, No. 2 (April 1969), p 167-247.
3. D. Mayers, and E. Rice. "Portfolio Performance, Residual Analysis and Capital Asset Pricing Model Tests." Unpublished manuscript.
4. R. Roll. "Ambiguity When Performance Is Measured by the Securities Market Line." *Journal of Finance*, September 1978, p. 1051-1068.
5. R. Roll. "A Reply to Mayers and Rice." Unpublished manuscript, January 1979.
6. W. Sharpe. "Mutual Fund Performance." *Journal of Business*, Vol. 39, No. 1, Part II (January 1966), p. 119-139.
7. S. Siegel. *Nonparameteric Statistics For the Behavioral Sciences* New York: McGraw-Hill Book Co. 1956, p. 229-39.
8. J. Treynor. "How to Rate Management of Investment Funds." *Harvard Business Review*, Vol. 43, No. 1 (January-February 1965), p. 63-76.