How to Beat the S&P 500 with Portfolio Optimization

Tal Schwartz, Ph.D.

DePaul University

March 15, 2000

Contact Information:
Tal Schwartz, Ph.D.
Department of Finance
1 East Jackson Boulevard
Chicago, IL 60604-2287
Phone: 312-362-6887
Fax: 312-362-6566
ts4u@earthlink.net

tschwart@wppost.depaul.edu

How to Beat the S&P500 with Portfolio Optimization

This paper examines the long-run efficiency of the most popular market benchmark, the Standard&Poors 500 index. We apply the Markowitz mean-variance optimization techniques to generate efficient frontier portfolios over the period of 1978 to 1998. Although only simple historical return estimates are used, the out-of-sample performance of the efficient frontier portfolios is superior to that of the S&P 500 in both return and risk. We conclude that the common practice of indexing portfolios to the market has been an inefficient investment strategy that can be dominated by investing in optimized portfolios.

1. Introduction

As of the end of 1999 institutional and individual investors have indexed a combined total of over $1 Trillion, up from $200 billion in 1990 and $9 billion in 1980. The magnitude and continuing trend toward indexing suggests that investors believe that this is an efficient investment strategy. That is, investors believe that alternative strategies with greater expected return and lower volatility aren't readily available.

The most popular market benchmark for indexation, the capitalization-weighted Standard&Poors 500 index, has outperformed a large majority of professionally managed funds. Over the past decade, the S&P 500 returns have been better than the results of 89% of all U.S. mutual funds and the average outperformance is reported to be 340 basis points (see Ellis (2000)). The professionals' performance lag has been even more exacerbated in recent years with the S&P 500 returning in excess of 26% in the 1995 to 1999 period.

As illustrated above, indexation has become a big industry causing increased demand for stocks belonging to major market indexes. Moreover, there has been an increasing emphasis on performance evaluation and the S&P 500 index is the most popular benchmark for performance evaluation. Chan and Lakonishok (1993) suggest that this extra demand translates into excess stock returns for stocks belonging to the S&P 500. They find that membership in the S&P 500 over the period 1980 to 1991 resulted in an excess return of 3.03% per year with a significant t-statistic of 3.90.

While the S&P 500 index has experienced significant excess returns, Haugen and Baker (1991) claim that capitalization-weighted market indexes, such as the S&P 500, are inefficient. The authors argued that a cap-weighted market portfolio will be inefficient because of taxes, restrictions on short-selling, heterogeneous expectations, and opportunity sets that differ across investors. They constructed a portfolio from the 1000 largest capitalization stocks with the minimum variance over the trailing twenty-four months.[1] This portfolio was then tracked for the next quarter, and reformed. Haugen and Baker found that the resulting simulation had lower variance and higher return than the Wilshire 5000 over the 1972-1989 period.[2]

In this article, we endeavor to answer two questions. First, given that over the past two decades beating the S&P 500 has been so difficult, are Haugen and Baker's claim that all capitalization-weighted portfolios are inefficient, holds for the S&P 500 index? Second, if Haugen and Baker are right, how can investors profit from this inefficiency and create portfolios that provide higher return and lower risk than the S&P 500?

We answer these questions by examining the long-run efficiency of the S&P 500 index. We apply the Markowitz mean-variance optimization techniques to generate efficient frontier portfolios and track their performance out-of-sample. Although only simple historical return estimates are used, the out-of-sample performance of the efficient frontier portfolios is superior to that of the S&P 500 in both return and risk. Our results confirm Haugen and Baker's assertions, the S&P 500 has been an inefficient index and indexing has been an inefficient investment strategy. Finally, we demonstrate how investors can consistently beat the S&P 500 index by investing in optimized portfolios.

2. Data and Methodology

We begin by identifying the stocks belonging to the S&P 500 from the fourth quarter reports of the Compustat Primary Annual Files over 1978 to 1997.[3] We than make every effort to locate the matching stocks in the CRSP database.[4] Daily return data is collected from CRSP over the

[1] Haugen and Baker constrained their optimization so that the minimum weight on any stock was zero (no short-selling) and the maximum was 1.5% of the portfolio and no more than 15% of the portfolio could be invested in any one industry.
[2] The Wilshire 5000 is the most comprehensive U.S. equity cap-weighted index.
[3] Compustat identifies stocks using non-unique cusip.
[4] Several Compustat stocks' cusips were missing from the CRSP files. Every effort was made to match the stocks precisely, but when there was ambiguity because of multiple stock classes, the most likely stock class was selected based on factors such as market cap.

calendar year in which the stocks are included in the S&P 500 index and for the following calendar year.[5] It is important to limit the number of missing returns because any missing data will be filled-in from the stock's statistical sample distribution.[6] To minimize the effects of missing data on the optimization, we drop stock's that are missing more than 10% of their daily return data in either calendar years. Table 1 shows that the stock sample contains most of the components of the S&P 500, with the number of stocks ranging between 415 and 467. The number of industry groups, as classified by their 2 digit SIC code, varies between 46 and 61. As expected, the average market cap grows over time from $1.322 billion in 1978 to $11.12 billion in 1997.

To test the efficiency of the S&P 500 we run the following simulation experiment over a twenty-year period. We implement the mean-variance (MV) optimization technique originally developed by Markowitz (1952, 1959) in which he showed that investors should hold mean-variance efficient portfolios. Starting in 1978, we estimate that year's stock sample mean return, variance and covariance as a rough forecast for the following year's values. These forecasts are than entered into the Markowitz mean-variance estimation algorithm producing the global minimum-variance portfolio, portfolio 1, and four additional portfolios, portfolios 2, 3, 4 and 5, along the efficient frontier.[7] The highest-return portfolio, portfolio 5, is calculated by setting its expected return to the 95th quantile of the forecasted stock returns. The remaining three portfolios are calculated by setting their expected returns so that each is evenly spaced between the returns of portfolio 1 and portfolio 5.

Our optimization is subject to constraints that insure diversification, and, hence, predictive power out-of-sample. We constrain the optimization so that the minimum weight of any given stock is zero[8], no more than 5% of the portfolio can be invested in any one stock[9] and no more

[5] For instance, for S&P 500 stocks from the fourth quarter of 1978 we collect daily returns for both 1978 and 1979.
[6] Filling-in missing data is necessary for estimation of the variance-covariance matrix.
[7] The global minimum-variance portfolio is the portfolio with the lowest possible variance.
[8] This no short-selling constraint is reasonable given the restrictions on money managers and the frictions faced by private investors.

than 20% can be invested in any one industry. The efficiency of the S&P 500 is compared to that of the five optimized portfolios by tracking their out-of-sample performance over the following year.[10] Because portfolio rebalancing only occurs once a year, we do not include transaction costs in the simulation experiment since these costs will not affect portfolio returns significantly.[11]

Figure 1 illustrates that, on average, over the 1979-1998 period the five optimized portfolios dominated the S&P 500 index. The portfolios were more efficient than the S&P 500 since they exhibited on average lower variance and higher return. These results are particularly impressive considering the simple forecasts used for mean, variance and covariance in the optimization. Since our efficient frontier portfolios on average dominate the S&P 500, the common practice of indexing portfolios to the market appears to have been an inefficient investment strategy.

Table 2 reports on the out-of-sample statistics for the five portfolios and the S&P 500. As one moves from the safer portfolio 1 to the riskier portfolio 5, the average return increases monotonically from 17.37% to 25.55% and standard deviation increases from 9.93% to 14.33%. Over the same period, the average S&P 500 return was 18.56% and standard deviation was 14.26%. The S&P 500's average capitalization was $4.376 billion, while the average capitalization of portfolio 1 was $4.154 billion, which decreases monotonically to $2.973 billion for portfolio 5. So the safer efficient portfolios on average contain smaller stocks, while the riskier S&P 500 contains larger stocks. The average number of stocks in portfolio 1 was 53, which decreases monotonically to 38.1 for portfolio 5 as compared with 443.2 for the S&P 500. Hence, in any given year, the number of stocks in the efficient portfolios is a small fraction of the number of stocks in the S&P 500 index.

---

[9] This upper bound on stock weight is well below the largest stock weight in the S&P 500 over our sample period (7.8%).

[10] The S&P 500 is proxied by a capitalization-weighted portfolio of our stock sample.

[11] Even if we assume an annual turnover rate of 100% and a transaction cost of 2 cents per share, the portfolio return would lose an insignificant 0.1% annually.

Figure 2 shows the twenty annual out-of sample graphs of return vs. standard deviation for the five efficient portfolios and the S&P 500. With the exception of 1997, the S&P 500 was dominated by at least one efficient portfolio in every year, but the shape of the efficient frontier deviated from the classic concave shape. In 15 out of the 20 years, the S&P 500 falls inside a well-defined efficient frontier. The five years in which the efficient frontier deviated significantly from a concave shape were: 1981, 1984, 1988, 1994, and 1997. The unusual shapes of the efficient frontier in these five years could only be caused by errors in forecasts of means, variances, and covariances. In the next section we explore how improvements in the forecasts' accuracy influence the efficient frontier portfolios.

3. Estimation Errors and Portfolio Efficiency

The literature examining the benefits and limitations of MV optimization has grown extensively over the years. Michaud (1989) pointed out that MV optimization tends to magnify the effects of errors in estimates. Chopra and Ziemba (1993) examined the relative impact of estimation errors in means, variances, and covariances. They found that errors in means are about eleven times as important as errors in variances and errors in variances are about twice as important as errors in covariances.[12] We examine Chopra and Ziemba's findings in the context of our simulation and we obtain consistent results.

We rerun our simulation for three scenarios with differing levels of perfect return foresight. In scenario A, we assume perfect out-of-sample foresight of expected stock returns, variances and covariances. That is we assume that all relevant information is known about stock returns apriori. In scenario B, we assume that only expected stock returns are known with perfect foresight. The variances and covariances are still estimated, as they were before, from the past year's returns. Finally in scenario C, we assume that variances and covariances are known with perfect foresight and expected returns are estimated as the mean of the past year's returns.

The results for scenario A are illustrated graphically in Figures 3 and 4. As expected, perfect forecasts of means, variances and covariances results in perfectly concave efficient frontiers and in the S&P 500 being grossly inefficient and easily dominated by the portfolios on the efficient frontier. Panel A in Table 3 reports average annual results for the simulation with return varying between 15.8% and 82.68% while standard deviation ranges from 6.83 to 11.66. These results serve as a benchmark against which scenarios B and C are compared.

Figures 5 and 6 illustrate the results for scenario B where expected stock returns are known with perfect foresight and the variance-covariance matrix is estimated from past year's returns. The shapes of the annual efficient frontiers are concave and clearly dominate the S&P 500 in every year. These efficient frontiers are a substantial improvements over the efficient frontier illustrated in Figures 1 and 2. Panel B in Table 3 demonstrates that the five efficient portfolios' returns correspond closely to that of Scenario A, while the standard deviations are closer to those of table 2. Although having perfect foresight of return is not realistic, it is possible to improve on our simplistic model for forecast returns. Implementing a more sophisticated model for forecasted stock returns (such as Fama and French (1993)) would improve our forecasted mean returns and help generate more efficient portfolios.

The results for scenario C are illustrated graphically in Figures 7 and 8. In this scenario we assume that variances and covariances are known with perfect foresight and expected returns are estimated from the mean of past year's returns. The shapes of the efficient frontier do not appear to improve on those in Figures 1 and 2. Not only do the shapes deviate from the classic concave, the efficient frontier portfolios fail to dominate the S&P 500 in four years: 1989, 1995, 1997, and 1998. The results from Scenario C appear to be worse than the simulation using only simple past estimation and no foresight. Panel C in Table 3 shows that mean returns for the five efficient portfolios correspond more closely to those of table 2, while the standard deviations

---

[12] The relative impact of errors depends on investor's risk tolerance, that is, the desired trade-off between extra return and extra risk. The reported results are for a risk tolerance level that corresponds the typical

are closer to those in Scenario A. Knowing the true variances and covariances did not improve our MV estimation significantly. Implementing a more sophisticated model for forecasting variances and covariances (see Chan, Karceski, and Lakonishok (1999)) would not necessarily improve our estimated efficient frontier by itself. Although in conjunction with a better estimation of forecasted mean returns, better variances and covariances would make a difference, as seen in Figures 3 and 4.

Our conclusions match those of Chopra and Ziemba, the effects of errors in variances and covariances seem to be much smaller when compared to the effects of errors in forecasted mean returns on the efficient frontiers. Given limited resources and the desire to move towards the efficient frontier, investors would be wise to emphasis acquiring better forecasts of assets' mean returns than in attempting to get better forecasts of variances and covariances.

4. Conclusions

In this article we examined the long-run efficiency of the most popular market benchmark, the Standard&Poors 500 index. We applied the Markowitz mean-variance optimization techniques to generate efficient frontier portfolios over the period of 1978 to 1998. Although only simple historical return estimates are used, the out-of-sample performance of the efficient frontier portfolios is superior to that of the S&P 500 in both return and risk. Improvements in forecasts of stock's mean returns appear to be more important for achieving more efficient portfolios than improving variance-covariance estimation. Since our efficient frontier portfolios on average dominate the S&P 500, the common practice of indexing portfolios to the market appears to have been an inefficient investment strategy.

---

portfolio allocations of large US pension funds and other institutional investors.

REFERENCES

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model," *Review of Financial Studies*, 12 (1999), pp. 937-974.

Chan, Louis K. C., and Josef Lakonishok, "Are the Reports of Beta's Death Premature?," *Journal of Portfolio Management*, Summer 1993, pp. 51-62.

Chopra, V. K., and William T. Ziemba, "The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, Summer 1993, pp. 6-11.

Ellis, Charles D., "Levels of the Game," *Journal of Portfolio Management*, Winter 2000, pp. 12-15.

Fama, E. F., and K. R. French, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1993), pp. 3-56.

Haugen, Robert A., and Nardin L. Baker, "The Efficient Market Inefficiency of Capitalization Weighted Stock Portfolios," *Journal of Portfolio Management*, Spring 1991, pp. 35-40.

Markowitz, Harry M., "Portfolio Selection," *Journal of Finance*, 7 (1952), pp. 77-91.

Markowitz, Harry M., *Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley & Sons, 1959.

Michaud, Richard O., "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analyst Journal*, 45 (January-February 1989), pp. 31-42.

Table 1: Summary Statistics for S&P 500 Stock Sample Over 1978-1997

Year	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Number Stocks	436	429	419	419	427	416	415	426	435	441
Number of Industries	46	47	47	47	48	48	49	50	50	52
Avg Market Cap ($M)	1322	1468	1915	1812	2082	2530	2335	2991	3451	3406
Max Market Cap ($M)	43571	37642	39714	47551	57113	75251	75527	95223	73904	71184
Min Market Cap ($M)	11.2	10.0	6.9	10.5	27.4	22.8	46.7	44.7	34.1	45.5

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Number Stocks	446	458	460	465	467	465	462	465	459	454
Number of Industries	54	59	60	60	60	61	61	60	59	56
Avg Market Cap ($M)	3654	4467	4152	5244	5538	6029	6090	8099	9820	11120
Max Market Cap ($M)	72696	57484	65008	71324	73962	90457	49386	75101	109978	99829

Min Market Cap ($M)	20.3	35.3	12.9	34.6	74.2	120.2	118.8	205.7	348.3	355.9

Table 2: Summary Statistics for Optimized Portfolios and S&P 500 over 1979-1998

	Average Return (%)	Average Standard Deviation (%)	Average Market Capitalization ($M)	Average Number of Stocks
Portfolio 1	17.37	9.93	4154	53.0
Portfolio 2	19.05	10.35	3401	52.5
Portfolio 3	21.77	11.41	3308	49.8
Portfolio 4	23.47	12.71	3038	43.3
Portfolio 5	25.55	14.33	2973	38.1
S&P 500	18.56	14.26	4376	443.2

Table 3: Summary Statistics for Optimized Portfolios over 1979-1998 for Three Forecast Scenarios

	Panel A: Future Mean Return and Variance-Covariance Matrix are Known		Panel B: Future Mean Return is Known		Panel C: Future Variance-Covariance Matrix is Known	
	Average Return (%)	Average Std. Deviation	Average Return (%)	Average Std. Deviation	Average Return (%)	Average Std. Deviation
Portfolio 1	15.80	6.83	17.37	9.93	15.80	6.83
Portfolio 2	32.75	7.23	33.77	10.43	16.92	7.27
Portfolio 3	49.58	8.23	50.05	11.39	18.37	8.31
Portfolio 4	66.24	9.72	66.16	12.69	20.28	9.78
Portfolio 5	82.68	11.66	82.14	14.19	22.24	11.64

Figure 1: Average Out-Of-Sample Efficiency Over 1979-1998 of Optimized Portfolios vs. S&P 500

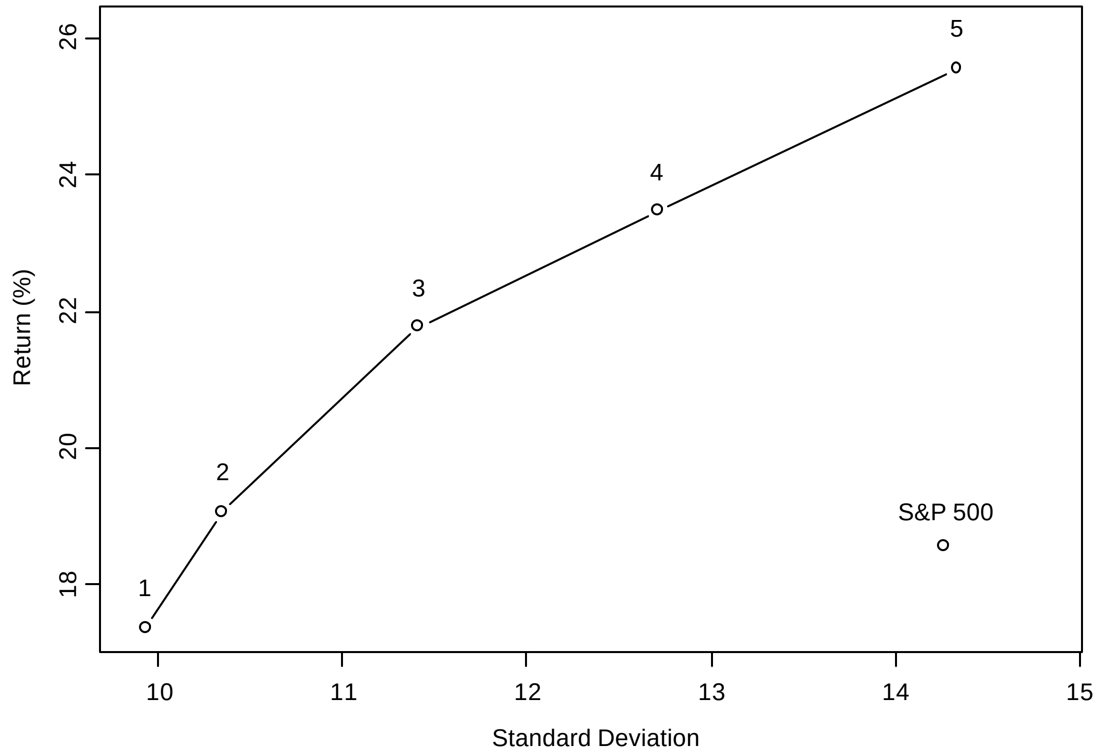

Figure 2: Annual Out-Of-Sample Efficiency of Optimized Portfolios vs. S&P 500

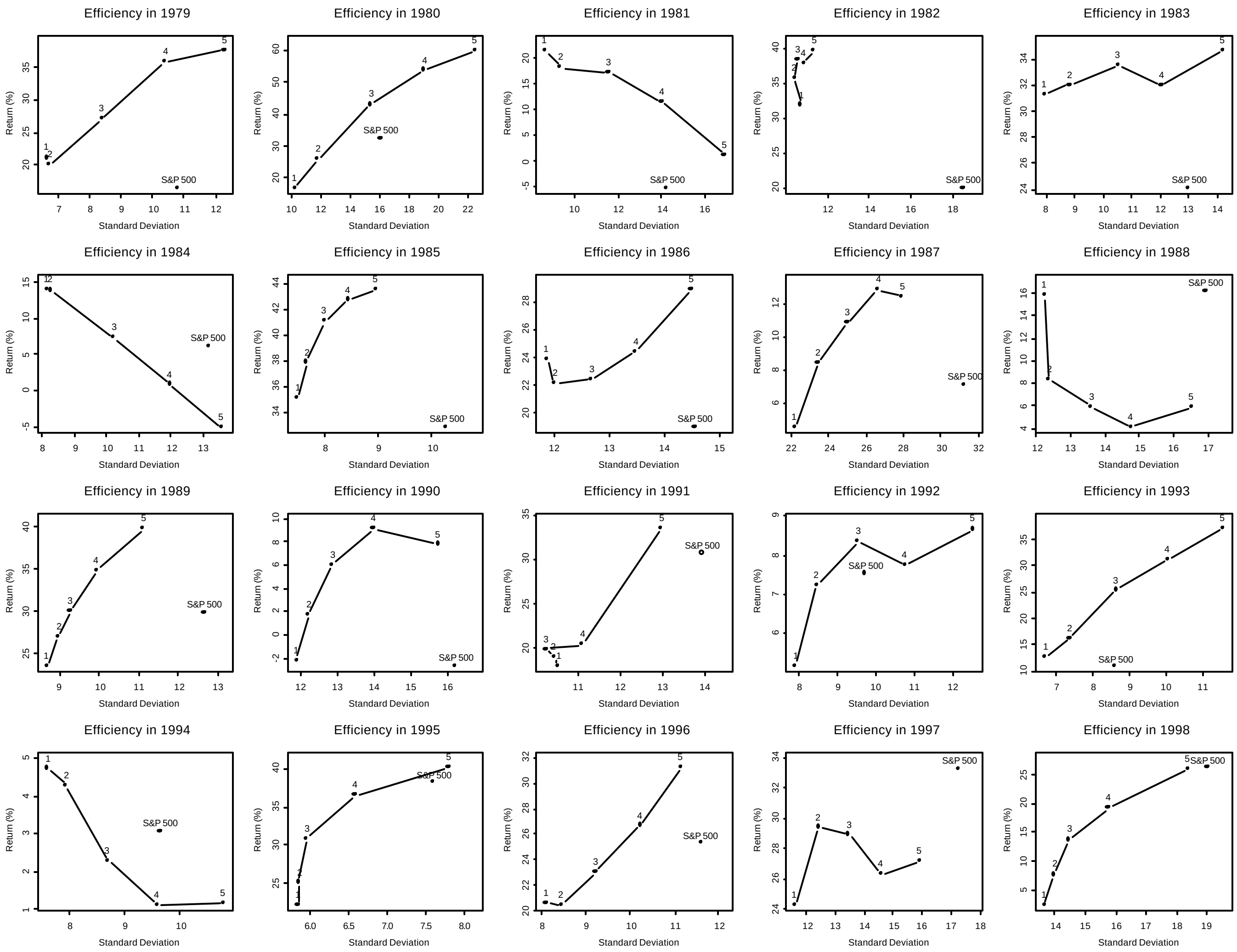

Figure 3: Average Out-Of-Sample Efficiency Over 1979-1998 when Future Mean Return and Variance-Covariance Matrix are Known

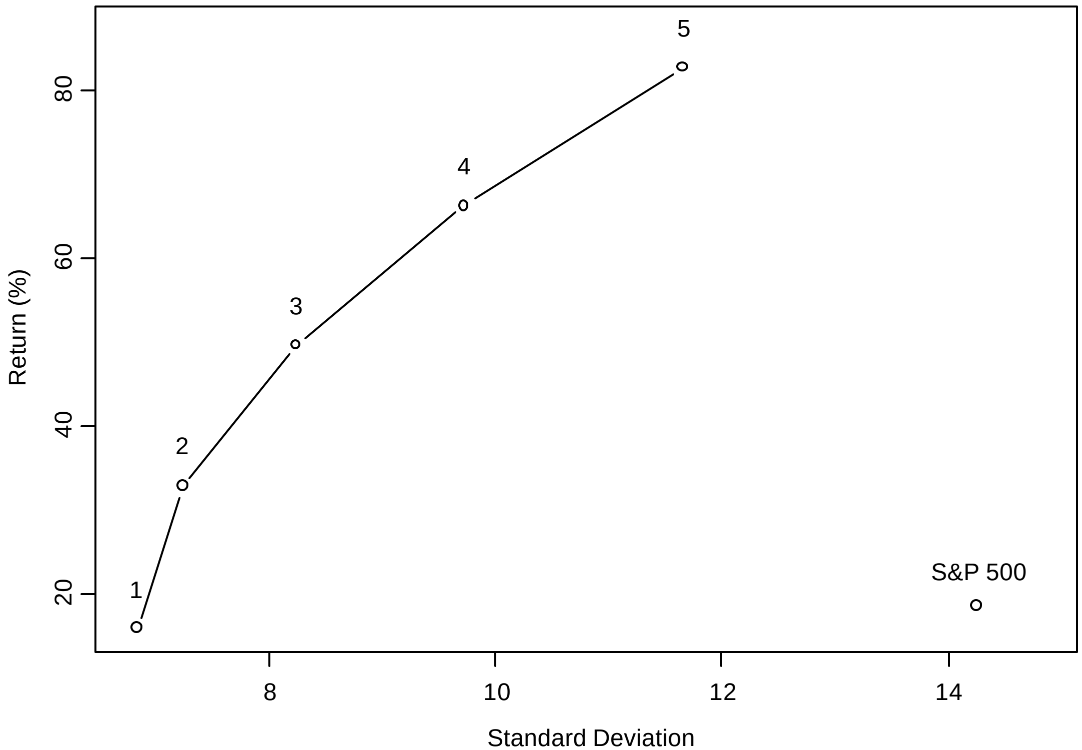

Figure 4: Annual Out-Of-Sample Efficiency when Future Mean Return and Variance-Covariance Matrix are Known

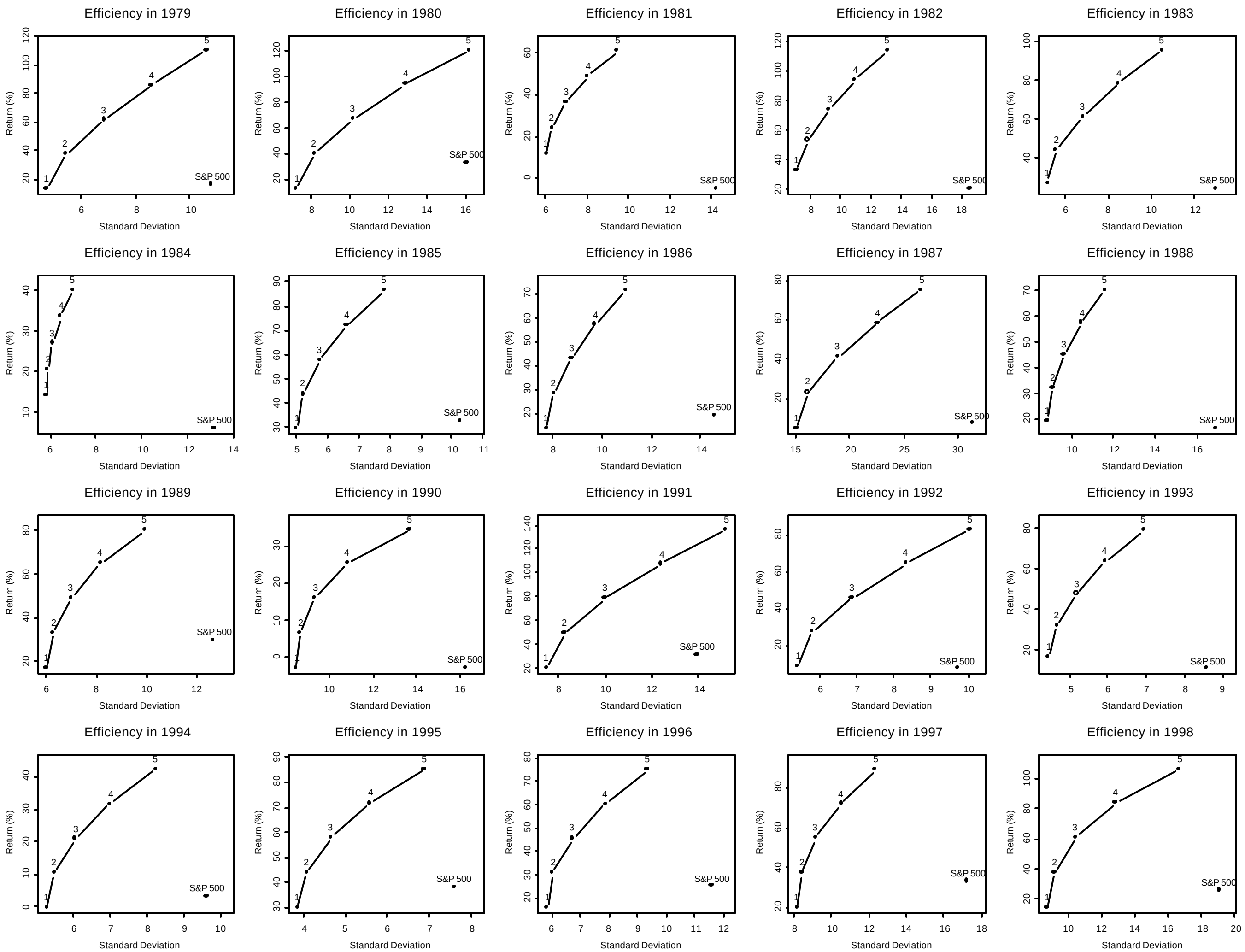

Figure 5: Average Out-Of-Sample Efficiency Over 1979-1998 when Future Mean Return is Known

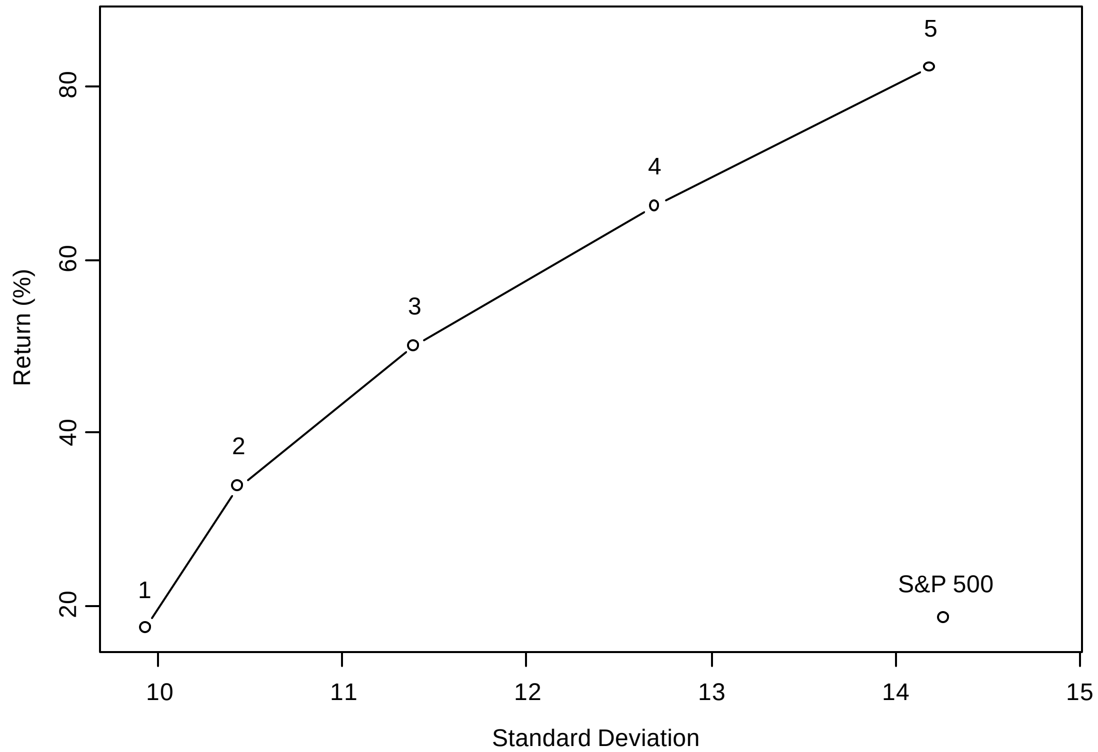

Figure 6: Annual Out-Of-Sample Efficiency when Future Mean Return is Known

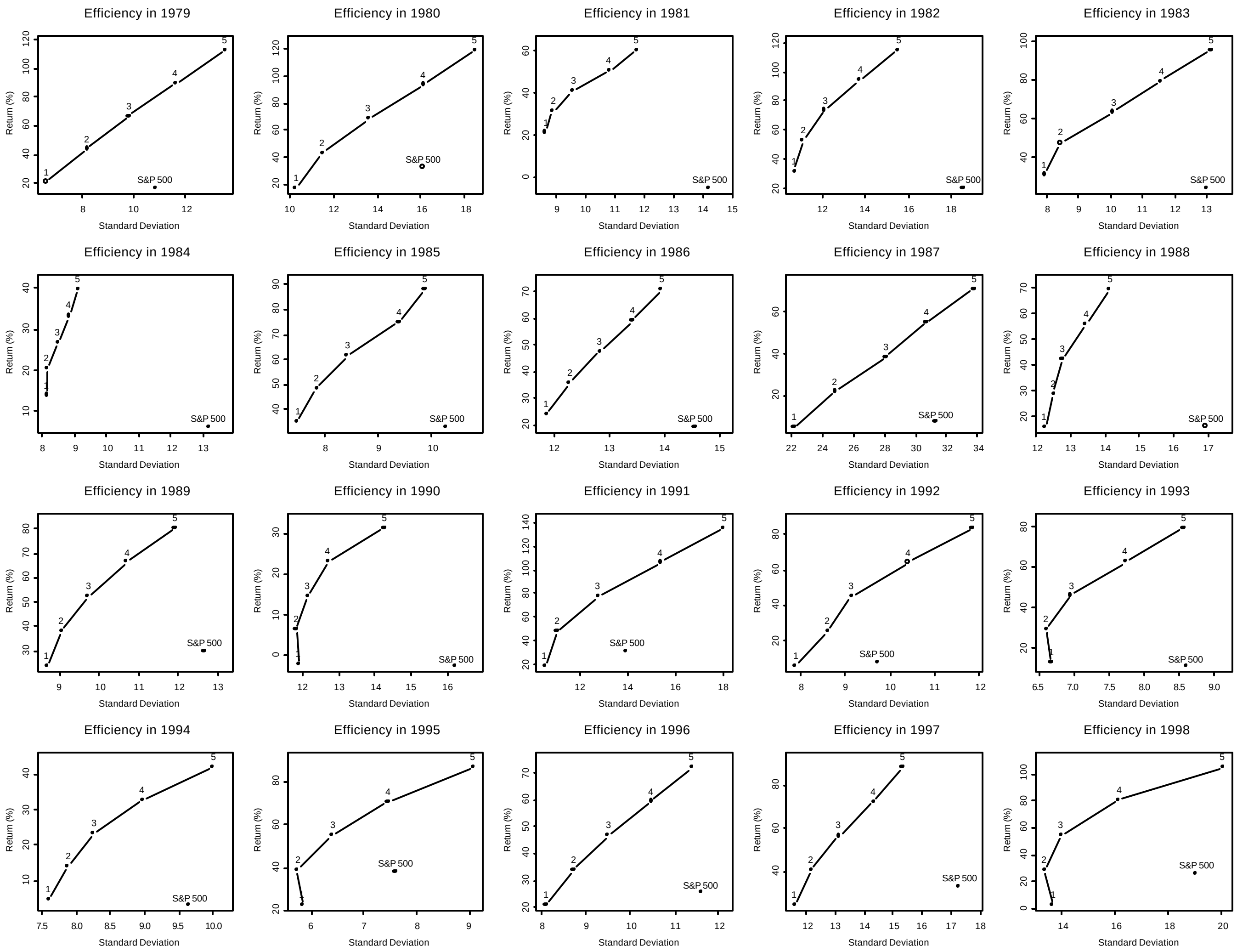

Figure 7: Average Out-Of-Sample Efficiency Over 1979-1998 when Future Variance-Covariance Matrix is Known

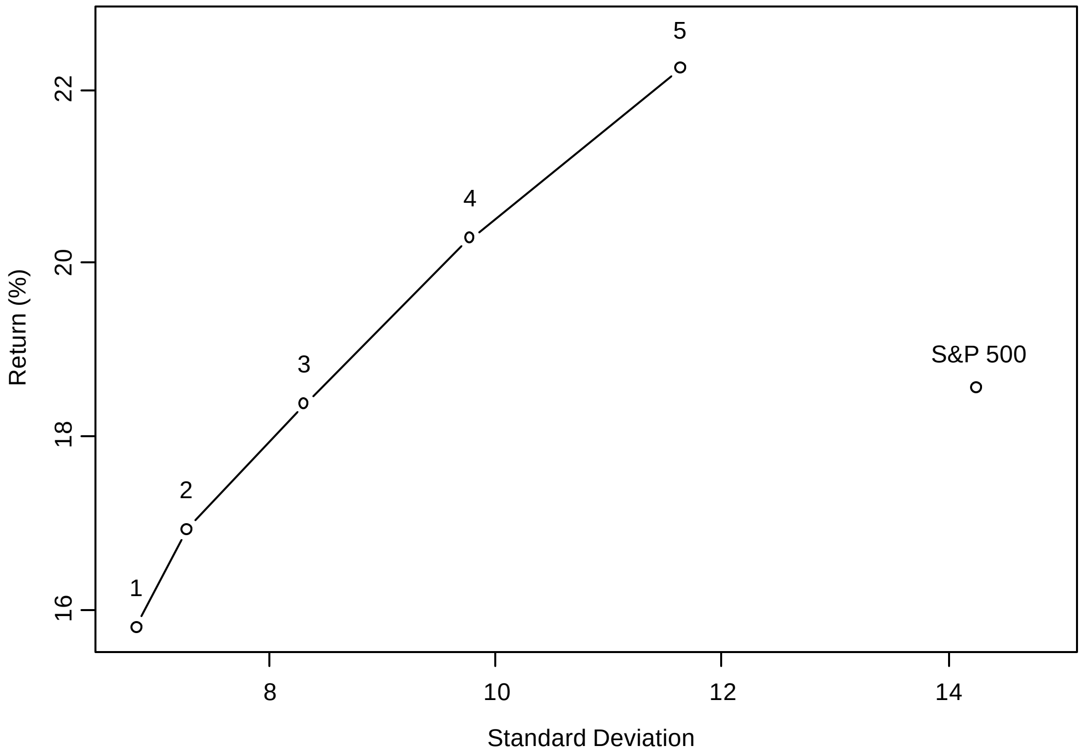

Figure 8: Annual Out-Of-Sample Efficiency when Future Variance-Covariance Matrix is Known