

On the Predictability of Stock Returns in Real Time*

Michael Cooper, Roberto C. Gutierrez Jr., and William Marcum

November 13, 2001

Forthcoming, *Journal of Business*

* Cooper is from the Krannert Graduate School of Management, Purdue University, West Lafayette, IN 47907-1310, 765-494-4438, mcooper@mgmt.purdue.edu. Gutierrez is from the Mays School of Business, Texas A&M University, 4218 TAMUS, College Station, TX 77843-4218, 979-845-1224, rcg@tamu.edu. Marcum is from the Calloway School of Business, Wake Forest University, Winston-Salem, NC 27109, 336-758-7166, marcumbm@wfu.edu. We appreciate helpful discussions with Brian Balyeat, Jennifer Conrad, Kent Daniel, David Denis, Diane Denis, David Downs, Ken French, Huseyin Gulen, Cam Harvey, Arvind Mahajan, Stewart Mayhew, Richard McEnally, Darius Miller, Christo Pirinsky, Raghu Rau, Henri Servaes, Steve Slezak and seminar participants at Arizona State University, Louisiana State University, Loyola University of Chicago, Purdue University, Texas A&M University, University of Georgia, University of Miami, University of North Carolina, University of Wisconsin, the Tenth Annual Conference on Financial Economics and Accounting at the University of Texas, and the 2001 Western Finance Association Meetings and the comments of an anonymous referee. We thank Mike Cliff for generously providing computational assistance.

On the Predictability of Stock Returns in Real Time

Abstract

Researchers have documented an abundance of evidence that stock returns are predictable *ex post*. We address in this study whether the cross section of stock returns is predictable *ex ante*. We ask if a real-time investor could have used book-to-market equity, firm size, and one-year lagged returns to forecast stock returns during the 1974 to 1997 period. Using a recursive out-of-sample method, we find that the market was difficult to beat in real time. Our findings suggest that the current notion of predictability in the literature is exaggerated.

In the last twenty years, financial economists have documented ample evidence that both the time series and the cross section of stock returns are predictable.¹ Recently, economists have even extended this evidence of predictability to pre-Compustat and foreign stock returns.² All of this evidence, however, is of *ex post* predictability. In other words, the patterns in stock returns were discovered with hindsight (full-period information). What we do not know from this evidence is if stock returns are predictable *ex ante*.

In this paper, we examine if cross-sectional patterns in stock returns were evident in real time without the benefit of hindsight. We develop a recursive out-of-sample method to assess the *ex ante* predictability of stock returns using three premier forecasters: book-to-market equity, size, and momentum.³ We do not contest that an investor *could* have implemented the book-to-market, size, and momentum strategies historically. We address if an investor *would* have chosen to implement them. Would an investor in real-time have found these market-beating strategies amidst the plethora of alternatives, or is the evidence that the market is beatable due only to the clarity of hindsight?

A secondary motivation for our study is to provide a potential resolution to a paradox observed in the literature. The current notion that the stock market is predictable stands in contrast to the well-documented inability of mutual funds to beat the market in real-time (Carhart (1997), Wermers (2000)). Some argue this inability is an agency cost whereby managers choose not to

¹ Basu (1977), Banz (1981), Chen, Roll, and Ross (1986), Keim and Stambaugh (1986), Campbell (1987), Fama and French (1988, 1992, 1993), Poterba and Summers (1988), Lakonishok, Shleifer, and Vishny (1994), Jegadeesh (1990), Jegadeesh and Titman (1993, 1995), Daniel and Titman (1997), and Conrad and Kaul (1998) are just a few.

² This evidence seems to mitigate, but does not eliminate, the concern that these effects are spurious. See Davis (1994), Davis, Fama, and French (1999), Chan, Hamao, and Lakonishok (1991), Harvey (1991), Bekaert and Hodrick (1992), Rouwenhorst (1998), and Fama and French (1998).

³ Fama and French (1992, 1996) show that book-to-market equity and size subsume other predictors of stock returns, but not yearly momentum.

fully exploit apparent mispricings (Lakonishok, Shleifer, and Vishny (1994), Del Guercio (1996), Shleifer and Vishny (1997), Chan, Chen, and Lakonishok (1999)). We suggest another explanation. Perhaps stock returns are not predictable *ex ante*. The performances of these funds then would be a result of managers simply being unable to see the book-to-market, size, and momentum effects coming. It is interesting to note that nearly all studies of real-time investment performances also fail to show that the market is clearly beatable. Barber and Odean (1999) find this for individual investors; Christopherson, Ferson, and Glassman (1998) find this for pension funds; Pirinsky (2001) finds this for banks, investment advisors, and insurance companies; Desai and Jain (1995) find this for “superstar” money managers; Metrick (1999) finds this for newsletter recommendations; Barber et. al. (2000) find this for analysts’ consensus recommendations. How easily the book-to-market, size, and momentum patterns could be detected and exploited *ex ante* is the focus of this study.

We examine if a real-time investor, without any prior beliefs in the efficacy of any specific strategy, would have discovered book-to-market, size, and momentum to be useful predictors of stock returns over the July 1974 to December 1997 period. We follow Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999) who note that allowing for alternative, competing variables is the crucial element of proper *ex ante* out-of-sample testing (model uncertainty). Investors in real time do not know *ex ante* which variables will and which variables will not be useful in capturing future profits. For our analysis, the investor may employ book-to-market equity, size, lagged annual return (momentum), beta, or a combination of these variables. We include beta in the analysis as an additional competing variable to increase the realism of the research design. Obviously, including just four variables understates the set of potential forecasters confronting the investor. Our conservative depiction of the variable set serves to bias our tests in favor of finding

out-of-sample predictability, since three of the four variables are the premier *ex post* cross-sectional predictor variables. Another distinguishing feature of our procedure is that we force the investor to decide *ex ante* how to employ the variables to select stocks (parameter uncertainty). For example, should he invest in a momentum or a contrarian strategy, or neither? We characterize the potential trading rules as cross-sectional sorts of all stocks based on each of the four variables.⁴

We test for predictability by analyzing whether a simulated real-time portfolio outperforms a passive index. The real-time portfolio is constructed each year by selecting the trading rules that perform best during the prior in-sample period. We examine three classes of real-time simulations, each characterized by the criterion employed to determine which rules are the best. The results indicate that if an investor used either mean monthly returns or terminal wealth as the criterion, the real-time portfolios were unable to beat the market. If an investor chose the Sharpe ratio as the criterion, the portfolio generated profits. Despite being handed three of the premier *ex post* variables, a real-time investor would not have easily beaten the market. The results stand in stark contrast to the in-sample results, which identify profits for each of the three simulations. These results are robust to many variations in the out-of-sample method.

More importantly, even the best-performing real-time portfolio generates profits that are only a fraction of what the literature has discovered *ex post*. Specifically, the best portfolio we form beats the market by 19 (10) basis points per month on average (after trading costs). Alternatively, a "hindsight" portfolio of stocks with high book-to-market ratios, low market

⁴ Pastor and Stambaugh (2000) examine the portfolio-allocation decision of a mean-variance optimizing investor in January 1998 who updates his beliefs with return data in a Bayesian fashion. They find that pricing uncertainty and short-sale constraints leave the investment portfolio nearly identical if the investor employs the Fama-French three-factor model of expected returns or the Daniel-Titman characteristic model. Employing the CAPM results in different allocations. Employing no model and relying solely on the prior

capitalization, and high lagged annual returns outperforms the market by 52 basis points per month on average.⁵

Our results complement recent examinations of the out-of-sample predictability of the *time-series* of stock returns. Bossaerts and Hillion (1999), Sullivan, Timmermann, and White (1999), and Goyal and Welch (1999) document substantial in-sample predictability in the time series of stock index returns but find no evidence of out-of-sample forecastability. Lo and MacKinlay (1997) find some evidence of market-timing profits from 1967 to 1993. However, Pesaran and Timmermann (1995) conclude that a real-time investor could have profited only during the 1970's from timing the stock market, not in the 1960's or the 1980's. Overall, these studies and ours highlight a marked difference between *ex ante* and *ex post* predictability.

The next section details the data and methodology; Section II reports the in-sample results of our method; Section III presents the out-of-sample results; Section IV summarizes the robustness checks; Section V concludes with several potential explanations for the differences in in-sample and out-of-sample results.

I. Data and Methodology

A. Independent and Dependent Variables

We use all NYSE, AMEX, and NASDAQ nonfinancial firms listed on the CRSP monthly stock return files and the COMPUSTAT annual industrial files from 1963 through 1997. To mitigate backfilling biases, a firm must be listed on Compustat for two years before it is included

data (i.e. an agnostic investor) results in a portfolio similar to that chosen under the first two models. Our study simulates the annual portfolio decisions of an agnostic investor from 1974 to 1997.

⁵ This estimate of the profits from hindsight is conservative since we employ quintile sorts and a five-dollar price screen. Depending on how you choose to define value, size, and momentum strategies, hindsight profits can easily exceed the market by 100 basis points per month.

in the data set (Fama and French, 1993). We exclude stocks priced below five dollars to alleviate the microstructure concerns associated with these securities (proportionally high transactions costs and illiquidity) (Ball, Kothari, and Shanken, 1995).

In accordance with Fama and French (1992), we form the book-to-market ratio of equity (B/M) by dividing the book value of a firm's equity at fiscal year-end $t-1$ from Compustat by the market value of equity from CRSP at the end of December of year $t-1$.⁶ SIZE is defined as the market value of the firm's equity from CRSP at the end of June of year t .

Our proxy for a stock's momentum is its one-year lagged return, which is consistent with the stock return models of Asness (1995) and Carhart (1997). Additionally, Jegadeesh and Titman (1993) and Conrad and Kaul (1998) find that momentum strategies are profitable when the conditioning set of lagged returns is defined as any horizon from three-months to one-year prior. Furthermore, a one-year horizon is consistent with our definitions of B/M and SIZE, which are also computed as annual variables. The one-year lagged holding-period returns (LAGRET) are calculated from the beginning of July of year $t-1$ to the end of May of year t . June returns in year t are excluded to mitigate the return biases due to the bid-ask bounce (Fama and French, 1996).⁷

Note that the investor in our analysis has some benefit of hindsight because we provide three variables that have been shown to be correlated (ex post) with returns over the sample period (Fama and French, 1992, Jegadeesh and Titman, 1993). In reality, unless an investor had strong prior beliefs, he would not ex ante employ only *these* specific variables to identify profitable trading rules. A real-time investor faces a much larger set of potential forecasting variables. In a

⁶ The book value of equity is defined as total shareholder's equity plus balance-sheet deferred taxes and investment tax credits minus the book value of preferred stock (valued at the redemption, liquidation, or par value, in that order as available). Firms with negative book values are eliminated.

⁷ Roll (1984) shows that the bid-ask spread induces negative autocorrelation in individual stock returns. "Skip-month" annual returns are used to prevent this spurious negative autocorrelation from reducing a momentum effect in LAGRET.

conservative effort to alleviate this structural advantage, we increase the variable set by including beta, which is very likely to have been considered as a forecast variable over the sample period.⁸ By including beta, we are examining if there is cross-sectional predictability due explicitly to differences in risk. It is unlikely that beta's poor performance in explaining stock returns would have been realized until the end of the sample period.

A beta is assigned to individual stocks in June of year t and is estimated using no more than 60 months and no less than 24 months of prior returns, employing the CRSP value-weighted index as the market's proxy. We define BETA as the sum of the coefficients in the regression of stock returns on lagged and contemporaneous market returns (Dimson (1979) and Fama and French (1992)).

While it is obvious that hindsight biases are not completely eliminated from our study, these biases are mitigated. The remaining bias will likely incline the out-of-sample test statistics toward falsely rejecting the null hypothesis of no predictability. Conrad, Cooper, and Kaul (2001) show that using randomly generated variables that have "worked" over the entire sample period biases the recursive out-of-sample performances of these variables towards providing evidence of predictability.⁹

In the next section, we detail the method used to examine the out-of-sample explanatory power of B/M, SIZE, LAGRET, and BETA for the cross section of monthly stock returns from July of year t to June of year $t+1$.

⁸ Other variables which could have been considered during this time period are P/E ratios, leverage, dividend yield, multifactor betas from factor analysis or Chen, Roll, and Ross (1986), six-month lagged returns, three-year lagged returns, to name just a few.

⁹ Specifically, they find that using the five best-performing randomly generated (spurious) variables generates recursive out-of-sample profits of fourteen basis points per month on average.

B. Out-of-Sample Methodology

To assess the forecasting value of B/M, SIZE, LAGRET, and BETA over the 1974 to 1997 period, we develop a recursive out-of-sample procedure to simulate an investor's real-time decision-making process. We will first describe the main specifications of the simulation, which are a culmination of the methods and results of the current predictability literature. Then, after reviewing the results of these simulations, we will investigate alternatives for each modeling choice we make. These robustness checks are discussed in Section IV.

The rule universe is constructed using all possible one-way and two-way independent sorts of the four variables' quintiles. There are a total of 20 quintiles of the four variables and 150 two-way combinations of the quintiles.¹⁰ Therefore, the investor considers 170 rules in each decision period. We exclude individual rules that identify more than one quintile of a particular variable.

The first in-sample (training) period extends from July 1964 to June 1974. We employ a rolling ten-year in-sample window as a reasonable tradeoff between reducing error in the estimation of the relations between stock returns and our choice variables and permitting regime switches in those relations.¹¹ Stocks are sorted into quintiles based on each variable (B/M, SIZE, LAGRET, and BETA) at the end of June of each year t of the in-sample window. The monthly equally-weighted returns for each of the 170 rules are calculated from July of each year t to June of each year $t+1$.

We follow Pesaran and Timmermann (1995) and Sullivan, Timmermann, and White (1999) and run three separate simulations, each one corresponding to a different criterion for

¹⁰ We also examine a much larger universe of 1,295 rules obtained by using all possible one-way, two-way, three-way, and four-way sorts. This specification is discussed in Section IV.

¹¹ Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Sullivan, Timmermann, and White (1999), and Goyal and Welch (1999) document nonstationarity in the time series relations between stock returns and numerous "predictor" variables. Daniel and Titman (1999) choose a ten-year rolling window as well. We also examine an alternative in-sample window that uses all available past data in Section IV.

identifying the best rules over the in-sample period. The three criteria we employ are the mean monthly return, the terminal wealth, and the Sharpe ratio. Consistent with the prior studies, we do not assert which criterion our investor would prefer a priori; we wish the real-time analysis to be reasonably broad in terms of investor types (i.e. utility functions).

The rules generating the highest (lowest) ten percent - 17 rules in this case - of mean monthly returns over the entire ten-year in-sample window form the investor's chosen LONG (SHORT) out-of-sample portfolio for July 1974 to June 1975. Similarly, the LONG (SHORT) portfolios for the terminal-wealth and Sharpe-ratio criteria are formed by selecting the rules that generate the highest (lowest) ten percent of terminal wealths and Sharpe ratios, respectively, over the in-sample period. For illustration, we provide three of the 17 LONG rules under the Sharpe-ratio criterion for the July 1974 to June 1975 out-of-sample period. The first rule indicates to buy all stocks that are in the lowest quintile of SIZE. The second rule indicates to buy all stocks that are both in the lowest quintile of SIZE and in the highest quintile of B/M. The third rule indicates to buy all stocks that are both in the lowest quintile of B/M and in the lowest quintile of BETA. These three rules are each one of the 17 (out of 170) rules which generated the highest Sharpe ratios during the prior ten-year in-sample period. Stocks selected for out-of-sample investment by more than one rule do not receive increased weighting. We choose the top (bottom) ten percent of the rules to ensure that we have reasonably diversified the noise in the relationships between the sort variables and stock returns. On average, the LONG (SHORT) portfolios for each of the three criteria contain over 500 (600) stocks, which is roughly equivalent to investing in a quintile of the available stocks.¹²

¹² For robustness, we also examine the performances of the top (bottom) rule only and the top (bottom) 5 percent of the rules. These specifications are discussed in Section IV.

Monthly equally-weighted returns are calculated for the LONG and SHORT portfolios over this out-of-sample period.¹³ At the end of this first out-of-sample period, June 1975, the in-sample window is rolled forward one year, and the process is repeated. This procedure produces a time-series of monthly out-of-sample LONG and SHORT returns from July 1974 to December 1997. While the investor's strategy is updated annually, the rules do not change dramatically from year to year. One year of new information is added to the previous nine years each time he updates his strategy. So small weight is given to the latest year's returns.

Note that we allow the entire cross section of stocks to be the investor's real-time opportunity set. We do not limit his rule universe to just the extreme quintiles of the predictor variables. A real-time investor would not likely have assumed a monotonic relationship between stock returns and each of the respective variables, B/M, SIZE, and LAGRET. Without any theoretical guidelines, an investor would most likely have relied on the empirical relationships to form his beliefs, as we have him do here.¹⁴ Our investor will invest in the extreme portfolios if prior returns to such strategies are best. Note that employing a regression-based analysis, instead of the sorting procedure, would be inconsistent with our assumption of no prior beliefs since linearity is imposed by the regressions.

To evaluate whether the cross section is predictable in real time, the returns of the active LONG and SHORT portfolios are compared to a passive benchmark. Finding that the LONG (SHORT) portfolio outperforms (underperforms) the benchmark is evidence of predictability. Because financial researchers disagree on how best to evaluate portfolio performance, we employ

¹³ The returns of the SHORT portfolio are constructed from a positive investment in the appropriate stocks; so profitable SHORT portfolio returns will be negative.

¹⁴ Nevertheless, as a robustness check, we examine in Section IV the performances of the strategies selected only from the first and fifth quintiles of each variable, i.e. imposing the prior belief that the relationships are all monotonic.

several measures. First, we compare the LONG (SHORT) portfolio's mean return to that of an equally-weighted market index (EW index), using a paired t-test. The EW index is composed of all stocks in the data set.¹⁵ If the cross section of stock returns is predictable ex ante, the LONG (SHORT) portfolio should generate a higher (lower) mean return. The second and third performance measures are risk-adjusted. We estimate a Jensen's alpha and a Sharpe Ratio respectively for the active portfolios. If the cross section of stock returns is predictable ex ante, the Jensen's alpha of the LONG (SHORT) should be positive (negative), and the Sharpe ratio of the LONG (SHORT) should be greater (less) than that of the passive EW index.

Notice that we consider only CAPM-based measures of risk-adjusted performance. This restriction is determined by the fact that the most common alternatives, namely the three-factor model of Fama and French (1993) or the four-factor model of Carhart (1997), utilize the same variables we are investigating here. Financial research labels these variables as "anomalous" because the CAPM is unable to explain the associated in-sample return patterns. For this reason, we employ the same measures to determine if the out-of-sample performance of B/M, SIZE, and LAGRET remains anomalous.

Our final test entails an examination of "zero-cost" COMBINED portfolios. The returns of the COMBINED portfolios are obtained by subtracting the returns of the SHORT portfolio from the returns of the LONG portfolio. We evaluate the COMBINED portfolio's performance by comparing its mean monthly return to zero and by estimating a Jensen's alpha (the risk-free rate is not subtracted from the COMBINED returns in the Jensen regression). If the COMBINED portfolio enjoys a mean return that is greater than zero or an alpha that is greater than zero, we view this as evidence of predictability.

¹⁵ We also examine a value-weighted index and the S&P 500 as alternative passive indices. These alternatives are discussed in Section IV.

Due to the practical limitations on an investor's ability to use the proceeds from short sales to fund long positions, as well as margin requirements and "haircuts", caution must be used when interpreting the returns to the COMBINED portfolios.¹⁶ Furthermore, not many investors would be allowed by their brokers to short an entire portfolio of hundreds of stocks, which the SHORT portfolio requires. For these reasons, we place the greatest emphasis in this study on the ability of the LONG portfolio to outperform the EW index.

It may be useful to note here that we consider a variety of alternatives to this procedure in Section IV. Changes in the in-sample window length, in the rule universe, and in the number of rules selected for investment do not alter the general performance of the method described in this section.

II. In-Sample Results

A. Returns to LONG, SHORT, and COMBINED Portfolios In-Sample

Figures 1 and 2 display the in-sample performances of the LONG and SHORT portfolios under the mean-return criterion. The LONG and SHORT portfolios both sustain remarkable in-sample performances. Figure 1 shows the spreads between the mean monthly returns of the LONG portfolio and the mean monthly returns of the EW index for each of the rolling ten-year in-sample periods. The time series of the in-sample spreads is quite smooth with the LONG portfolio outperforming the EW index by an average of 0.55% per month. The LONG alpha is 0.56% per month on average and indicates that the LONG easily outperforms the index on a risk-adjusted basis as well.

¹⁶ See Alexander (2000) for a detailed discussion and analysis of the regulatory constraints that preclude "zero-cost" strategies.

Figure 2 plots the in-sample performance of the mean-return-criterion SHORT portfolio relative to the EW index. The performance of the SHORT portfolio mirrors the performance of the LONG portfolio. The average in-sample underperformance of the SHORT portfolio relative to the EW index is 0.51% in raw returns with an average alpha of -0.38% . Interestingly, the small time-series variations in the market-adjusted returns of the LONG and of the SHORT are nearly identical.

Using the ten percent of rules that generate the highest terminal wealths, the LONG portfolio outperforms the EW index by 54 basis points per month on average. Using the rules generating the worst terminal wealths, the SHORT underperforms the index by 49 basis points per month on average. Furthermore, these portfolios generate average alphas of 0.58% and -0.41% respectively. Although not shown, the times-series plots of average in-sample LONG and SHORT market-adjusted returns using the terminal-wealth criterion are nearly identical to Figures 1 and 2.

Employing the Sharpe ratio to select the investment rules results in a slightly less impressive in-sample performance. The average monthly return for the LONG in this case exceeds the average return of the index by 30 basis points, while the return of the SHORT falls below the return on the index by 45 basis points per month on average. Their average alphas are 0.59% and -0.41% respectively. Again, the time series variations in these in-sample market-adjusted returns is small (figures not shown), but the pattern of the changes is very similar to that of the previous two simulations, only the magnitude of the spreads is noticeably different.

We draw attention to the level of in-sample returns exhibited by the real-time simulations because they are consistent with the in-sample performances documented for B/M, SIZE, and LAGRET by previous research (Fama and French (1992), Lakonishok, Shleifer, and Vishny (1994), and Jegadeesh and Titman (1993)). Second, the persistence of the in-sample returns is

quite dramatic. In fact, the *minimum* market-adjusted in-sample return for the LONG portfolio is 35, 33, and 9 basis points per month for the mean-return, terminal-wealth, and Sharpe-ratio criteria, respectively. Furthermore, these minima all occur in 1992. The only difference that is apparent so far between the three specifications is the weaker performance of the LONG for the Sharpe-ratio criterion. The examination of the rules that are selected for the LONG and SHORT portfolios in the next section will reveal why this difference occurs.

B. Rule Compositions of LONG and SHORT Portfolios

For each of the three criteria, two-way rules dominate the investment choices with at least 90% of the selected LONG and SHORT rules being two-way rules. In addition, the LONG and the SHORT portfolios consist of between 558 and 666 stocks on average under each of the three criteria. The EW index has 2,646 stocks on average in each year. Most important though is that *each* of the four variables is employed frequently in the construction of the LONG and SHORT portfolios. Of the total number of investment rules selected (17 rules per year across 24 selection periods) for the portfolios, each of the variables is typically employed in about 40% to 55% of these rules when using the three in-sample selection criteria.

It is interesting to note how the hypothetical investor is using the four variables. Table I provides the mean quintiles of each variable selected for investment in the LONG and SHORT portfolios. For the mean-return and terminal-wealth criteria, the average quintiles selected for B/M and SIZE conform to our hindsight. The LONG (SHORT) portfolio tends to invest in high-B/M (low-B/M) and small-cap (high-cap) stocks. There however seems to be no discernable tendency in the LONG or SHORT portfolios under the mean-return and terminal-wealth criteria toward a momentum or a BETA strategy. For both LAGRET and BETA, the entire spectrum of

quintiles is employed as evidenced by the relatively high standard deviations of selected quintiles reported in Table I.

The quintiles selected for investment under the Shape-ratio criterion are noticeably different. Table I shows that while the LONG and the SHORT tend to be composed of high-B/M and low-B/M stocks respectively, the tendency for the LONG (SHORT) to be in low-cap (high-cap) stocks diminishes. Instead, the selected quintiles under the Sharpe-ratio criterion demonstrate a pronounced tendency to invest in a particular BETA style. Particularly interesting is that the LONG portfolio tends to be comprised of low-BETA stocks and the SHORT of high-beta stocks. A momentum effect is not employed under this criterion either.

We see now why the in-sample returns of the mean-criterion and terminal-wealth criterion are so similar; they are in general selecting the same characteristics for B/M, SIZE, LAGRET, and BETA. The in-sample returns of the Sharpe-ratio criterion are noticeably different than the other two criteria because it selects markedly different BETA characteristics and, to a lesser degree, different SIZE characteristics.

Despite the noted deviations from the ex post best rules, the real-time simulations are still able to detect a high level of in-sample "predictability" consistent with the prior literature. In the next section, we document the out-of-sample performances of these real-time portfolios to see if the large in-sample returns persist into the next period. In other words, we examine the viability of B/M, SIZE, LAGRET, and BETA as predictors of stock returns in real time.

III. Out-of-Sample Results

A. Mean-Return Criterion

Table II presents statistics for the 23.5 out-of-sample years extending from July 1974 to December 1997. Without adjusting for trading costs, the LONG portfolio earns a mean monthly return of 1.54%, exceeding the EW index by 13 basis points, which is a statistically significant difference at the ten-percent level.¹⁷ However, the Jensen's alpha and the Sharpe ratio of the LONG portfolio fail to detect risk-adjusted abnormal returns. Additionally, the SHORT and COMBINED portfolios do not perform statistically differently than their respective benchmarks on a raw or risk-adjusted return basis.

We adjust the portfolio returns for trading costs using the findings of Keim and Madhavan (1997). These cost estimates are stock-specific and are a function of the price of the stock, the exchange where the stock trades, and the market value of the stock. The Keim-Madhavan estimates are for 1991 to 1993. We adjust these cost estimates for time series variations using the results of Stoll (1995). The method for estimating trading costs is detailed in Appendix A. Once the returns of these mean-return-criterion portfolios are adjusted for trading costs, there is not a

¹⁷ The t -statistics for the means tests and the Jensen's alpha tests are calculated using the HAC covariance estimator of Gallant (1987). In Section B of the Appendix, we describe the procedure we employ for selecting the bandwidths, suggested by Andrews (1991). Note that the inferences throughout this study are robust to alternative choices of the bandwidths. These alternatives are also suggested by Andrews (1991) and are described in Appendix B as well. The t -statistics that compare the Sharpe ratio of the LONG or SHORT portfolio to the Sharpe ratio of the EW market portfolio are estimated via the "delta method" (Greene (1997), Theorem 4.16, p. 124). For any two portfolios, we estimate the mean and variance of the excess returns as well as the covariance matrix of these four parameters using GMM with the robust HAC covariance estimator. The asymptotic distribution of the difference between the Sharpe ratios of the two series, which is a function of the four parameters, is given in Theorem 4.16 of Greene (1997).

single piece of evidence in favor of predictability.¹⁸ An investor employing the mean-return criterion to select his investment strategies would not have outperformed the passive index.

Figure 3 plots the out-of-sample versus the in-sample performances of the LONG portfolio under the mean-return criterion. The out-of-sample returns are quite volatile. The LONG performs well from 1974 to 1978 but underperforms the EW index from 1983 to 1991 in all but one year. Similarly, Figure 4 illustrates the poor and volatile performance of the SHORT portfolio, which never enjoys a sustained period of market underperformance. Nonetheless, Figure 5 shows that the COMBINED portfolio generated reasonable returns over the 1974 to 1982 period, based primarily on the strength of the LONG portfolio. During this period, we observe a negative mean monthly return to the COMBINED portfolio only in 1979. After 1983, extreme variability and negative returns mark the out-of-sample performance of the COMBINED portfolio.

B. Terminal-Wealth Criterion

Table III shows that the terminal-wealth criterion performs better than the mean-return criterion. While the LONG portfolio unadjusted for trading costs beats the EW index by 13 basis points again and at a ten-percent level of significance, the LONG in this case displays some ability to outperform the index on a risk-adjusted basis. The LONG alpha is 14 basis points per month, which is significant at the ten-percent level, and the Sharpe ratio of the LONG portfolio is 0.18, which is significantly different from the EW index's Sharpe ratio of 0.15. The other difference between this case and the mean-return case is that the COMBINED here has an alpha of 28 basis

¹⁸ On average, the estimated roundtrip cost of the LONG (SHORT) under the mean-return criterion is 265 (180) basis points.

points that is significant at the five-percent level. Even so, all evidence of predictability dissipates when trading costs are considered.¹⁹

Figure 6 shows that the LONG starts off well but performs erratically after 1979. The SHORT, in Figure 7, never enjoys any sustained underperformance. Figure 8 illustrates that the spread between the LONG and the SHORT is positive for the latter half of the 1970's, but it becomes erratic after then.

C. Sharpe-Ratio Criterion

The performance of the Sharpe-ratio criterion better supports the view that stock returns are predictable. As shown in Table IV, the monthly mean return of the LONG without trading costs exceeds the EW index by 19 basis points and generates an alpha of 28 basis points per month on average. The Sharpe-ratio of the LONG is 0.21. All of these measures indicate that this specification generates performance superior to the EW index. The mean monthly return of the SHORT without trading costs does not differ from the mean return of the index statistically, but it does underperform the EW index on a risk-adjusted basis according to both the Jensen's alpha and the Sharpe ratio. The Sharpe-ratio criterion also generates a COMBINED portfolio that averages 29 basis points per month with an alpha of 47 basis points, both statistically significant.

The most notable point of the Sharpe-ratio case, however, is that the evidence of predictability is not completely eroded by trading costs.²⁰ In fact, after trading costs are considered, the LONG still statistically outperforms the mean on a risk-adjusted basis with an average monthly alpha of 19 basis points and a significantly higher Sharpe ratio (0.19) than that of

¹⁹ On average, the estimated roundtrip cost of the LONG (SHORT) under the terminal-wealth criterion is 253 (196) basis points.

²⁰ On average, the estimated roundtrip cost of the LONG (SHORT) under the Sharpe-ratio criterion is 241 (206) basis points.

the EW index. The COMBINED has a significant alpha of 28 basis points. The raw returns are no longer distinguishable from the benchmarks however. This simulation therefore provides evidence that a Sharpe-ratio investor could have beaten the EW index in real time.

Figure 9 shows that the LONG portfolio constructed using the Sharpe-ratio criterion performs more consistently out-of-sample than the previous two criteria. There is a string of positive market-adjusted returns from 1981 through 1990. However, there is a tremendous below-market return of 89 basis points per month on average in 1991. Figure 10 indicates that the SHORT performs well from 1981 to 1986. The COMBINED, shown in Figure 11, performs remarkably well from 1981 through 1990, but it takes extraordinarily large losses in 1980 and 1991.

D. Subperiod Results

Although the performances of the first two criteria are poor overall, we do see time series variation in the performances of their active portfolios, illustrated in Figures 4 through 9. Specifically, this cursory evidence suggests that the mean-return and terminal-wealth simulations perform best in the 1970's. Interestingly, Pesaran and Timmermann (1995) find out-of-sample *time-series* predictability of stock index returns in the 1970's using macroeconomic variables, but find little evidence of out-of-sample predictability in the 1960's or 1980's. However, we find that for the mean-return and terminal-wealth criteria, there is no evidence of predictability in the 1970's, 1980's, or 1990's after adjusting for trading costs.

As in the overall period, only the Sharpe-ratio criterion provides trading-cost-adjusted evidence of predictability in the subperiods. An examination of Figures 9 through 11 suggests that the Sharpe-ratio case performs well in the 1980's. In fact, even after trading costs are considered,

the LONG, SHORT, and COMBINED provide evidence of predictability in the 1980's. In the 1970's and 1990's, however, no evidence is found after adjusting for trading costs. So the profits uncovered in the real-time simulation under the Sharpe-ratio criterion are not persistent; they are concentrated in the 1980's.

E. Overview of the Out-of-Sample Performances

The most important observation to be made about the out-of-sample performances of these simulations is the dramatic difference between the in-sample and out-of-sample returns. The monthly average return spread between the LONG and the EW index declines from in-sample to out-of-sample by 37% (11 basis points) for the Sharpe-ratio criterion and by 76% for each of the other two criteria (42 basis points for the mean-return criterion and 41 basis points for the terminal-wealth criterion). The erosion in alphas is just as startling. The LONG alpha declines out-of-sample by 53% (31 basis points) for the Sharpe-ratio criterion, by 70% (39 basis points) for the mean-return criterion, and by 76% (44 basis points) for the terminal-wealth criterion. These results suggest that the predictability of stock returns has been vastly overstated in the current literature.

Additionally, the ability of an investor to outperform the index in real time is dubious when using B/M, SIZE, and LAGRET. Of the three criteria applied for selecting the best in-sample rules, only one resulted in abnormal profits after trading costs are considered. Recalling that we give the investor three (ex post) premier cross-sectional variables to use makes these results even more striking. We conclude that beating the market in real time is difficult to do.

Perhaps even more interesting is the finding that the best simulation is only able to generate profits that are a fraction of the revenues commonly believed to be attainable with B/M,

SIZE, and LAGRET. In other words, all of our simulations indicate that the evidence on the predictability of the cross section of stock returns is exaggerated. Figure 12 makes this point emphatically. We plot the terminal wealth generated by a "hindsight" portfolio against the terminal wealths generated by the mean-return criterion and the Sharpe-ratio criterion. The hindsight portfolio is comprised of all stocks that are simultaneously in the highest quintile of B/M, the lowest quintile of SIZE, and the highest quintile of LAGRET. As the evidence in Section II.B indicates, this portfolio is discovered only through the hindsight of financial researchers.

Figure 12 shows that the hindsight portfolio easily dominates the real-time simulations and the EW index. As the prior literature does, we ignore the trading costs of the hindsight portfolio and the EW index. One dollar invested in the hindsight portfolio on July 1, 1974 generates a staggering \$111 by the end of December 1997. One dollar invested in the EW index generates \$35 over the same period. One dollar invested in the mean-return LONG portfolio, incurring trading costs along the way, generates only \$32; one dollar invested in the Sharpe-ratio LONG portfolio, also incurring trading costs along the way, generates only \$49. It is easy to see that the prior evidence of predictability is exaggerated.

IV. Robustness of the Results

To determine if the real-time out-of-sample results are sensitive to the design of the method, we examine several variations in the procedure. Table V identifies the alternative specifications that we consider. Each of the stated changes is employed separately to the base case described in Section I. These alternatives are examined under each of the three selection criteria, mean-return, terminal-wealth, and Sharpe-ratio. We essentially find that our results are robust across a wide variety of alterations in the test design. Specifically, the mean-return and the

terminal-wealth criteria perform poorly after including trading costs while the Sharpe-ratio criterion provides evidence of predictability even after trading costs are considered. The best-performing specifications still generate dramatically lower profits than the hindsight portfolio of Figure 12. One particular result that should be mentioned is that adjusting the out-of-sample profits for a 1.00% one-way trading cost eliminates all evidence of the LONG portfolio outperforming the market under *any* criterion.

V. Conclusion

Researchers have documented an abundance of evidence that stock returns are predictable *ex post*. In this study, we address whether stock returns are predictable *ex ante*. We ask if a real-time investor could have used book-to-market equity, firm size and one-year lagged returns to forecast returns over the 1974 to 1997 period. Using a recursive out-of-sample method, we find that the market was difficult to beat despite being given three of four variables which are the *ex post* premier cross-sectional variables. Moreover, the specification with the highest level of real-time profits falls far short of the in-sample evidence of predictability. The results are robust to variations in the procedure.

The poor out-of-sample performance of book-to-market, size, and momentum suggests that the notion of predictability that is currently in the literature is exaggerated. This has several implications for financial economics. First, the strong debate over whether predictability is due to mispricing or to risk seems potentially misplaced. At a minimum, the findings soften the debate given that the level of predictability is markedly lower than previously believed. Second, the well-documented performances of real-time investors seem consistent with our findings. This suggests that agency costs in the money management industry might not be as high as some economists

have recently suggested (Lakonishok, Shleifer, and Vishny (1994), Del Guercio (1996), Shleifer and Vishny (1997), Chan, Chen, and Lakonishok (1999)). Lastly, the cross-sectional out-of-sample results of our study and the times-series out-of-sample results of others (Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Sullivan, Timmermann, and White (1999), and Goyal and Welch (1999)) suggest that financial economists should perhaps be (re)focusing on understanding why the level of predictability in stock returns is so low.

On this last issue, we mention several (non-mutually exclusive) ideas for why the out-of-sample evidence of stock return predictability so strikingly different from the in-sample evidence but give no particular credence to any. First, the poor out-of-sample performance we document is consistent with the notion that the book-to-market, size, and momentum effects are spurious. In truth, this particular issue might never be settled, except perhaps after observing the next fifty years of stock returns. This is because a second possibility is that one or more of the variables we examine here may be in fact truly correlated with expected returns. Their explanatory power for the cross section of stock returns, however, may simply be quite limited. To identify these variables and their relationships to stock returns “cleanly” in real time may require many future decades of return data.

These are by no means the only potential interpretations of our results. One alternative explanation is offered by Sullivan, Timmermann, and White (1999). They suggest that opportunities may disappear out-of-sample because of increased efficiency in the markets over time, such as lower transactions costs and increased liquidity. This does not seem to be the predominant case, however, since we find little evidence of predictability in the 1970's, but we find the strongest evidence in the 1980's (under the Sharpe-ratio criterion). Alternatively, Goyal and Welch (1999) point out that the lack of out-of-sample predictability may be a consequence of

learning in the marketplace. That is, the best in-sample investment strategies may not persist into subsequent periods because the market adjusts to the new information.

Finally, on quite a different note, Lewellen and Shanken (2001) and Bossaerts and Hillion (2001) argue that the Bayesian learning of economic agents can generate ex post predictable patterns that are ex ante rational and therefore not real-time tradable opportunities. In this case, predictability is just an ex post illusion. For example, suppose you *know* that the time-series of stock returns is mean-reverting. In real time, you still do not know if stock prices will be higher or lower next period because you do not know the true mean of the distribution. Nonetheless, a pattern of mean reversion is easily detected ex post relative to the *sample* mean. Disentangling the potential explanations for poor out-of-sample predictability may provide fruitful avenues for future research.

REFERENCES

- Alexander, G. J., 2000, "On Back-Testing 'Zero-Investment' Strategies", *Journal of Business* 73 (2), 255-278.
- Andrews, D. K., 1991, "Heteroskedastic and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.
- Asness, C., 1995, "The Power of Past Stock Returns to Explain Future Stock Returns," working paper, Goldman Sachs.
- Ball, R., S. P. Kothari, and J. Shanken, 1995, "Problems in Measuring Portfolio Performance: An Application to Contrarian Investment Strategies," *Journal of Financial Economics* 38, 79-107.
- Banz, R. W., 1981, "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics* 9, 3-18.
- Barber, B., R. Lehavy, M. McNichols, and B. Trueman, 2000, "Can Investors Profit from the Prophets?: Security Analysts' Recommendations and Stock Returns," forthcoming *Journal of Finance*.
- Barber, B. M. and T. Odean, 2000, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance* 55, 773-806.
- Basu, S., 1977, "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance*, 32, 663-682.
- Bekaert, G. and R. J. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," *Journal of Finance* 47, 467-509.
- Bossaerts, P., and P. Hillion, 2001, "IPO Post-Issue Markets: Questionable Predilections but Diligent Learners," *Review of Economics and Statistics* 83, 1-15.
- Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria To Select Return Forecasting Models: What do we learn?" *Review of Financial Studies* 12, 405-428.
- Campbell, J., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373-399.
- Carhart, M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.
- Chan, L. K. C., Y. Hamao, and J. Lakonishok, 1991, "Fundamentals and Stock Returns in Japan," *Journal of Finance* 46, 1739-1764.

- Chan, K. C., H. Chen, and J. Lakonishok, 1999, "On Mutual Fund Investment Styles," NBER working paper 7215.
- Chen, N., R. Roll, and R. Ross, 1986, Economic Forces and the Stock Market, *Journal of Business*, 59, 383-403.
- Christopherson, J. A., W. Ferson, and D. A. Glassman, 1998, "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance," *Review of Financial Studies* 11, 111-142.
- Conrad, J., Cooper, M. and G. Kaul, 2001, "Value Versus Glamour," unpublished manuscript, University of Michigan.
- Conrad, J., and G. Kaul, 1998, "An Anatomy of Trading Strategies," *Review of Financial Studies*, 11, 489-519.
- Daniel, K. and S. Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* 52, 1-33.
- Daniel, K. and S. Titman, 1999, "Market Efficiency in an Irrational World," *Financial Analysts Journal* 55 (6), 28-40.
- Davis, J. L., 1994, "The Cross Section of Realized Stock Returns: The Pre-Compustat Evidence," *Journal of Finance* 49, 1579-1593.
- Davis, J. L., E. F. Fama, and K. R. French, 1999, "Characteristics, Covariances, and Average Returns: 1929 to 1997," forthcoming *Journal of Finance*.
- DeBondt, W. and R. Thaler, 1985, "Does the Stock Market Overreact?," *Journal of Finance* 40, 793-805.
- Del Guercio, D., 1996, "The Distorting Effect of the Prudent-Man Laws on Institutional Equity Investors," *Journal of Financial Economics* 40, 31-62.
- Desai, H. and P. Jain, 1995, "An Analysis of the Recommendations of the 'Superstar' Money Managers at Barron's Annual Roundtable," *Journal of Finance* 50, 1257-1274.
- Dimson, E., 1979, "Risk Measurement when Shares are Subject to Infrequent Trading," *Journal of Financial Economics* 7, 197-226.
- Fama, E. F. and K.R. French, 1988, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy* 96, 246-273.
- Fama, E. F., and K. R. French, 1992, "The Cross Section of Expected Stock Returns," *Journal of Finance* 47, 427-465.

- Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.
- Fama, E. F., and K. R. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance* 51, 55-84.
- Fama, E. F., and K. R. French, 1998, "Value versus Growth: The International," *Journal of Finance* 53, 1975-1999.
- Gallant, A. R., 1987, *Nonlinear Statistical Models*, New York: Wiley.
- Goyal, A. and I. Welch, 1999, "The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium?" unpublished manuscript.
- Greene, W. H., 1997, *Econometric Analysis*, New Jersey: Prentice Hall.
- Harvey, C., 1991, "The World Price of Covariance Risk," *Journal of Finance* 46, 111-157.
- Jegadeesh, N., 1990, "Evidence of Predictable Behavior of Security Returns," *Journal of Finance* 45, 881-898.
- Jegadeesh, N. and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 65-91.
- Jegadeesh, N. and S. Titman, 1995, "Overreaction, Delayed Reaction, and Contrarian Profits," *The Review of Financial Studies* 8, 973-993.
- Keim, D., and A. Madhavan, 1997, "Execution Costs and Investment Performance: An Empirical Analysis of Institutional Equity Trades," *Journal of Financial Economics*, 46, 265-292.
- Keim, Donald B. and Robert F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17, 357-390.
- Lakonishok, J., A. Shleifer, and R. W. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance* 49, 1541-1578.
- Lewellen, J. and J. Shanken, 2001, "Learning, Asset-Pricing Tests, and Market Efficiency," forthcoming *Journal of Finance*.
- Lo, A. W., and A. C. MacKinlay, 1997, "Maximizing Predictability in the Stock and Bond Markets," *Macroeconomic Dynamics* 1, 102-134.
- Metrick, A., 1999, "Performance Evaluation with Transactions Data: The Stock Selection of Investment Newsletters," *Journal of Finance* 54, 1743-1775.
- Moskowitz, T. J., 2000, "Discussion", *Journal of Finance* 55, 1695-1703.

- Odean, P., 1999, "Do Investors Trade Too Much?", *American Economic Review* 89, 1279-1298.
- Pastor, L. and R. Stambaugh, 2000, "Comparing Asset Pricing Models: an Investment Perspective," *Journal of Financial Economics* 56, 335-381.
- Pesaran, M. and A. Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance* 50, 1201-1228.
- Pirinsky, C., 2001, "Are Financial Institutions Better Investors?," unpublished manuscript, Texas A&M University.
- Poterba, J. and L. Summers, 1988, "Mean Reversion in Stock Prices: Evidence and Implications," *Journal of Financial Economics* 22, 27-59.
- Roll, R., 1984, "A Simple Implicit Measure of the Bid-Ask Spread in an Efficient Market," *Journal of Finance* 39, 1127-1139.
- Rouwenhorst, K. G., 1998, "International Momentum Strategies," *Journal of Finance* 53, 267-284.
- Shleifer, A. and R. W. Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance* 52, 32-55.
- Stoll, H. R., 1995, "The Importance of Equity Trading Costs: Evidence From Securities Firms' Revenues," in *Global Equity Markets: Technological, Competitive, and Regulatory Challenges*, ed. by R. A. Schwartz (Irwin Professional Publishing, Chicago).
- Sullivan, R., A. Timmermann, and H. White, 1999, "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance* 54, 1647-1691.
- Wermers, R., 2000, "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *Journal of Finance* 55, 1655-1695.

Appendix

A. Trading Cost Estimates

Since we are interested in quantifying the extent to which stock returns are predictable in real time, we must account for the trading costs that investors incur. This is difficult to estimate since each investor faces different costs for each transaction. In a recent study, Keim and Madhavan (1997) estimate the trading costs for 21 institutions from January 1991 through March 1993. Using 62,333 trades, they find that the average roundtrip total cost of equity trading is 146 basis points. This amount includes commissions paid as well as an estimate of the price impact of the trade. More importantly, the authors show the tremendous variation that exists in trading costs across institutions, investment styles, trade difficulty, and exchanges.²¹

Keim and Madhavan regress total trading costs on several characteristics of the trade and the traded stock. As Wermers (2000) does in a recent study, we employ the regression results of Keim and Madhavan to estimate trading costs for each stock transaction in our simulation. Since we wish to be *conservative* in our cost estimates, we assume that our hypothetical investor is a trader of the type that Keim and Madhavan classify as "value" (long-term traders). They find that this group incurs lower trading costs than the "technical" and "index" traders, possibly due to the "value" traders' lower demand for immediacy in the execution of their trade orders. We also are conservative in that we set the trade size equal to zero. Not surprisingly, they find that, as the size of the trade increases, the trading costs increase.

²¹ For example, they find that "value" (long-term) traders average 47 basis points roundtrip while "index" traders and "technical" traders average 108 and 205 basis points respectively. The average trading costs for the stocks in the smallest (largest) quintile of market capitalization are 383 (56) basis points for NYSE/AMEX and 578 (40) basis points for NASDAQ.

Using the results in Table 5 of Keim and Madhavan, and setting their technical-trader and index-trader dummies to zero as well as trade size to zero, we obtain our estimates of buyer and seller trading costs

$$\hat{C}^{Buy}_i = 0.767 + 0.336D^{NASDAQ} - 0.084Logmcap + 13.807\left(\frac{1}{P_i}\right) \quad (A.1)$$

$$\hat{C}^{Sell}_i = 0.505 + 0.058D^{NASDAQ} - 0.059Logmcap + 6.537\left(\frac{1}{P_i}\right) \quad (A.2)$$

where \hat{C}^{Buy}_i and \hat{C}^{Sell}_i are the estimated total trading costs for stock i in percent for either a buyer-initiated or seller-initiated order, respectively, D^{NASDAQ} is equal to one if stock i is a NASDAQ-traded stock and zero if stock i is traded on NYSE or AMEX, $Logmcap$ is the logarithm of the market value of outstanding stock i measured in thousands of dollars, and P_i is the price per share of stock i . All of these variables are obtained from CRSP. For the LONG portfolio, we use \hat{C}^{Buy}_i to open the positions in the component stocks and \hat{C}^{Sell}_i to close the positions, vice versa for the SHORT portfolio.

Equations (A.1) and (A.2) however do not reflect the substantial decline in trading costs over the 1974 through 1997 period. Using the revenues from broker and dealer firms to estimate trading costs, Stoll (1995) finds that costs have declined over 40% from 1980 to 1992. As Wermers (2000) does, we use the year-by-year results of Stoll to adjust equations (1) and (2) for changes in costs over time.²² The adjusted costs are

$$A\hat{C}^{Buy}_i = Y_t^e \hat{C}^{Buy}_i \quad (A.3)$$

²² Wermers does not use our equations (1) and (2) exactly. First, in his cost estimates, he adjusts the intercept to account for the higher costs incurred by index and technical traders, and second, he considers trade size, which he observes in his mutual fund data.

$$A\hat{C}^{Sell}_i = Y_t^e \hat{C}^{Sell}_i \quad (A.4)$$

where $A\hat{C}^{Buy}_i$ and $A\hat{C}^{Sell}_i$ are the adjusted estimates of the respective buyer and seller trading costs for stock i , Y_t^e is the yearly scale factor for year t and exchange e (either NYSE/AMEX or NASDAQ), and \hat{C}^{Buy}_i and \hat{C}^{Sell}_i are from equations (A.1) and (A.2) respectively. The yearly scale factor, Y_t^e , is calculated by dividing the year t cost estimate of Stoll by the 1992 estimate (Stoll, 1995, Table 7-4). Stoll separates the yearly costs into exchanges and NASDAQ. The NYSE/AMEX stocks that our simulations select are adjusted using the exchange scale factors, and the NASDAQ stocks using the NASDAQ scale factors. These yearly scale factors are provided in Table A.

Table A
Yearly Scale Factors for Trading Cost Estimates

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Exchanges	2.426	2.225	2.176	1.827	1.673	1.525	1.335	1.106	1.211	1.250	0.986	1.194	1.000
NASDAQ	2.196	2.076	1.856	1.802	1.850	1.685	1.428	1.232	1.386	1.188	1.097	1.029	1.000

Stoll only provides cost estimates for 1980 through 1992. Therefore, we adjust trading costs for each year 1974 through 1979 by employing the appropriate exchange's 1980 scale factor and adjust each year from 1993 through 1997 by employing the appropriate 1992 scale factor.

This likely results in our underestimating the costs in the early years and overestimating the costs in the later years.²³

It is important to note that the estimates we employ for equity trading costs are those of institutional investors. Recent studies have addressed the trading costs of individual investors. Odean (1999) finds that the average cost for 10,000 individuals at a discount brokerage firm from 1987 through 1993 is 5.93% roundtrip (of which 4.99% is commissions). Barber and Odean (2000) find that the average trading cost of 66,465 individuals at a discount brokerage house from 1991 through 1996 is 4.03% roundtrip (of which 3.03% is from commissions). It is sufficed to say that these cost levels preclude all of our real-time simulations from outperforming the market.

B. Automatic Bandwidth Selection for HAC Estimation of the Covariance Matrix

We calculate standard errors using the covariance estimator suggested by Gallant (1987), which is robust to heteroskedasticity and autocorrelation. To select the bandwidth, we follow the suggestions of Andrews (1991). Let $y_t = \mathbf{X}_t' \boldsymbol{\theta} + \varepsilon_t$, where y_t and ε_t are scalars, \mathbf{X}_t and $\boldsymbol{\theta}$ are $(k \times 1)$ vectors. We parameterize $(\hat{\varepsilon}_t, \mathbf{X}_t)$ as an AR(1) model with no drift term. Let $(\hat{\phi}_a, \hat{\sigma}_a^2)$ be the Least Squares estimates of the autoregressive and innovation variance parameters for the AR(1) model of series a , where a denotes ε_t and each elemental series of \mathbf{X}_t .

We plug these parameters into the following equation to obtain an estimate of α for the kernel suggested by Gallant (1987) (equation 6.4 of Andrews, 1991).

²³ We contacted the SEC to acquire the data necessary to extend the estimates of Stoll (1995) through 1997; but, the data are unavailable.

$$\hat{\alpha} = \frac{\sum_a w_a \frac{4\hat{\phi}_a^2 \hat{\sigma}_a^2}{(1-\hat{\phi}_a)^8}}{\sum_a w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\phi}_a)^8}} \quad (\text{A.5})$$

The value of $\hat{\alpha}$ from equation (A.5) is then used to calculate the optimal bandwidth, S_T , for the kernel suggested by Gallant (1987) (equation 6.2 of Andrews, 1991).

$$\hat{S}_T = 2.6614(\hat{\alpha}T)^{1/5} \quad (\text{A.6})$$

Andrews (1991) shows that S_T is optimal under a mean-squared-error loss function. However, he suggests estimating the covariance matrix with several alternative bandwidths centered about the optimal bandwidth. To examine the robustness of our results with respect to the bandwidth selection, we follow his recommendation and calculate these alternative bandwidths by replacing $\hat{\phi}_a$ with its estimated value plus or minus one and two standard deviations of its value ($1/\sqrt{T}$).

Note that this procedure uses real values for the bandwidths; Gallant (1987) uses integers. Also, in this notation, S_T equals one plus the bandwidth parameter defined by Gallant (1987).

**Market-Adjusted In-Sample Returns
of LONG Portfolio
under the Mean-Return Criterion**

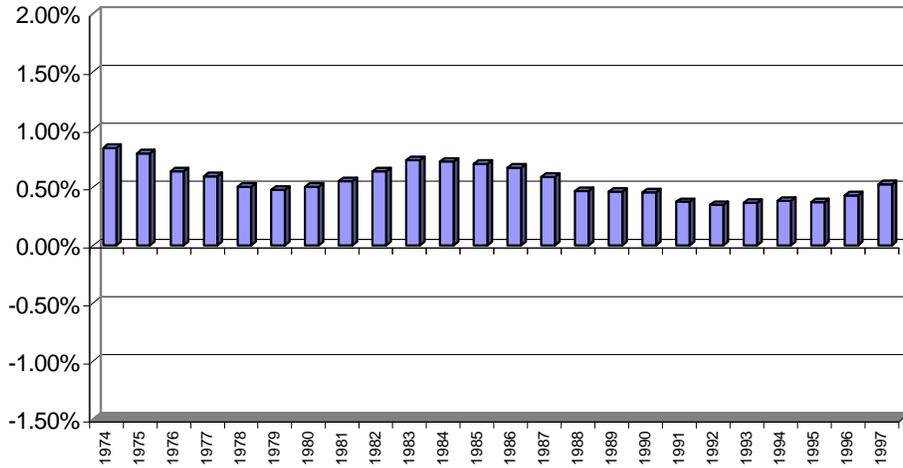


Figure 1: The differences between the mean in-sample monthly returns of the LONG portfolios under the mean-return criterion and the contemporaneous mean monthly return on the EW index are plotted for each rolling ten-year in-sample period. The date indicates the last year of the corresponding in-sample period.

**Market-Adjusted In-Sample Returns
of SHORT Portfolio
under the Mean-Return Criterion**

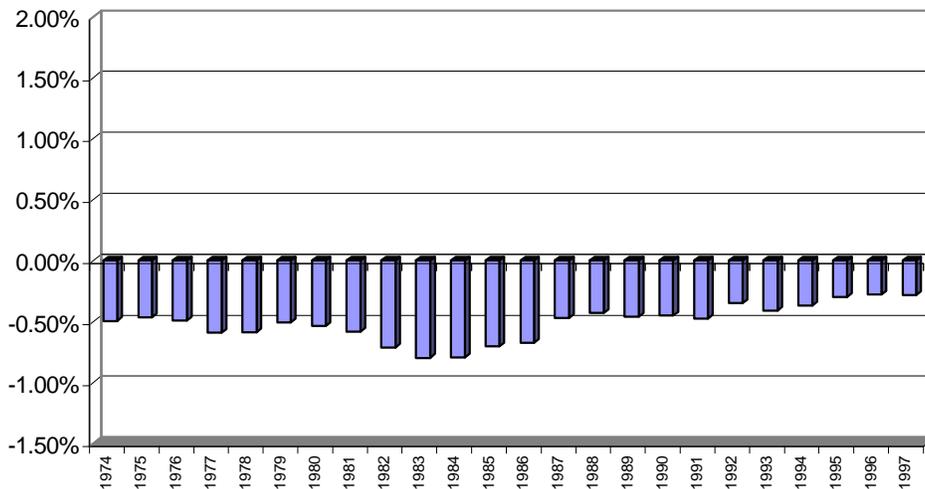


Figure 2: The differences between the mean in-sample monthly returns of the SHORT portfolios under the mean-return criterion and the contemporaneous mean monthly return on the EW index are plotted for each rolling ten-year in-sample period. The date indicates the last year of the corresponding in-sample period.

Table I**Mean Quintiles Selected for Investment**

Using the four variables, B/M, SIZE, LAGRET, and BETA, investment portfolios are identified in June of each year from 1974 to 1997. Consider the first in-sample period, which extends from July of 1964 to June of 1974. At the beginning of July of each year 1964 to 1973, all NYSE-, AMEX-, and -NASDAQ-listed stocks are sorted into quintiles based on the four variables separately. Equally-weighted returns to each of the resulting 170 one-way and two-way rules are calculated for each month of the ten-year, in-sample period. The rules in the highest (lowest) decile of the respective performance criteria (mean return, terminal wealth, and Sharpe ratio) over the in-sample period identify the stocks selected for investment in the LONG (SHORT) portfolio in the first out-of-sample period from July 1974 to June 1975. The ten-year in-sample window is then rolled forward one year, and the process is repeated. Panel A reports the mean quintiles selected for the LONG portfolio under each criterion, and Panel B for the SHORT portfolio. Standard deviations of selected quintiles are given in parentheses.

	<u>Mean Quintiles Selected</u>			
	B/M	SIZE	LAGRET	BETA
A. LONG portfolio				
Mean-return criterion	4.3 (1.0)	1.3 (0.7)	3.2 (1.3)	3.5 (1.3)
Terminal-wealth criterion	4.4 (0.9)	1.3 (0.7)	3.2 (1.3)	3.0 (1.3)
Sharpe-ratio criterion	4.6 (0.7)	2.3 (1.3)	2.8 (0.9)	1.5 (0.6)
B. SHORT Portfolio				
Mean-return criterion	1.4 (0.6)	4.2 (0.8)	2.3 (1.3)	3.5 (1.3)
Terminal-wealth criterion	1.3 (0.5)	3.9 (0.9)	2.3 (1.4)	3.9 (1.2)
Sharpe-ratio criterion	1.3 (0.4)	3.7 (1.1)	2.3 (1.5)	4.3 (1.0)

Table II

Out-of-Sample Performances under the Mean-Return Criterion

Using the four variables, B/M, SIZE, LAGRET, and BETA, a time series of monthly out-of-sample returns is generated from July 1974 through December 1997. Consider the first in-sample period, which extends from July of 1964 to June of 1974. At the beginning of July of each year 1964 to 1973, all NYSE-, AMEX-, and -NASDAQ-listed stocks are sorted into quintiles based on the four variables separately. Equally-weighted returns to each of the resulting 170 one-way and two-way rules are calculated for each month of the ten-year, in-sample period. The rules in the highest (lowest) decile of mean monthly returns over the in-sample period identify the stocks selected for the out-of-sample LONG (SHORT) portfolio in the first out-of-sample period from July 1974 to June 1975. Monthly returns are calculated for the LONG (SHORT) portfolio over the out-of-sample period. The ten-year in-sample window is then rolled forward one year, and the process is repeated for each of the 23.5 out-of-sample years (July 1974 to December 1997). The out-of-sample COMBINED portfolio is defined as the LONG minus the SHORT.

The mean equally-weighted monthly returns (standard deviation of monthly returns) to the LONG, SHORT, and COMBINED portfolios are reported below unadjusted and adjusted for trading costs. The method of estimating the dynamic, stock-specific trading costs is described in Section I.C. The LONG (SHORT) portfolio's mean monthly return is compared to the return of the equally-weighted (EW) market index. The COMBINED portfolio's mean monthly return is compared to zero. The Sharpe ratios of the LONG and SHORT are compared to the Sharpe ratio of the EW index. The *t*-tests for comparing mean returns and Sharpe ratios are robust to heteroskedasticity and autocorrelation (Gallant, 1987). The Jensen's alpha and Sharpe ratio are reported for the LONG and SHORT portfolios; only the Jensen's alpha is reported for the COMBINED.

“_” indicates that the measure is not applicable.

	Unadjusted			Adjusted for Trading Costs		
	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe Ratio	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe Ratio
EW Market	1.41 (5.40)	_	0.15	_	_	_
LONG	1.54* (5.76)	0.09	0.17	1.41 (5.82)	-0.05	0.14
SHORT	1.38 (5.47)	-0.02	0.15	1.46 (5.47)	0.05	0.16
COMBINED	0.16 (1.93)	0.12	_	-0.05 (1.99)	-0.10	_

***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

**Market-Adjusted In-Sample and Out-of-Sample
Returns on the LONG Portfolio
under the Mean-Return Criterion**

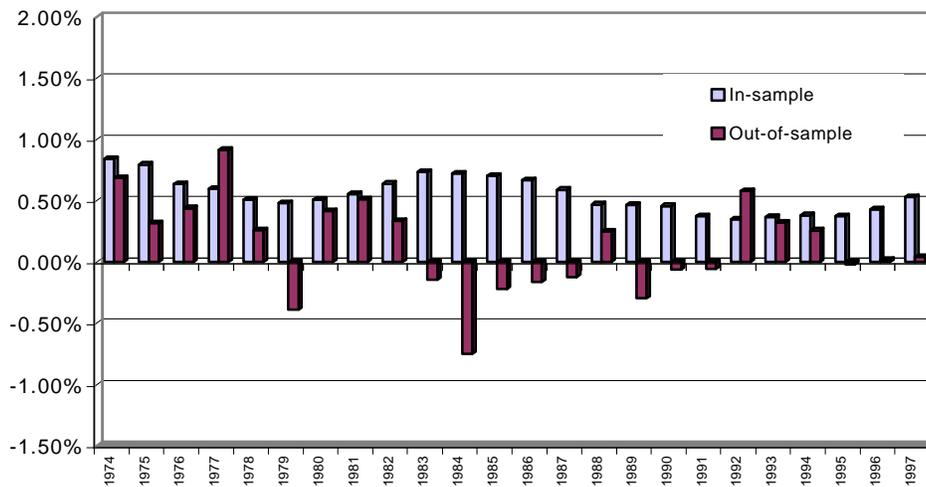


Figure 3: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the LONG portfolio under the mean-return criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

**Market-Adjusted In-Sample and Out-of-Sample
Returns on the SHORT Portfolio
under the Mean-Return Criterion**

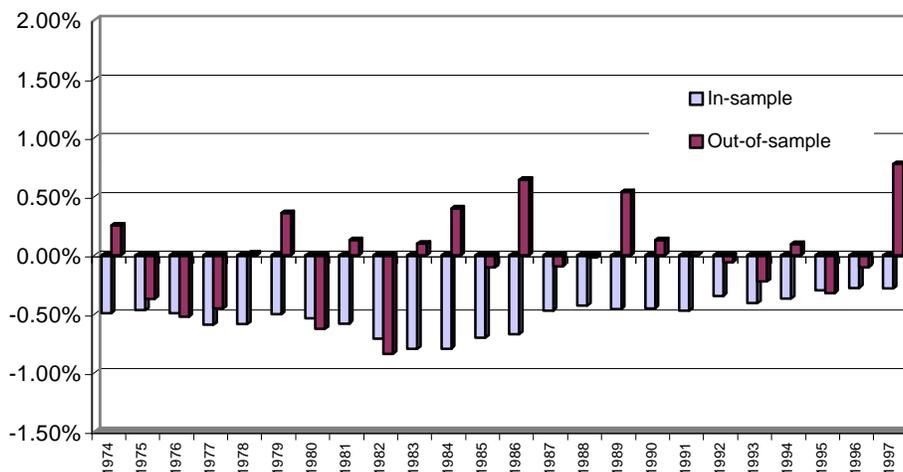


Figure 4: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the SHORT portfolios under the mean-return criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

**In-Sample and Out-of-Sample
Returns on the COMBINED Portfolio
under the Mean-Return Criterion**

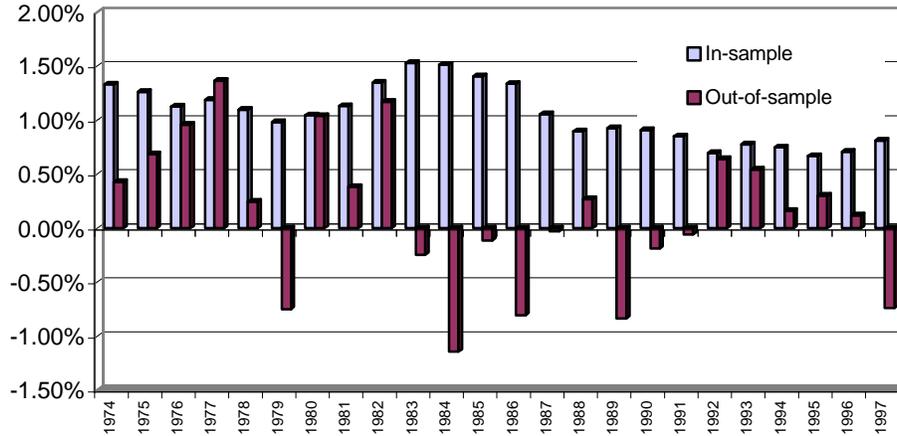


Figure 5: The in-sample and corresponding out-of-sample mean monthly returns to the COMBINED portfolios under the mean-return criterion are plotted for each rolling ten-year in-sample period. The date indicates the last year of the in-sample period.

Table III**Out-of-Sample Performances under the Terminal-Wealth Criterion**

Using the four variables, B/M, SIZE, LAGRET, and BETA, a time series of monthly out-of-sample returns is generated from July 1974 through December 1997 as described in Table II, with one exception; the LONG (SHORT) is comprised of the rules in the top (bottom) decile of terminal wealth in-sample.

“_” indicates that the measure is not applicable.

	Unadjusted			Adjusted for Trading Costs		
	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe's Ratio	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe's Ratio
EW Market	1.41 (5.40)	–	0.15	–	–	–
LONG	1.54* (5.48)	0.14*	0.18*	1.42 (5.33)	0.00	0.15
SHORT	1.36 (5.77)	-0.09	0.14	1.44 (5.78)	-0.01	0.15
COMBINED	0.18 (1.87)	0.23**	–	-0.02 (1.91)	0.02	–

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

**Market-Adjusted In-Sample and Out-of-Sample
Returns on the LONG Portfolio
under the Terminal-Wealth Criterion**

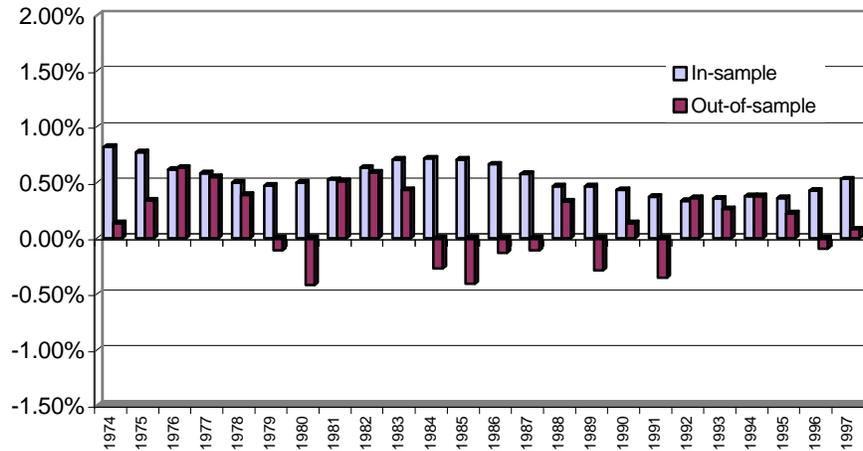


Figure 6: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the LONG portfolio under the terminal-wealth criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

**Market-Adjusted In-Sample and Out-of-Sample
Returns on the SHORT Portfolio
under the Terminal-Wealth Criterion**

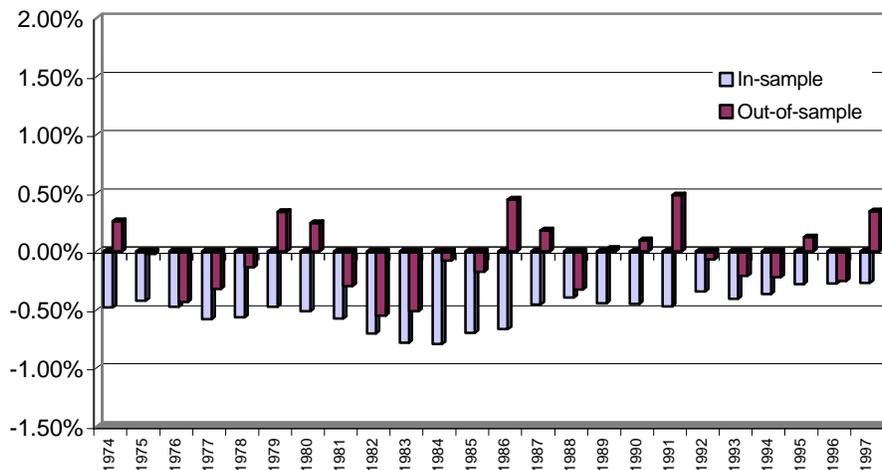


Figure 7: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the SHORT portfolios under the terminal-wealth criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

In-Sample and Out-of-Sample Returns on the COMBINED Portfolio under the Terminal-Wealth Criterion

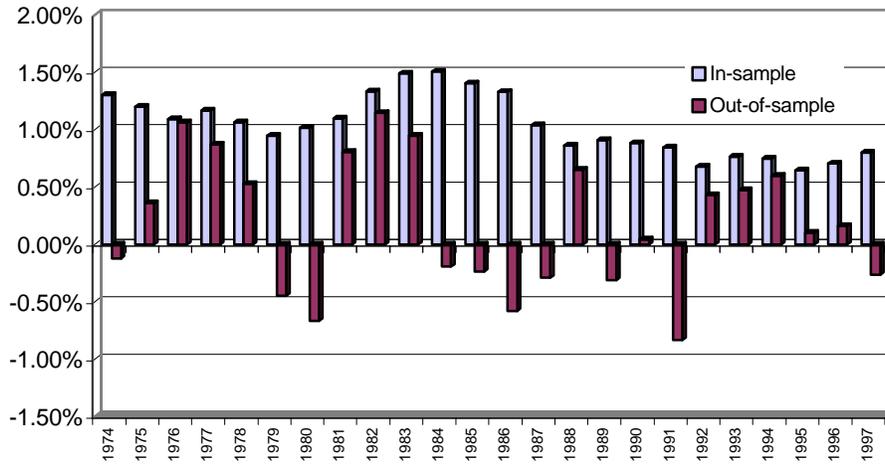


Figure 8: The in-sample and corresponding out-of-sample mean monthly returns to the COMBINED portfolios under the terminal-wealth criterion are plotted for each rolling ten-year in-sample period. The date indicates the last year of the in-sample period.

Table IV**Out-of-Sample Performances under the Sharpe-Ratio Criterion**

Using the four variables, B/M, SIZE, LAGRET, and BETA, a time series of monthly out-of-sample returns is generated from July 1974 through December 1997 as described in Table II, with one exception; the LONG (SHORT) is comprised of the rules in the top (bottom) decile of Sharpe ratios in-sample.

“_” indicates that the measure is not applicable.

	Unadjusted			Adjusted for Trading Costs		
	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe's Ratio	Mean Monthly Return (%) (std. dev.)	Jensen's Alpha	Sharpe's Ratio
EW Market	1.41 (5.40)	_	0.15	_	_	_
LONG	1.60** (4.88)	0.28***	0.21***	1.51 (4.91)	0.19***	0.19***
SHORT	1.31 (6.02)	-0.18**	0.12**	1.40 (6.01)	-0.09	0.14
COMBINED	0.29** (2.19)	0.47***	_	0.11 (2.21)	0.28**	_

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

**Market-Adjusted In-Sample and Out-of-Sample Returns
on the LONG Portfolio
under the Sharpe-Ratio Criterion**

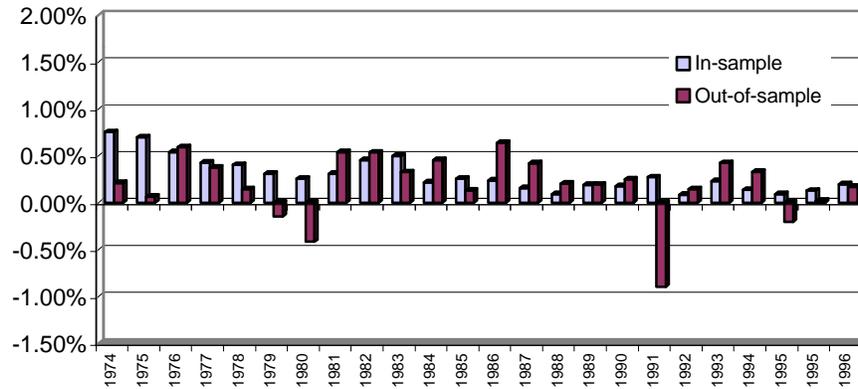


Figure 9: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the LONG portfolio under the Sharpe-ratio criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

**Market-Adjusted In-Sample and Out-of-Sample Returns on the SHORT Portfolio
under the Sharpe-Ratio Criterion**

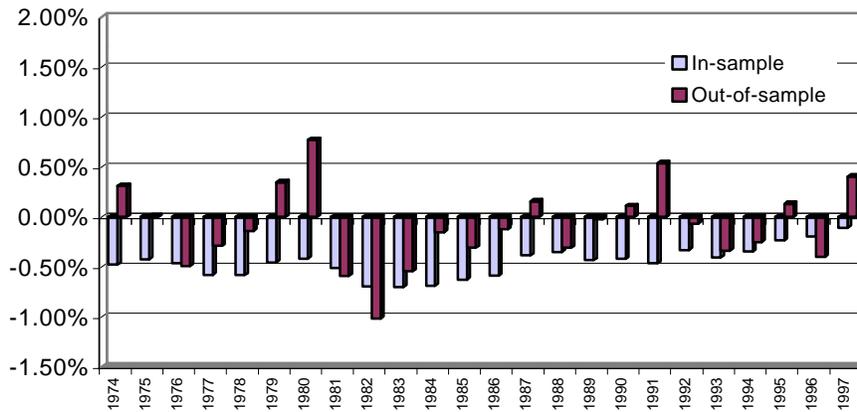


Figure 10: The in-sample and corresponding out-of-sample mean monthly market-adjusted returns to the SHORT portfolios under the Sharpe-ratio criterion are plotted for each rolling ten-year in-sample period. The contemporaneous monthly return to the EW index is subtracted from the in-sample and out-of-sample mean monthly returns, respectively. The date indicates the last year of the in-sample period.

**Market-Adjusted In-Sample and Out-of-Sample
Returns on the COMBINED Portfolio
under the Sharpe-Ratio Criterion**

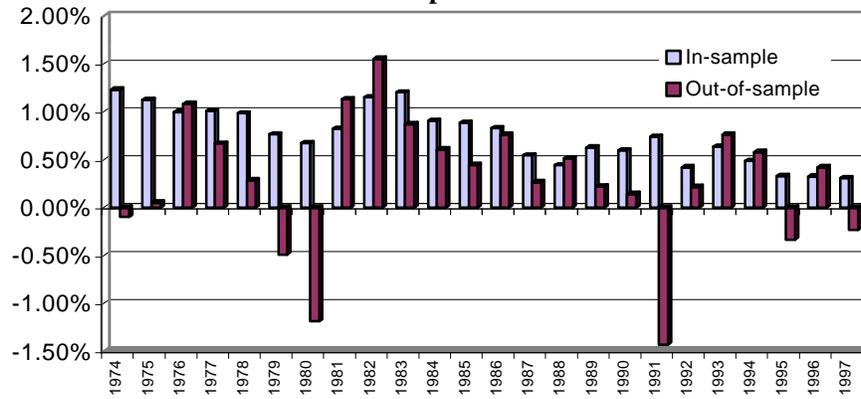


Figure 11: The in-sample and corresponding out-of-sample mean monthly returns to the COMBINED portfolios under the Sharpe-ratio criterion are plotted for each rolling ten-year in-sample period. The date indicates the last year of the in-sample period.

**Terminal Wealths
of \$1 invested in each portfolio on July 1, 1974**

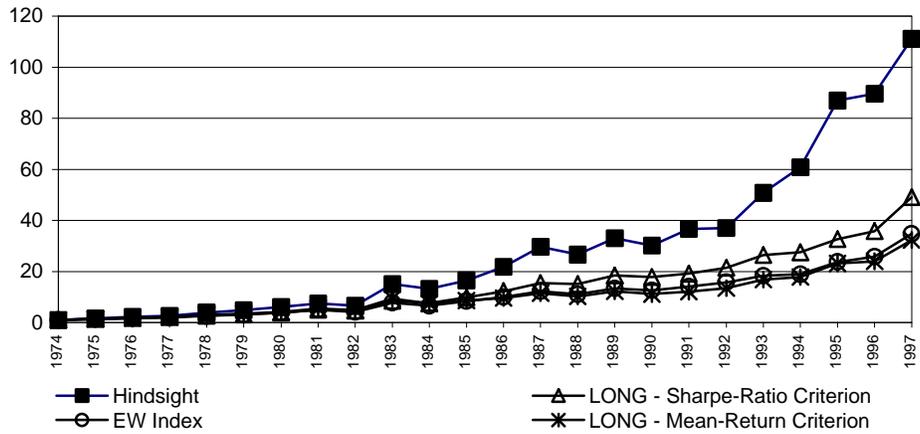


Figure 12: The terminal wealths of one dollar invested in each portfolio on July 1, 1974 are plotted through December 1997. The Hindsight portfolio is comprised of all stocks that simultaneously are in the highest quintile of B/M, the lowest quintile of SIZE, and the highest quintile of LAGRET. The LONG portfolios under the mean-return criterion and the Sharpe-ratio criterion are described in Tables II and IV, respectively, and are adjusted for trading costs. The equally-weighted (EW) index is comprised of all available stocks in our dataset.

Table V

Robustness Checks

Listed below are the specifications, identified by categories, which are examined as alternatives to the base-case out-of-sample method employed earlier (Tables II, III, and IV). Each of these alternatives is employed separately to the base case under each of the three selection criteria (mean-return, terminal-wealth, and Sharpe-ratio).

-
- A. In-sample window length
 - Expanding, all prior years

 - B. Rule universe
 - First and fifth quintiles only (32 rules)
 - Three-way and four-way rules also included (1295 rules)

 - C. Number of rules selected
 - Best five percent of the rules
 - Best rule

 - D. Trading costs
 - 0.50% one-way
 - 1.00% one-way

 - E. Passive Benchmarks
 - Value-weighted CRSP index
 - S&P 500 index
-