Financial Decisions, Spring 2003, Article 4.

The Improper Use of Dartboard Portfolios as Performance Benchmarks

Randall A. Heron, Glen A. Larsen, Jr., and Bruce G. Resnick
Indiana University, Indiana University, Wake Forest University, respectively.

Abstract

In this paper we empirically study the improper use of dartboard portfolios (i.e., randomly selected portfolios of stocks) as "passive" investment performance benchmarks for active financial decision-making investment strategies. The problems associated with dartboard benchmark portfolios arise because of systematic risk differences, price pressures, and benchmarking errors. Simulations show that dartboard portfolios are, in terms of risk-adjusted performance, dominated by the broader market index fund constructed from the universe of stocks from which the dartboard portfolios are selected. The dartboard portfolios suffer from the lack of diversification due to size and errors in security selection. The dominance of the index fund persists even when we employ Markowitz optimization techniques to identify ex ante efficiently diversified portfolios of the randomly selected stocks.

JEL classification: G10, G14

I. Introduction

In his popular book titled *A Random Walk Down Wall Street,* which has used a picture of a dart sticking in the stock listings on its cover in each edition since 1985, Malkiel [1996, p. 24] states "taken to its logical extreme, the random walk theory means that a blindfolded monkey throwing darts at a newspaper's financial pages could select a "dartboard portfolio" that would do just as well as one carefully selected by the experts." Although Malkiel ultimately advocates investment in unmanaged index funds as opposed to actually creating investment portfolios by throwing darts, his discussion highlights the widely-held belief that in efficient markets, dartboard portfolios will perform as well as portfolios constructed by so-called experts or by using advanced portfolio selection and optimization techniques.

The purpose of this analysis is to highlight the key problems associated with using dartboard portfolios to benchmark return performance under the premise of market efficiency. Although there are many instances where such randomly selected stocks are used as benchmarks, in most cases, the structure of the dartboard selection and evaluation process violates well-established financial decision-making theory. Although it is without question entertaining, *The Wall Street Journal* dartboard contest is a prime example.

Since January 1990, the WSJ has held a monthly contest in which the six-month average performance of four stocks selected randomly by throwing a dart at the stock listings of the WSJ is compared to the average performance of a portfolio of four stocks, where each is selected by a different market professional, and also to the performance of the Dow Jones Industrial Average of 30 stocks.[1] The purpose of the contest is to test the concept of market efficiency. Although the contest design may not be sufficient in this regard, its unique nature has prompted numerous studies directed at both research and practitioner audiences. The intent of such studies ranges from promoting market efficiency to studying how price pressures affect stock prices to reinforcing basic finance theory with regard to the relationship between risk and return. The objective of this study is to provide empirical evidence that highlights the improper use of dartboard portfolios to benchmark the performance of stocks chosen by investment professionals.

Recall that market efficiency implies that nobody can consistently beat the market. Because the top two (out of 4 total) contestants in the WSJ stock-picking contest are invited to participate in the contest again, one can empirically test whether the frequency of repeat winners differs from what would be expected if the success of the contestants resulted from equal probabilities. By equal probabilities, we mean the luck of the draw – as opposed to a true competitive advantage resulting from actual stock-picking expertise. Studies by Metcalf and Malkiel [1994] and Greene and Smart [1999] conduct tests of this nature. Both studies find the observed distribution of repeat winners does not differ than what would be expected by pure chance, or the luck of the draw. These results cast doubt on the notion that some experts have superior stock picking ability and thus support the premise of market efficiency that nobody can consistently outperform dartboard portfolios.

Thomas and Ghani [1996] provide further evidence that suggests that there is little or no additional information contained in the stock picks made by the investment professionals in the WSJ contest. If the investment professionals truly had superior information regarding the future prospects of the stocks they select, one would expect capital markets would recognize this over time. This would, in turn, lead to upward revisions in the forecasted earnings for the companies chosen by the investment professionals. However, Thomas and Ghani [1996] find no evidence that earnings estimates are revised upward in response to the recommendations. This lends strong support to the concept of market efficiency – suggesting that there is an insignificant amount of incremental "information" contained in the selections of the investment professionals.

Viewed collectively, the evidence discussed above builds a strong case for market efficiency and suggests that any 'apparent' superior stock price performance on behalf of the investment professionals in the WSJ dartboard contest is associated with using dartboard portfolios as benchmarks to gauge investment performance.

[1] The dartboard contests actually started in October of 1988. Although the contest horizons were originally 1 month in length, they were changed to six-months in length because price pressures generated significant short-term abnormal returns for the stocks picked by investment professionals (see Barber and Loeffler [1993] and Greene and Smart [1999] for further details).

II. Using Dartboard Portfolios as Performance Benchmarks

The remainder of this paper expands on why one cannot make any inferences regarding market efficiency based on a simple comparison of the average returns of (a) the investment professionals, (b) the stocks selected by throwing darts, and (c) the Dow Jones Industrials as reported in the WSJ's dartboard contest. Dartboard portfolios are improper benchmarks for financial investment decision-making because of systematic risk differences, price pressures, and benchmarking errors. In the end, we conclude that the proper manner with which to benchmark investment performance is to use a risk-adjusted approach where the strategy of "passive investment" is represented by a broad market index return rather than portfolios of randomly selected stocks.

The Role of Systematic Risk

Sharpe [1964] introduced the theory behind the Capital Asset Pricing Model (CAPM) in 1964 in a classic *Journal of Finance* article titled "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." Included among the contributions of this study is the notion that unsystematic risk can be diversified away in a portfolio context, leaving only systematic risk. Because investors can construct portfolios where unsystematic risk has been effectively eliminated, the compensation for risk in the capital market should depend solely upon the component of risk that is not diversifiable -- in other words, systematic risk. This is indeed the insight that is captured in Sharpe's CAPM, which specifies that expected returns are a function of systematic risk.

In this regard, neither the dartboard portfolio, nor the portfolio of stocks chosen by the investment professionals in the WSJ contest is a well-diversified portfolio that eliminates most unsystematic risk. Although the work of Evans and Archer [1968], Elton and Gruber [1977], and Statman [1987] shows that total portfolio risk (variance) is reduced by naïve diversification of unsystematic risk, Statman shows that the number of randomly selected stocks needed to form a portfolio with little remaining unsystematic risk is approximately 30 to 40. Thus, the 4-stock portfolios compiled by averaging the picks of the investment professionals or by the dartboard selections still contain significant amounts of unsystematic risk and would therefore exhibit much greater return volatility than a diversified portfolio. Moy [1994] provides evidence of this when he reports that for the WSJ dartboard contests held between 1990 and 1993, the DJIA, which has less unsystematic risk because of greater diversification than either the portfolios selected by the investment professionals and by the darts, was much more likely to finish second in the dartboard contest than the riskier (higher variance) 4-stock portfolios of selected by either the professionals or the darts. Specifically, the DJIA finished 2nd in roughly two-thirds of the dartboard contests analyzed, where performance is based solely on average returns un-adjusted for risk.

While the portfolios of the investment professionals and of the dartboard selections contain significant amounts of unsystematic risk, it is the *systematic risk* differential between the picks of the investment professionals, the picks made by darts, and the DJIA that prevents any meaningful comparisons on the basis of average returns for each category. Fama [1965a, 1965b], who perhaps first popularized the idea of randomly selected stock portfolios in his early work on the behavior of security prices, was careful to note that a proper test of the performance of a randomly selected portfolio versus the performance of one selected by a professional money manager would have to be done in such a way that both portfolios had the same degree of riskiness. In the WSJ contest, there is no mechanism to insure that the dartboard and the professionals' portfolios will have the same level of systematic risk. Fama [1976] reports that a randomly selected portfolio of 30 stocks has a portfolio beta of approximately unity. Based on this, one would expect that on average, the beta of the dartboard portfolios in the WSJ contest would be close to one. However, the logical strategy for an investment professional who is being judged on the basis of their stock picking ability is to pick a stock with a high beta because higher systematic risk levels imply higher expected returns.

Metcalf and Malkiel (1994) show that over the period from 1990 to 1992, the stocks chosen by the investment professionals do in fact have higher levels of systematic risk than do the dartboard portfolios. Specifically, they report an average beta of the expert's selections of 1.401, which is significantly different (t-statistic of 4.46) from a beta of unity that would be expected for a well-diversified portfolio or perhaps a broad market index. Moreover, the average beta estimate (1.065) for the dartboard selections did not statistically differ from unity (t-statistic of 0.75). We estimate betas for these two groups of stocks over a much larger time horizon and find similar, quite predictable results. Specifically, over the period from January 1990 through December 2000, the *aggregate* beta for the professionals' portfolio was 1.20 when regressed against the CRSP value-weighted market index. By comparison, the dartboard portfolios had an aggregate beta of 0.97 when regressed against the same benchmark. Collectively, these findings imply that the higher average return performance of the stocks selected by the investment professionals is an artifact of their rational choice to enhance their likelihood of winning the contest by choosing stocks with higher levels of systematic risk – and thus higher expected returns. This strategy is obviously predictable based on an application of Sharpe's Capital Asset Pricing Model (CAPM).

Moy (1994) provides further evidence that the stocks chosen by the investment professionals have higher levels of systematic risk when he reports that the win percentage of investment professionals is much higher in up markets than it is in down markets. Such a finding also buttresses arguments in favor of market efficiency. The reason is simple: if the investment professionals in the contest had better information regarding the future direction of the overall market, they would rationally pick low beta stocks either during, or leading up to, periods of market decline. Such a strategy would lower exposure to downward market forces and thereby increase the likelihood of winning the contest in periods of low overall market performance. As Moy points out, the lack of evidence in this regard suggests that the investment professionals that participate in the dartboard contest are not adept in market timing.

The Role of Price Pressure

A second problem with the WSJ dartboard contest, and for that matter, potentially any contest that measures performance *after* publicly disclosing the picks of "investment professionals" arises because of price pressures in the stocks picked by the professionals. Specifically, the public disclosure of the stocks chosen by the investment professionals creates a demand for the stock by investors who believe that the investment professionals are convinced that the stock will take off in the near future. This heightened demand for the picks of the investment professionals creates price pressure that produces 'abnormal returns' for their stocks. There is obviously no reason for this phenomenon to occur with the stocks selected randomly by darts and used as a performance benchmark.

Barber and Loeffler [1993] were perhaps the first to document this phenomenon as it relates to the WSJ dartboard contest. Although several researchers have documented abnormal returns for stocks that are recommended by analysts, Barber and Loeffler show that the abnormal returns associated with the picks of the investment professionals in the WSJ dartboard contest (approximately 4%) are roughly twice as large as what is found in typical studies of analyst recommendations. Moreover, Barber and Loeffler document a doubling in the demand for the stocks chosen by the investment professionals in the two days following their revelation in the WSJ contest. Thus, an atypical demand for the stocks, resulting from the public disclosure of the investment professionals' choices, creates price pressure that generates an abnormal return.

Barber and Loeffler further document that most of the abnormal return performance dissipates within 25 days of the announcement. This leads them to conclude that although there may be information content in the picks of the investment professionals, much of the positive returns to the stocks chosen by the investment professionals result from naïve buying pressure, and are thus, transitory in nature. As we note earlier, the significant short-term price pressures such as those discussed by Barber and Loeffler are one of the primary reasons the WSJ contest period was lengthened from 1 month to 6 months in 1990.

Greene and Smart [1999] analyze the price and volume behavior of the stocks chosen by the investment professionals in the WSJ dartboard contest to test theoretical models that predict a positive relation between noise trading and liquidity. Like Barber and Loeffler, they report that the majority of abnormal returns experienced by the stock picks of the investment professionals disappear within a few weeks. However, they also find that both abnormal trading volumes and abnormal returns are greater for the stocks chosen by investment professionals who have previously won the WSJ contest – even though, as we note earlier, there is no compelling evidence to suggest that previous contest winners fare better than their first-time counterparts. Greene and Smart also report a positive relation between the magnitude of abnormal returns and the extent of abnormal trading volume. Thus, if Moy's [1994] suggestion that the investment professionals will tend to pick smaller stocks is accurate, because smaller stocks tend to have lower trading volumes, the price pressures that produce abnormal returns when the stock picks of investment professionals are revealed would be exacerbated.

What's the Appropriate Benchmark?

In addition to the benchmarking problems we discuss earlier that result from comparing the performance of the stocks picked by the investment professionals to that of randomly selected stocks, additional benchmarking problems result when the returns to the stocks picked by the investment professionals are compared to the returns to the Dow Jones Industrial Average (hereafter, DJIA). First, although the DJIA is a price-weighted index of 30 stocks and it would have much less diversifiable risk than the 4-stock portfolios of the investment professionals or those created by the darts, it is still not an appropriate benchmark because the stocks that comprise the DJIA represent only a small subset of the potential stocks that can be selected by the professionals or by throwing darts. Specifically, the selections made by the experts and those made randomly by "throwing darts" come from the universe of stocks listed on the NYSE, AMEX, or trading on NASDAQ. One consequence is that the DJIA would represent the performance of stocks that are on average, much larger in size than the typical selections made by both the investment professionals and by the process of throwing darts. In addition, the DJIA would obviously have a lower level of systematic risk than the selections made by the investment professionals, which would imply higher expected returns for the investment professionals.

As evidence in this regard, when we estimate the *aggregate* beta for the professionals' portfolios from January 1990 through December 2000, using the DJIA as the proxy for the market index, we find a beta of 1.28. This implies that the stock picks of the investment professionals have 28% more systematic risk, and therefore, should obviously command a higher expected return than the DJIA. Viewed in this light, the fact that the investment professionals have a higher win percentage than the DJIA (via comparison of raw returns) is consistent with the predictions of Sharpe's CAPM and is therefore evidence in favor of market efficiency. As we argue below, a better benchmark of overall market performance would be a risk-adjusted measure relative to that for the entire universe of the firms from which the investment professionals can choose.

III. Randomly Selected Portfolios vs. Market Indices

If capital markets are relatively efficient, do randomly selected portfolios of reasonable size provide good benchmarks? Recall that the theory underlying Sharpe's CAPM implies that the most efficient portfolio is the overall market portfolio where each asset's portfolio weight is based on its relative market value. Because investors can passively invest in broad index funds (value-weighted) with minimal transaction costs, it is natural to consider such an index fund to be the "benchmark" when comparing the risk-return efficiency of both actively managed portfolios and when testing whether valid comparisons of investment performance can be made using randomly selected portfolios (perhaps constructed by throwing darts) of reasonable size.

It follows that if well-diversified dartboard portfolios are to be used as representative benchmarks of investment performance, their risk-return performance should parallel that of common index funds. Although several studies suggest that most of the risk-reduction benefits of diversification occur with modestly sized portfolios, with perhaps as few as 30 to 40 securities (see Statman [1987] for further details), overall market returns are increasingly being driven by a handful of top-performing stocks. As an example, Ibbotson Associates calculated how much lower the return on the Standard and Poor's 500 index would have been if the 50 top performing stocks were eliminated (see Clements [1998]). Over the 12-year period studied, on an annual basis, without the 50 top-performing stocks, the index's return would have been reduced by approximately 50%. Similarly, according to figures attributed to Ned Davis Research and cited by Brammer (2000), over the period from June 30th, 1999 through March 10th, 2000, the extreme performance of only 10 individual stocks accounted for 44% of the total returns to the Russell 2000 growth index.

This phenomenon, where broader market returns are driven by a relatively small subset of stocks, arises because of the considerable skewness in the individual stock returns when they are measured over investment horizons such as years. In simple terms, the skewness results from the fact that the most an investor can lose from investing in a stock is 100% of the original investment, whereas the upside potential is infinite. Because of this, the return on a typical, randomly selected stock (i.e., a stock that exhibits the median level of observed performance) will tend to be less than the return to a broad market index. As evidence of this phenomenon relating to the WSJ dartboard contest, in the 116 contests from January 1991 through December 2000, the return of the dartboard portfolios exceeded that of the DJIA in only 39% of the contests (45 out of 116) and that of the broader market – as measured by the CRSP market-value-weighted index in only 34% (39 out of 116) of the contests. Assuming a uniform distribution (i.e., normal returns with no skewness) and no significant difference in systematic risk, one would expect 50%, or about 58 of the dartboard portfolios to produce a return that is superior to the index.

Carhart's (1997) analysis of the persistence in mutual fund performance indicates that Jegadeesh and Titman's (1993) one-year momentum in stock returns accounts for Hendricks, Patel, and Zechhauser's (1993) hot hands effect in mutual fund performance. Funds that earn higher one-year returns, however, do so not because fund managers successfully follow momentum strategies, but because some mutual funds just happen by chance to hold relatively larger positions in last year's winning stock. Ikenberry, Shockley, and Womack (1998) provide convincing evidence that in addition to transactions costs, size-effects and return skewness contribute to the fact that most active fund managers underperform passive benchmarks such as the S&P 500.[2] Specifically, with regard to the effects of skewness, they simulate "managed" portfolios of various sizes selected from the S&P 500 universe and report that for portfolios of 15 stocks, the mean return exceeds the median return by an average of approximately 40 basis points per year. Such a finding shows that skewness does play a role in the return performance

[2] Clements [2000] reports the results of Vanguard research that finds that only 28% of managed U.S. stock funds outperformed the Wilshire 5000 index.

of portfolios that hold a limited number of stocks – as the typical (i.e., median) small portfolio is likely to underperform a broader market index when the portfolio's members are randomly selected from the underlying benchmark.

In the following simulations, we shed additional light on the importance of return skewness and how it relates to benchmarking portfolios by analyzing the total return performance and the risk-adjusted performance of randomly selected (dartboard) portfolios versus the performance of the broader market index from which the stocks were selected. For comparison purposes, we also examine the performance of ex ante efficiently diversified portfolios of stocks selected at random from the same population. Our simulation results show that dartboard portfolios of reasonable size are, in terms of risk-adjusted performance, dominated by a broader market index constructed from the universe of stocks from which the dartboard portfolios can be selected. The dartboard portfolios suffer from the lack of diversification due to size and errors in security selection. This implies that when testing the performance of stock picks made by investment professionals, the appropriate benchmark representing a passive investment strategy should be a broad-market index representing the universe of stocks from which the investment professionals can make their choices (rather than dartboard portfolios), and that the resulting comparisons should be on a risk-adjusted basis. We also include comparable performance figures for ex ante efficiently diversified portfolios of randomly selected stocks. In this regard, our results raise questions as to whether Markowitz optimization can be successfully implemented with individual securities to overcome the risk-adjusted return performance advantage of the broader market index that arises because of return skewness.[3]

IV. Data and Methodology

Simulation Data

The data we use in our simulations are from the University of Chicago's Center for Research in Security Prices (CRSP). From CRSP, we obtained monthly returns on all New York Stock Exchange (NYSE) listed firms for which returns existed every month over the 20-year period beginning in January 1981 and ending December, 2000.[4] This produced a sample of 543 firms. We also obtained the corresponding monthly Treasury bill returns (30-day) and the monthly portfolio returns for the population of NYSE-listed firms (market-value weighted). The market value weighted returns for the population of NYSE-listed firms (hereafter referred to as MVW index) serves as the benchmark "passive" portfolio as all stocks that can be selected in our simulations are drawn from this population set.

[3] Larsen and Resnick [2001] suggest that it is possible that ex ante efficiently diversified portfolios can consistently achieve enhanced returns at much the same level of return per unit of risk as an index fund if the constituent assets, instead of being randomly selected, are concentrated by some characteristic of the assets that generates return persistence, such as firm size.

[4] We analyze only those firms with continuous monthly returns because the optimization techniques we investigate require no missing data.

Simulations

We performed 100 simulations per year, over a fifteen year period, where in each simulation, portfolios were constructed by randomly selecting securities from the 543 stocks using the IMSL FORTRAN subroutine RNUND, which generates pseudorandom numbers from a uniform distribution. In each simulation, we constructed portfolios of 5, 20, and 40 securities. The portfolios were constructed so that the first five securities in the 20-stock portfolio were the same as the five securities in the respective 5-stock portfolio, etcetera, and the first 20 stocks in a 40-stock portfolio were the same as the 20 stocks in the respective 20-stock portfolio.

For each random selection of 5, 20, and 40 stocks, we constructed four different portfolios: an equally weighted portfolio, and three Markowitz tangency portfolios. The equally weighted portfolio (EQW) produces the returns that would result from an investment of equal dollars in stocks selected randomly, say for example, from throwing darts at a list of the NYSE stocks. One of the three Markowitz tangency portfolios was constructed based on using historical estimates (HIS) of the input parameters to calculate the ex ante optimal investment weights, whereas the other two recognize parameter uncertainty, or estimation error. Estimation error in the input parameters was controlled by using the minimum-variance portfolio (MVP) technique of Jobson and Korkie [1980, 1981] and the Bayes-Stein tangency (BST) technique derived by Jorion [1985, 1986]. The MVP technique selects the historical minimum-variance portfolio's investment weights as the ex ante optimal investment weights. The BST technique uses an empirical Bayes technique for estimating the expected return vector in calculating the ex ante input parameters to solve the portfolio problem. Next, we tabulated the return performance for the simulated portfolios over 15 non-overlapping annual periods beginning with January, 1986 and ending in December, 2000. This produced a total of 1500 simulations (15 years * 100 simulations per year). Input parameters for the Markowitz tangency portfolios were estimated using the 60 months of returns immediately prior to the assumed portfolio inception date.

V. Empirical Results and Analysis

Tables 1, 2, and 3 present the average results for the 1500 simulations (100 in each of the 15 non-overlapping, out-of-sample, annual holding periods) for the 5-stock (Table 1), 20-stock (Table 2), and 40-stock portfolios (Table 3) constructed according to the EQW, HIS, MVP, and BST techniques. These results can be directly compared to the portfolio return (market-value weighted) of the overall population from which all securities were selected (NYSE-listed firms) -- referred to as the MVW index. The figures we report in each table include, for each portfolio selection technique; the average monthly portfolio return standard deviation (σ_p), the average monthly portfolio excess return (R_p) -- where excess return is defined as the return in excess of the risk-free return on 30-day T-bills, the minimum and maximum monthly portfolio excess return from the 1500 simulations (15 years x 100 simulations per year), the number of times out of 1500 that the mean monthly portfolio excess return was greater than the corresponding MVW index mean monthly excess return (R_I), the average Sharpe (SHP_p) measure of portfolio performance, the minimum and maximum SHP_p values, and the number of times out of 1500 total that the portfolio's Sharpe ratio (SHP_p) was greater than the Sharpe ratio for the MVW index (SHP_I).

Table 1 presents the results when five stocks are randomly selected. As is evident in the table, the average monthly portfolio return standard deviation for each of the four portfolio construction techniques is significantly larger (< 1% level of significance) than the average monthly return standard deviation of 3.8424 percent for the MVW index. Among the portfolio selection techniques, the HIS technique produces the largest average portfolio standard deviation, whereas the MVP technique produces the smallest. The average monthly return standard deviation of the MVW index is only .76 (= 3.8424/5.0639) of the average for the four portfolio construction techniques, again confirming that regardless of how they are constructed, portfolios of only five stocks contain substantial diversifiable risk. Of the four portfolio construction techniques, the EQW had the largest mean monthly portfolio excess return, .9024 percent per month, and the HIS technique had the largest range of portfolio excess returns. By comparison, the MVW index had a mean excess return of .9582 percent. Moreover, all of the four portfolio selection techniques were also dominated by the MVW over the majority of the simulations. Under the null hypothesis where the likelihood that a portfolio selection technique will beat the MVW index for any given draw is 0.50, one can reject the null at the 5% level when the number of times that the portfolio beats the MVW index is less than 711 (out of 1500). The corresponding figure for a 1% level of significance is 700 (out of 1500).

In terms of risk-adjusted performance, each portfolio construction technique produced a substantially smaller average SHP ratio than the average for the MVW index. The best technique was the MVP, which generated a larger SHP in comparison to the index in 559 out of 1,500 simulations. Overall, the annual results for 5-stock portfolios are far inferior to investment in a well-diversified index fund. All statistics related to comparisons of Sharpe ratios between the 5-stock portfolios and the MVW index indicate significance (i.e., dominance of the index) at beyond the 1% level.

Table 2 shows the results when 20 stocks are randomly selected. The average portfolio return standard deviations for the EQW, HIS, and the BST portfolio construction techniques are all only slightly larger than the return standard deviation for the MVW index, whereas for the MVP technique it is significantly smaller (1% level). The average standard deviation of the MVW index over the 15 annual holding periods is .96 (= 3.8424/4.0093) of the average over the total of 6,000 simulations for the four portfolio construction techniques.

Despite the similarity in return standard deviations, the MVW index dominates all of the 20 stock portfolios in terms of mean excess returns (.9582% mean excess return per month). Among the portfolio selection techniques, the EQW technique had the largest mean monthly portfolio excess return, .8720 percent per month, whereas the HIS technique had the largest range of monthly portfolio excess returns. All of the monthly return deficiencies for the 20-stock portfolios relative to the MVW index are significant at 5% levels or beyond. In addition, none of the four techniques yielded a greater monthly excess return than the MVW index in the majority of the simulations. In this regard, the best selection technique was the EQW, which outperformed the index in 696 out of the 1,500 simulations. This is still significantly less than the 750 (1500*.5 = 750) that would be expected under the assumption of equal probabilities at the 1% level of statistical significance (z-statistic = -2.58).

Although the simulation results are consistent with the extant literature with respect to randomly selected portfolios of 20 stocks having little remaining diversifiable risk, the 20-stock portfolios do not compare favorably to the MVW index in terms of risk-adjusted performance. As evidence, regardless of the final portfolio construction technique (equal weight or Markowitz optimization), all of the 20-stock portfolios had significantly lower (1% level of significance) Sharpe ratios than the MVW index. The best portfolio selection technique in terms of the number of times it had a higher SHP ratio than the MVW index was the MVP optimization, which generated a larger SHP in comparison to the index in 639 out of 1,500 simulations. Nonetheless, this fraction is still significantly less than the null under the assumption of equal probabilities at beyond the 1% level.

Table 3 presents the corresponding results when 40 stocks are randomly selected. For the sake of brevity, and because the conclusions are quite similar, we broadly interpret the results in Table 3 as opposed to discuss the specific details as we did in the preceding tables. Overall, the figures reported in Table 3 confirm the fact that it is exceedingly difficult to outperform a well-diversified index fund on a risk-adjusted basis with portfolios of randomly selected stocks. This conclusion remains robust when the number of securities in the randomly selected portfolios is large enough so that diversifiable risks have been virtually eliminated, and also when Markowitz optimization techniques are employed in an attempt to enhance performance.

VI. Summary and Conclusion

This study highlights the pitfalls of using dartboard portfolios to represent "passive" investment performance when benchmarking the returns to active investment strategies – such as the stock picks of investment professionals. We begin with a consolidated discussion of the extant literature relating to *The Wall Street Journal's* monthly dartboard contest. We discuss three major pitfalls associated with such comparisons, including 1) no controls for the levels of systematic risk, 2) price pressures in the stock picks of investment professionals, and 3) the use of inappropriate benchmark returns rather than the returns to "passive" investment in a broad market index. With regard to the latter, we perform simulations that show that dartboard portfolios (i.e., randomly selected) of reasonable size are, in terms of risk-adjusted performance, dominated by the broader index fund constructed from the universe of stocks from which the dartboard portfolios are selected. We attribute much of the index fund's dominance relative to dartboard portfolios to the skewness in stock returns. The dominance of the "index fund" persists even when we employ Markowitz optimization techniques to identify ex ante efficiently diversified portfolios of the randomly selected stocks.

Our results imply that when evaluating the performance of portfolios of stocks chosen by investment professionals, the “passive” benchmark should be a market index that spans the population of stocks from which investment professionals can make their choices (rather than dartboard portfolios or the DJIA – as are used in the WSJ’s dartboard contests), and the resulting comparisons should be on a risk-adjusted basis. Moreover, our findings raise questions as to whether Markowitz optimization techniques can be used with individual securities to construct ex ante efficiently diversified portfolios that overcome the risk-adjusted return performance advantage of the broader market index that appears to be attributable to return skewness.

References

Barber, Brad M. and Douglas Loeffler (1993), "The Dartboard Column: Second-Hand Information and Price Pressure." *Journal of Financial and Quantitative Analysis* 28, pp. 273-284.

Bodie, Zvi, Alex Kane and Alan J. Marcus (1999), *Investments*, 4th edition. Boston: Irwin/McGraw-Hill.

Brammer, Rhonda (2000), "Mourning After." *Barron's*, April 24.

Carhart, Mark M. (1997), "On Persistence in Mutual Fund Performance." *Journal of Finance* 52, pp.57-82.

Clements, Jonathan (1998), "Highfliers Propel the Market Averages, But Trying to Pick Them Can be Perilous." *The Wall Street Journal*, September 15, pp. c.1.

Clements, Jonathan (2000), "Criticisms of Indexing Don't Hold Up." *The Wall Street Journal*, April 25, pp. c.1.

Elton, Edwin J. and Martin J. Gruber (1977), "Risk Reduction and Portfolio Size: An Analytical Solution." *Journal of Business* 50, pp. 415-437.

Evans, John L. and Stephen H. Archer (1968), "Diversification and the Reduction of Dispersion: An Empirical Analysis." *Journal of Finance* 23, pp. 761-767.

Fama, Eugene F. (1965a), "The Behavior of Stock Market Prices." *Journal of Business* 38, pp. 34-105.

Fama, Eugene F. (19165b), "Random Walks in Stock Market Prices." *Financial Analysts Journal* 21, pp. 55-59.

Fama, Eugene F. (1976), *Foundations of Finance*. New York: Basic Books.

Greene, Jason and Scott Smart (1999), "Liquidity Provision and Noise Trading: Evidence from the 'Investment Dartboard' column." *Journal of Finance* 54, pp. 1885-1899.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser (1993), "Hot Hands in Mutual Funds: Sort-Run Persistence of Performance." *Journal of Finance* 48, pp. 93-130.

Ikenberry, David, Richard Shockley, and Kent Womack (1998), "Why Active Fund Managers Often Underperform the S&P 500: The Impact of Size and Skewness." *Journal of Private Portfolio Management* 1, pp. 13-26.

Jegadeesh, Narasimhan, and Sheridan Titman (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48, pp. 65-91.

Jobson, J. D. and Bob M. Korkie (1980), "Improved Estimation and Selection Rules for Markowitz Portfolios." Paper presented at the Western Finance Association Annual Meetings.

Jobson, J. D. and Bob M. Korkie (1981), "Putting Markowitz Theory to Work." *Journal of Portfolio Management* 7, pp. 70-74.

Jorion, Philippe (1985), "International Portfolio Diversification with Estimation Risk, *Journal of Business* 58, pp. 59-78.

Jorion, Philippe (1986), "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis* 21, pp. 279-92.

Larsen, Glen A. and Bruce G. Resnick (2001), "Parameter Estimation Techniques, Optimization Frequency, and Equity Portfolio Return Enhancement, *Journal of Portfolio Management*, 27, forthcoming.

Malkiel, Burton G. (1996), *A Random Walk Down Wall Street*. New York: W. W. Norton.

Metcalf G.E. and Burton G. Malkiel (1994), "The Wall Street Journal Contests: The Experts, the Darts, and the Efficient Market Hypothesis." *Applied Financial Economics* 4, pp. 371-374.

Moy, Ronald (1994), "Pros, Darts, and the Dow: Portfolio Lessons from *The Wall Street Journal*." *Financial Practice and Education* 4, pp. 144-147.

Sharpe, William F (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19, pp. 425-442.

Sharpe, William F (1970), *Portfolio Theory and Capital Markets*. New York: McGraw-Hill.

Statman, Meir (1987), "How Many Stocks Make a Diversified Portfolio?" *Journal of Financial and Quantitative Analysis* 22, pp.353-363.

Thomas, Martin R. and Waqar I. Ghani. "The Dartboard Column: Analysts' Earnings Forecast and the Information Content of Recommendations." *Journal of Business and Economic Studies* 3, pp. 33-42.

Table 1

Summary Results from 1500 Simulated Portfolios of 5 Randomly Selected Stocks

The table presents average monthly summary figures for 1500 simulated portfolios of 5 stocks randomly selected from the NYSE. The simulation of annual portfolio holding period returns (100 per year * 15 years) begins in 1986 and extends through 2000. *EQW* represents an equally weighted portfolio. *HIS* represents an ex ante optimized portfolio based on historical estimates of the input parameters. *MVP* represents an ex ante optimized portfolio where estimation error in the input parameters was controlled for using the minimum variance portfolio technique of Jobson and Korkie [1980, 1981]. *BST* represents an ex ante optimized portfolio where estimation error in the input parameters was controlled for using Bayes-Stein tangency technique derived by Jorion [1985, 1986]. *Total* presents the average across all portfolio selection techniques. *MVW Index* represents the value-weighted index of NYSE-listed stocks. σ_p is the average monthly portfolio return standard deviation, R_p is the average monthly portfolio excess return, R_I represents the excess return on the MVW NYSE index, and SHP_p and SHP_I represent the Sharpe ratios of return performance for the individual portfolios and the MVW NYSE index, respectively. [a] and [b] designate a significant difference from the corresponding figure for the MVW Index at the 1% and 5% levels, respectively. Standard deviations are in parentheses.

Technique	σ_p	R_p	*Min* R_p	*Max* R_p	*No.* $R_p > R_I$	SHP_p	*Min* SHP_p	*Max* SHP_p	*No.* $SHP_p > SHP_I$
EQW	5.2523[a] (2.0906)	0.9024 (1.4294)	-5.4048	8.1402	700[a]	0.2097[a] (0.3033)	-0.8241	1.3296	544[a]
HIS	5.5024[a] (2.3017)	0.8789 (1.6586)	-6.0393	9.1177	701[b]	0.1904[a] (0.3189)	-0.8517	1.3982	502[a]
MVP	4.5446[a] (1.7357)	0.7910[a] (1.3230)	-5.3020	5.3208	651[a]	0.2185[a] (0.3279)	-0.8397	1.3194	559[a]
BST	4.9565[a] (2.2125)	0.8528[b] (1.4825)	-5.6498	9.3554	654[a]	0.2084[a] (0.3258)	-0.8183	1.4093	542[a]
Total	5.0639[a] (2.2158)	0.8563[b] (1.4786)	-6.0393	9.3554	2,706[a]	0.2068[a] (0.3192)	-0.8517	1.4093	2,147[a]
MVW Index	3.8424[a] (1.7838)	0.9582 (0.8426)	-0.8434	2.1030	N/A	0.3240 (0.3389)	-0.1640	1.3364	N/A

Table 2

Summary Results from 1500 Simulated Portfolios of 20 Randomly Selected Stocks

The table presents average monthly summary figures for 1500 simulated portfolios of 20 stocks randomly selected from the NYSE. All values shown in Table 2 are presented in exactly the same manner as previously described in Table 1.

Technique	σ_p	R_p	*Min* R_p	*Max* R_p	*No.* $R_p > R_I$	SHP_p	*Min* SHP_p	*Max* SHP_p	*No.* $SHP_p > SHP_I$
EQW	4.2673 (1.7976)	0.8720 (1.0843)	-3.3570	3.8603	696	0.2680 (0.3084)	-0.6378	1.6036	630
HIS	4.2830 (1.5001)	0.8014 (1.3781)	-4.3051	7.0809	640	0.2130 (0.3453)	-0.9376	1.4467	545
MVP	3.6180 (1.3147)	0.7468 (1.1472)	-3.3588	3.6084	589	0.2620 (0.3681)	-0.8458	1.6510	639
BST	3.8690 (1.3701)	0.7636 (1.2531)	-3.8095	5.6076	598	0.2330 (0.3595)	-0.9612	1.8932	559
Total	4.0093 (1.5327)	0.7960 (1.2214)	-4.3051	7.0809	2,523	0.2440 (0.3467)	-0.9612	1.8932	2,373
MVW Index	3.8424 (1.7838)	0.9582 (0.8426)	-0.8434	2.1030	N/A	0.3240 (0.3389)	-0.1640	1.3364	N/A

Table 3

Summary Results from 1500 Simulated Portfolios of 40 Randomly Selected Stocks

The table presents average monthly summary figures for 1500 simulated portfolios of 40 stocks randomly selected from the NYSE. All values shown in Table 3 are presented in exactly the same manner as previously described in Table 1.

Technique	σ_p	R_p	*Min* R_p	*Max* R_p	*No.* $R_p > R_I$	SHP_p	*Min* SHP_p	*Max* SHP_p	*No.* $SHP_p > SHP_I$
EQW	4.0730 (1.7792)	0.8732 (1.0241)	-2.7792	3.2111	673	0.2890 (0.3142)	-0.4909	1.5957	641
HIS	3.9743 (1.3902)	0.7870 (1.3183)	-3.2408	5.5177	624	0.2219 (0.3558)	-0.9371	1.3436	571
MVP	3.3515 (1.2195)	0.7304 (1.1023)	-2.9037	3.4519	586	0.2786 (0.3893)	-0.6921	1.7404	671
BST	3.6471 (1.2851)	0.7617 (1.2297)	-2.8319	5.5177	598	0.2406 (0.3681)	-0.9524	1.5805	608
Total	3.7615 (1.4625)	0.7881 (1.1750)	-3.2408	5.5177	2,481	0.2575 (0.3589)	-0.9524	1.7404	2,491
MVW Index	3.8424 (1.7838)	0.9582 (0.8426)	-0.8434	2.1030	N/A	0.3240 (0.3389)	-0.1640	1.3364	N/A