Federal Reserve Bank of St. Louis

Working Paper —99-015

Using Genetic Algorithms to Find Technical Trading Rules:
A Comment on Risk Adjustment

Christopher J. Neely

September 16, 1999

Abstract: Allen and Karjalainen (1999) used genetic programming to develop optimal ex ante trading rules for the S&P 500 index. They found no evidence that the returns to these rules were higher than buy-and-hold returns but some evidence that the rules had predictive ability. This comment investigates the risk-adjusted usefulness of such rules and more fully characterizes their predictive content. These results extend Allen and Karjalainen's (1999) conclusion by showing that although the rules' relative performance improves, there is no evidence that the rules significantly outperform the buy-and-hold strategy on a risk-adjusted basis. Therefore, the results are consistent with market efficiency. Nevertheless, risk-adjustment techniques should be seriously considered when evaluating trading strategies.

Senior Economist, Research Department
Federal Reserve Bank of St. Louis
St. Louis, MO 63011
(314) 444-8568 (o), (314) 444-8731 (f), neely@stls.frb.org

Primary Subject Code: G0 - Financial Economics
Secondary Subject Code: G14 - Information and Market Efficiency
Keywords: technical analysis, genetic programming, trading rules, stock prices,

---

The views expressed are those of the author and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, or the Federal Reserve System. The author thanks Kent Koch for research and programming assistance, Chuck Whiteman and Paul Weller for comments, Franklin Allen and Risto Karjalainen for making their programs available and for correspondence on the programs. Any errors are my own.

Using a technique known as genetic programming (Koza, 1992), Allen and Karjalainen (1999)—hereafter AK—searched for optimal ex ante technical trading rules on daily S&P500 data over the period 1929 through 1995. They found that the transactions cost-adjusted returns to these rules failed to exceed the returns to a buy-and-hold strategy—despite the exclusion of dividends from the stock return—and that the market was efficient in this sense.[1] There was, however, some evidence of predictability in returns as the rules tended to be in the market during periods of high returns and out of the market during periods of low returns. Although AK attributed this predictability to low order serial correlation in the stock index, they speculated that the rules might be useful on a risk-adjusted basis despite their lower returns.

The goal of this comment is two-fold: to examine the value of genetic-programming rules with three common methods of risk adjustment and to more fully characterize the predictability found by the rules. Risk adjustment is essential both for evaluating the usefulness of trading rules and for measuring the consistency of results with market efficiency (Sharpe, 1966; Jensen, 1968; Kho, 1996; Brown, Goetzmann, and Kumar, 1998; Ready, 1998). To evaluate risk-adjusted returns, new sets of rules that maximize risk-adjusted measures like the Sharpe ratio (Sharpe, 1966) and the X* statistic (Sweeney and Lee, 1990) are generated. Also, tests of market timing formally quantify the predictability found by the rules (Cumby and Modest, 1987).

The rules fail to consistently and significantly outperform the buy-and-hold strategy by any risk-adjusted measure. Thus, this exercise extends AK's results to find that risk-adjusted rule returns are consistent with market efficiency. The facts that the market indices used exclude dividends and that some predictability may be due to spurious autocorrelation, only reinforce the negative results.

---

[1] The return to a dynamic strategy—moving in and out of the market—will be reduced less by the exclusion of dividends than will the return to a buy and hold strategy.

METHODOLOGY

Genetic programming is a nonlinear search procedure for problems in which the solution may be represented as a computer program or decision tree (Koza, 1992). Like its cousin, the genetic algorithm (Holland, 1975), genetic programming uses the principles of parallel search and natural selection to search for candidate solutions to problems of interest.[2] Essentially, a computer randomly generates a population of candidate solutions—expressible as decision trees—to a problem of interest. The rules are required only to be well defined and to produce output appropriate to the problem of interest—a buy/sell decision in the present case. Of course, most of these random solutions will be quite poor, but some, purely by chance, will "fit" the in-sample data reasonably well, generating excess returns. The computer then allows the population to "evolve" using reproduction and mutation operators. Reproduction mixes subtrees of the population while mutation replaces subtrees with new, randomly generated subtrees. More fit (profitable) members of the population have a greater chance to reproduce while less fit members have a greater chance of being replaced. In this way the genetic program searches promising areas of the solution space by evolving a population of rules that tends to become more adept at solving the problem in successive generations.

Genetic programming minimizes—but does not eliminate—the problem of "data snooping" by searching for optimal ex ante rules, rather than rules known to be used by traders. Ready (1998), for example, argues that testing rules known to be widely used by technical traders—as done by Brock, Lakonishok and Lebaron (1992)—is a form of data snooping. This

[2] Genetic algorithms require the solution to the problem to be encoded as fixed length character strings rather than as decision trees or computer programs as in genetic programming.

practice is likely to produce spurious evidence of technical trading profits because the rules are widely used precisely because they would have been profitable on past data.[3]

This paper uses programs made publicly available by AK to maintain maximum comparability to their results.[4] One difference between AK's procedures and those used here should be noted: Interest rates are treated differently. AK's code attributes one day's (1/365) interest rate to the rules during each business day—not calendar day—they are out of the market. This practice understates the returns to the genetic programming rules by 0.5 percent or less. In this paper, rules earn interest on calendar days—not business days—they are out of the market.[5] Table 1 summarizes some of the important parameters of interest chosen by AK for their implementation of the genetic program. AK provide more information on the program.

AK used genetic programming to construct trading rules on daily data from the S&P500 from 1929 to 1995, using ten overlapping in-sample estimation periods (1929-35, 1934-40, 1939-45… 1974-80). Each in-sample period of seven years was broken down into a training period (five years) and a selection period (two years) to alleviate the problem of overfitting the data. Ten independent rules were generated for each set of in-sample data. Rules with positive excess returns over the buy-and-hold strategy in the training period were saved for out-of-sample testing over the remainder of the data (1936-95, 1941-95…1981-95).

Each day, the trading rules generated by the genetic program observe prices and generate a buy or sell signal indicating the position to take (the same day). The buy and sell signals are used along with stock prices and 30-day T-Bill interest rates to compute the continuously

[3] Neely, Weller and Dittmar (1997) and Neely and Weller (1999b) have applied genetic programming to find trading rules in the dollar foreign exchange market and the European Monetary System, respectively. Neely and Weller (1999a) have also permitted genetic programs to use additional information—central bank intervention—as inputs to the trading rule.
[4] Programs written by Rob Dittmar produced results similar to those generated by the AK programs.
[5] The author thanks Kent Koch for observing this and Risto Karjalainen for confirming it in private communication.

compounded excess return of the rule over the return to a buy-and-hold strategy in the stock market. This excess return over the buy-and-hold strategy at time t is given by:

$$xsr_t = (z_t - 1)\left[\ln\left(\frac{P_{t+1}}{P_t}\right) - \ln(1 + i_t)\right] \tag{1}$$

where z_t is an indicator variable taking the value 1 if the rule is in the market or 0 if the rule is in T-Bills, P_t is the stock index and i_t is the interest rate on the 30-day Treasury Bill earned from business day t to business day $t+1$. The cumulative excess return—also called the "fitness"—for a trading rule from time zero to time T is the sum of the daily excess returns less a proportional transactions cost. AK considered transactions costs of 0.1 percent, 0.25 percent and 0.5 percent. For brevity's sake, this comment concentrates on transactions costs of 0.25 percent.

RESULTS

Comparison with AK's Results

Table 2 shows the out-of-sample results from implementing a uniformly weighted portfolio based on all the good rules found in-sample.[6] This is similar to AK's baseline case. The rules are assessed a 0.25 percent transactions cost for changing positions and information on day t is used to trade the same day. As in AK (compare to Table 2, Panel A in AK), the rules generally failed to produce positive excess returns over the buy-and-hold strategy in the sample. With the exception of the period 1949-55, for which no good in-sample rules were found, the out-of-sample performance was similar to that found by AK.[7] While AK found only one period in which the mean excess return over the buy-and-hold strategy was positive, the current exercise found two such periods. The rules were long in the market about 50 percent of the time and

[6] Results for median portfolio rules are broadly similar to—slightly better than—those of the uniform portfolio rules. For the sake of brevity, they will not be reported separately. The median portfolio rule goes into the market if most of the N rules are in the market, otherwise it stays out of the market.

traded 7.7 times a year, on average, though the figures varied widely with the in-sample period. The 1974-80 period produced uninteresting rules that stayed out of the market almost all the time.

Column 4 of Table 2 shows the mean annual return to the market when the rules are in the market less the mean annual market return when the rules are out of the market (r_b-r_s). Although there is no measure of statistical significance, positive numbers favor the proposition that the rules have some market timing ability. While AK found that rules from 7 of 10 in-sample periods had market timing ability by this measure, the results in this paper are slightly more pessimistic, showing that only 5 of 9 have positive r_b-r_s. Because the rules' buy/sell decisions could be closely replicated by moving average rules, AK concluded that the genetic programming rules were taking advantage of low-order serial correlation. AK speculated that the rules might be of use to a risk-averse speculator, but did not seriously explore that possibility.

Risk Adjustment

The criterion of judging the rules to be useful only if they generate a return that exceeds the buy-and-hold return is neither necessary nor sufficient to conclude that the rules do not violate the efficient markets hypothesis (EMH).[8] The EMH is usually interpreted as meaning that asset prices reflect information to the point where the potential risk-adjusted excess returns do not exceed the transactions costs of acting (trading) on that information (Jensen, 1978). This is potentially important because dynamic strategies, such as those found by the genetic program, are often out of the market and therefore may bear much less risk than the buy-and-hold strategy.

---

[7] There are two reasons why the results will not exactly replicate those found by AK: 1) Genetic programming is inherently stochastic, generating and recombining populations probabilistically; and 2) interest rate returns were treated differently in this analysis.

[8] Brown, Goetzmann, and Kumar (1998) find that risk adjustment is crucial in evaluating Dow Theory recommendations.

Although there is no universally accepted method of adjusting returns for risk, this paper will employ three commonly used techniques: the Sharpe ratio, the X* measure, and Jensen's α.

The Sharpe ratio—the expected excess return per unit of risk for a zero-investment strategy (Campbell, Lo and MacKinlay, 1997)—is usually expressed in annual terms as the annual excess return over the riskless rate to a portfolio over that excess return's annual standard deviation. The excess return over the riskless rate to the rules at time *t* is given by:

$$r_t = z_t\left[\ln\left(\frac{P_{t+1}}{P_t}\right) - \ln(1+i_t)\right] \tag{2}$$

where z_t is an indicator variable that takes the value 1 when the rule is in the market and 0 otherwise. Although the rules may have lower returns than the buy-and-hold strategy, lower volatility may permit the returns to be leveraged up to exceed the buy-and-hold return with similar risk.[9]

The average Sharpe ratio of the transactions cost-adjusted genetic programming rules is about 0.02, lower than the average 0.13 Sharpe ratio—the index doesn't include dividends—for the buy-and-hold strategy over the ten subsamples.[10] Therefore, positive returns in excess of a buy-and-hold strategy could not be generated by leveraging up the sizes of positions held by the genetic programming rule.[11]

Of course, the rules trained on an excess return criterion may not be the best risk-adjusted rules. To determine whether technical trading rules can produce better risk-adjusted returns than the buy-and-hold strategy, ideally we must <u>train</u> a set of rules using the Sharpe ratio as the fitness

[9] Ready (1998) has questioned whether the strategy of leveraging returns is implementable, as the investor would have to know—or predict—the ex post moments to compute the proper amount of leverage.

[10] Jorion and Goetzmann (1999) estimate that dividends made up much of the total return to U.S. equities over the period 1921 to 1995. The average Sharpe ratio for the buy-and-hold strategy over the 10 overlapping out-of-sample subsamples is 0.13 while the Sharpe ratio from 1929 through 1995 is 0.06.

criterion. The results of this exercise are shown in Table 3. The rules trained on Sharpe ratios failed to produce higher Sharpe ratios on average but they did show greater predictive ability by the standard of the r_b-r_s and X* statistics. They also spent less time in the market (24 percent long).

Sweeney and Lee (1990) developed another risk-adjustment strategy, the X* measure, in the context of the foreign exchange market that may be even more appropriate for equity markets.[12] They show that, in the presence of a constant risk premium, an equilibrium daily risk-adjusted return to a trading rule would be given by:

$$(3) \qquad X^* = \frac{1}{T}\sum_{t=0}^{T-1}\left[z_t \ln\left(\frac{P_{t+1}}{P_t}\right) + (1 - z_t)\ln(1 + i_t)\right] + \frac{n}{2T}\ln\left(\frac{1-c}{1+c}\right) - \left[\frac{p_1}{T}\sum_{t=0}^{T-1}\ln\left(\frac{P_{t+1}}{P_t}\right) + \frac{p_2}{T}\sum_{t=0}^{T-1}\ln(1 + i_t)\right]$$

where z_t, P_t and i_t are defined as before, T is the number of observations, n is the number of one-way trades, c is the proportional transactions cost, p_1 is the proportion of the time spent in the market and p_2 is the proportion of the time spent in T-Bills ($p_1 + p_2 = 1$). Note that the sum of the third and fourth terms estimates the expected return to a zero transactions-cost strategy that randomly is in the market on a fraction p_1 of the days, earning the market premium, and in T-Bills otherwise. The risk-adjusted return—under the null of no timing ability—is the actual return less the expected return. Positive X* statistics are interpreted as evidence of superior risk-adjusted returns.

Most of the annualized X* statistics—net of transactions costs—in Table 2 and Table 3 are negative, indicating that the rules would not have been useful, even by this risk-adjusted

---

[11] Because dynamic strategies are at an inherent disadvantage, as the market return will, on average, exceed the riskless return, Bessembinder and Chan (1998) pursue another strategy to compare trading rules to a market return. They permit rules to use double leverage during periods in which they are in the market.

[12] Sweeney (1988) uses the X* measure in the equity market. Ready (1998) constructs a statistic similar to Sweeney and Lee's (1990) X*. In turn, the test statistic of X* proposed by Sweeney and Lee (1990), is virtually equivalent to the test statistic of the coefficient β_1 in the Cumby-Modest test of market timing if transactions costs are omitted from the X* calculation.

measure. Almost all the X* statistics would have been positive though, if transactions costs had not been netted out. This supports the evidence of predictability suggested by the r_b-r_s statistics.

Table 4 shows the X* statistics from rules trained to maximize X* as the in-sample fitness criteria. There are no trivial X* rules and the rules are very even handed; there are no cases in which the rules are always in or always out of the market. The results are generally superior to those of the rules trained on excess returns. The annualized excess return over the buy and hold is greater than in the benchmark case and the average Sharpe ratio is about the same as the average buy and hold Sharpe ratio over all sample periods (0.12 vs. 0.13). The mean annualized X* statistic is also slightly positive and higher than the average X* statistics from the rules trained with excess returns and the Sharpe ratio as the fitness criterion. However, it should be noted that even positive X* results may be consistent with the EMH in the presence of a time-varying risk premium.

The final risk-adjustment measure considered is Jensen's (1968) α, the return in excess of the riskless rate that is uncorrelated with the excess return to the market.

$$(4) \qquad z_t\left[\ln(P_{t+1}/P_t)-\ln(1+i_t)\right]-\frac{n}{2T}\ln\left(\frac{1+c}{1-c}\right)=\alpha+\beta_M\left[\ln(P_{t+1}/P_t)-\ln(1+i_t)\right]+\varepsilon_t$$

If the intercept in equation (4)—α—is positive and significant, then the trading rule produces excess returns that cannot be explained by correlation with the market. To measure Jensen's α, returns to the market and to the trading rules were aggregated over nonoverlapping 30-day periods and regression (4) was performed by OLS using annualized returns. Results for each set of rules are shown in the 9th and 10th columns of Table 2 through Table 4. Again, the only set of rules for which the average α is positive are those trained on X*, and these are never significant at conventional levels.

CHARACTERIZING LOW ORDER SERIAL CORRELATION

After finding that moving average rules could closely approximate the GP rules' buy/sell behavior, AK attributed the predictability found by their GP trading rules to "low order serial correlation" in the returns (Campbell, Lo and MacKinlay 1997). One might speculate that a simple time series model of returns could produce better decisions than the GP. To test this prediction and to attempt to better characterize the nature of the predictability found by AK, a variety of ARMA models were fit to the in-sample excess returns and the best in-sample models and parameters were chosen by the Akaike, Schwarz and excess return criteria. The best models were used to generate trading signals—like the genetic programs—during the out-of-sample periods. Table 5 shows that the non-trivial ARIMA models are even less successful than the rules constructed by genetic programming. If low-order serial correlation generates the predictability, the genetic rules are apparently more successful at estimating it than are standard ARIMA models.

Finally, Cumby-Modest tests of market timing ability are used to more formally determine whether the rules have predictive content. The statistical significance of the coefficient (β_1) in the regression of excess returns on signals from the trading rule summarizes the rules' one-day ahead timing ability:

$$250 \cdot 100 \cdot \left[\ln\left(\frac{P_{t+1}}{P_{t+1}} \right) - \ln(1 + i_t) \right] = \beta_0 + \beta_1 z_t + \varepsilon_t \,. \tag{5}$$

Table 6 presents strong evidence that the rules do possess predictive ability: 22 of the 25 available β_1 coefficients are positive and 14 of those are significant at the 5 percent level. Of the three fitness criteria, the X* criteria seems to have produced the rules with the most predictive content. These results illustrate the well-known result that profitability is not necessary for a rule to have predictive content.

CONCLUSION

This paper has investigated the results of AK (1999) to determine if ex ante optimal rules created by genetic programming are useful on a risk-adjusted basis. Although risk-adjustment improves the relative attractiveness of the rules, neither Sharpe ratios nor Sweeney and Lee's X* statistic, nor Jensen's α provide evidence that rules developed by genetic programming would have been useful even to risk-averse speculators, contrary to AK's reasonable speculation. Rules trained on X* measures had the best risk-adjusted performance by all the measures, approximately equaling the buy-and-hold return performance. Of course, risk is difficult to measure and any risk adjustment is subject to criticism. Nevertheless, this comment argues that trading rule results must be carefully interpreted in light of risk adjustment.

It is likely that the inclusion of dividends in the stock index, the removal of spurious autocorrelation from the index returns, or accounting for price slippage would only strengthen the negative results of this exercise.

References

Allen, F., Karjalainen, R., 1999. Using Genetic Algorithms to Find Technical Trading Rules. Journal of Financial Economics 51, 245-271.

Bessembinder, H., Chan, K., 1998. Market Efficiency and the Returns to Technical Analysis. Financial Management 27, 5-17.

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. Journal of Finance 47, 1731-1764.

Brown, S.J., Goetzmann, W.N., Kumar, A., 1998. The Dow Theory: William Peter Hamilton's Track Record Reconsidered. Journal of Finance 53, 1311-1333.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Cumby, R.E., Modest, D.M., 1987. Testing for Market Timing Ability: A Framework for Forecast Evaluation. Journal of Financial Economics 19, 169-189.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI.

Jensen, M.C., 1968. Problems in Selection of Security Portfolios: The Performance of Mutual Funds in the Period 1945-1964. Journal of Finance 23, 389-416.

Jensen, M.C., 1978. Some Anomalous Evidence Regarding Market Efficiency. Journal of Financial Economics 6, 95-101.

Jorion, P., Goetzmann, W.N., 1999. Global Stock Markets in the Twentieth Century. Journal of Finance 54, 953-980.

Kho, B.C., 1996. Time-Varying Risk Premia, Volatility, and Technical Trading Rule Profits: Evidence from Foreign Currency Futures Markets. Journal of Financial Economics 41, 249-290.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA.

Neely, C.J., Weller, P.A., 1999a. Technical Analysis and Central Bank Intervention. Federal Reserve Bank of St. Louis Working Paper 97-002B.

Neely, C.J., Weller, P.A., 1999b. Technical Trading Rules in the European Monetary System. Journal of International Money and Finance 18, 429-458.

Neely, C.J., Weller, P.A., Dittmar, R.D., 1997. Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach. Journal of Financial and Quantitative Analysis 32, 405-426.

Ready, M.J., 1998. Profits from Technical Trading Rules. Unpublished manuscript. University of Wisconsin, Madison.

Sharpe, W.F., 1966. Mutual Fund Performance. Journal of Business 39, 119-138.

Sweeney, R.J., 1988. Some New Filter Tests, Methods and Results. Journal of Financial and Quantitative Analysis 23, 285-300.

Sweeney, R.J., Lee, E.J.Q., 1990. In: Aggarwal R., and Lee C.F. (Eds.), International Dimensions of Securities and Currency Markets. Advances in Financial Planning and Forecasting Series Vol. 4, Part A. JAI Press, Greenwich, CT, pp. 55-79.

Table 1: Genetic programming parameters of interest for AK's implementation

Parameter	AK's Choice
Size of a generation	500
Termination criterion	50 generations or no improvement for 25 generations
Probability of selection for reproduction	$\sqrt{\frac{\text{rank in population}}{\text{size of population}}} - \sqrt{\frac{\text{rank in population} - 1}{\text{size of population}}}$
arithmetic functions	+, -, *, /, norm, constant between (0,2)
Boolean operators	"if-then", "and", "or", "<", ">", "not", "true", "false"
functions of the data	"moving average", "local maximum", "local minimum", "lag of stock index", "current stock index"

Table 2: Uniform portfolio results from the benchmark case

In-sample period	# of good rules in-sample	Annualized Excess over buy-and-hold	Annualized rb-rs	Sharpe ratio	X* statistic	Trades per year	% long	alpha	s.e.	Annualized In-sample B&H	Annualized Out-of-sample B&H
1929-35	10.00	-2.49	-13.12	-0.02	-0.46	1.85	0.17	-0.70	0.29	-8.53	6.34
1934-40	10.00	-0.83	-11.26	0.27	0.27	2.84	0.65	0.63	0.79	0.85	7.35
1939-45	7.00	-3.28	-0.64	-0.12	-1.26	4.32	0.18	-2.06	1.09	3.98	7.10
1944-50	3.00	0.64	15.41	0.29	1.12	5.45	0.81	1.82	1.06	8.01	7.52
1949-55	0.00	NA	NA	NA	NA	NA	NA	NA	NA	15.63	6.47
1954-60	10.00	0.24	24.82	0.08	0.29	1.87	0.93	0.42	0.44	12.06	6.70
1959-65	8.00	-0.77	44.89	-0.08	-0.79	22.63	0.85	-1.15	1.28	7.31	6.28
1964-70	10.00	-1.56	24.42	-0.06	-1.29	21.54	0.70	-1.77	1.32	2.96	7.52
1969-75	10.00	-3.29	102.98	-0.18	-1.19	8.36	0.22	-1.38	0.72	-2.00	9.50
1974-80	10.00	-3.42	-5.96	NA	-0.01	0.27	0.00	NA	NA	4.67	9.97
mean	7.80	-1.64	20.17	0.02	-0.37	7.68	0.50	-0.52	0.87	4.49	7.47

Notes: Column 2 provides the number of rules (out of 10 trials) that had positive training period returns. Column 3 is the annualized out-of-sample excess return, net of transactions cost, to the portfolio rule while column 4 is the mean difference between average market returns on days that the rules were in the market and the days that they were out of the market. The portfolio mean return over the riskless rate, net of transactions cost, divided by the standard deviation of the portfolio return is in column 5. Column 6 shows the annualized X* risk-adjusted return measure, net of transactions cost. Columns 7 and 8 show the mean number of trades per year and the mean proportion of time spent in the market. Jensen's alpha and its standard error are in columns 9 and 10. The annualized buy-and-hold returns are shown in columns 11 and 12.

Table 3: Results generated using the Sharpe ratio as the fitness criterion

In-sample period	# of good rules in-sample	Annualized Excess over buy-and-hold	Annualized rb-rs	Sharpe ratio	X* statistic	Trades per year	% long	alpha	s.e.	Annualized In-sample B&H	Annualized Out-of-sample B&H
1929-35	10.00	-2.47	-9.25	NA	-0.02	0.08	0.00	NA	NA	-8.53	6.34
1934-40	10.00	-2.85	113.55	0.11	-0.09	1.52	0.11	-0.08	0.34	0.85	7.35
1939-45	10.00	-2.96	6.72	-0.24	-0.62	1.66	0.05	-0.87	0.34	3.98	7.10
1944-50	10.00	-1.63	64.01	0.25	0.61	0.97	0.09	1.05	0.75	8.01	7.52
1949-55	10.00	-0.17	4.87	0.07	-0.03	2.58	0.86	-0.05	0.62	15.63	6.47
1954-60	10.00	-0.35	5.99	0.06	0.08	6.01	0.44	0.21	1.43	12.06	6.70
1959-65	10.00	-0.39	10.07	-0.07	-0.46	11.40	0.51	-0.68	1.34	7.31	6.28
1964-70	10.00	-1.32	22.20	-0.06	-0.76	9.20	0.39	-1.07	0.89	2.96	7.52
1969-75	10.00	-2.69	78.37	NA	-0.01	1.01	0.00	NA	NA	-2.00	9.50
1974-80	10.00	-3.42	-9.97	NA	0.00	0.00	0.00	NA	NA	4.67	9.97
mean	10.00	-1.82	28.66	0.02	-0.13	3.44	0.24	-0.21	0.82	4.49	7.47

Notes: see the notes to Table 2.

Table 4: Results generated using the X* measure as the fitness criterion.

In-sample period	# of good rules in-sample	Annualized Excess over buy-and-hold	Annualized rb-rs	Sharpe ratio	X* statistic	Trades per year	% long	alpha	s.e.	Annualized In-sample B&H	Annualized Out-of-sample B&H
1929-35	10.00	-1.37	1.66	0.20	0.26	3.11	0.33	0.67	0.78	-8.53	6.34
1934-40	10.00	-0.77	6.62	0.29	0.70	3.19	0.53	1.29	1.17	0.85	7.35
1939-45	10.00	-2.49	-13.12	0.00	-0.93	2.34	0.37	-1.60	0.76	3.98	7.10
1944-50	10.00	-0.07	18.09	0.30	1.02	3.87	0.56	1.81	1.01	8.01	7.52
1949-55	1.00	-0.57	2.72	0.05	-0.16	4.00	0.58	-0.15	1.47	15.63	6.47
1954-60	10.00	-0.50	6.19	0.04	-0.02	5.93	0.37	0.07	1.35	12.06	6.70
1959-65	10.00	0.92	16.27	0.08	0.87	4.83	0.66	1.25	0.93	7.31	6.28
1964-70	10.00	-1.40	20.23	-0.06	-0.96	14.94	0.53	-1.34	1.08	2.96	7.52
1969-75	10.00	-2.35	-1.16	0.10	-0.25	2.13	0.22	-0.16	0.61	-2.00	9.50
1974-80	10.00	-2.23	-1.68	0.23	-0.03	2.53	0.36	0.12	0.69	4.67	9.97
mean	9.10	-1.08	5.58	0.12	0.05	4.69	0.45	0.20	0.99	4.49	7.47

Notes: see the notes to Table 2.

Table 5: Results from ARIMA rules

In-sample Period	Search Criterion	AR Order	MA Order	Daily Dummy	Annualized Excess over buy-and-hold	Annualized r_b-r_s	Sharpe ratio	X* statistic	Trades per year	% long
1929-35	AIC	5	5	2	-34.19	15.36	-3.04	-32.62	145.56	0.5750
	SC	1	0	1	-3.69	NA	NA	0.00	0.00	0.0000
	Excess Return	1	2	1	-3.89	46.81	-0.07	-0.22	1.44	0.0030
1934-40	AIC	2	2	1	-34.40	5.21	-3.00	-32.23	134.14	0.5149
	SC	2	2	1	-34.40	5.21	-3.00	-32.23	134.14	0.5149
	Excess Return	1	0	1	-0.17	-2187.39	0.32	-0.17	0.04	0.9999
1939-45	AIC	5	5	3	-34.05	-10.33	-2.90	-32.27	118.79	0.5486
	SC	1	0	1	-0.58	-610.46	0.25	-0.58	0.36	0.9992
	Excess Return	1	0	1	-0.58	-610.46	0.25	-0.58	0.36	0.9992
1944-50	AIC	5	4	2	-25.34	13.85	-2.03	-23.78	108.27	0.6174
	SC	2	0	1	-24.35	-0.26	-1.78	-23.14	92.43	0.7031
	Excess Return	1	0	1	-0.20	-2187.25	0.30	-0.20	0.04	0.9999
1949-55	AIC	3	5	2	-30.43	15.35	-2.62	-29.54	131.82	0.6703
	SC	2	0	2	-22.37	21.52	-1.68	-21.74	102.44	0.7669
	Excess Return	1	0	1	-0.24	-1095.88	0.19	-0.24	0.10	0.9998
1954-60	AIC	4	1	3	-28.12	18.57	-2.25	-27.29	125.33	0.6850
	SC	2	0	2	-24.67	21.02	-1.87	-24.00	112.11	0.7448
	Excess Return	1	0	1	-0.97	-107.73	0.12	-0.96	2.06	0.9958
1959-65	AIC	2	3	2	-25.61	25.65	-2.28	-24.91	123.85	0.6256
	SC	1	1	2	-23.71	22.48	-1.86	-23.27	109.16	0.7680
	Excess Return	1	0	1	-7.68	-11.64	-0.46	-7.55	27.50	0.9346
1964-70	AIC	5	5	3	-33.53	13.82	-3.04	-32.16	142.47	0.5396
	SC	3	2	2	-30.84	10.80	-2.71	-29.56	128.99	0.5708
	Excess Return	2	2	1	-16.05	-11.26	-1.38	-14.61	47.38	0.5150
1969-75	AIC	5	5	2	-24.25	18.42	-1.96	-21.75	105.42	0.4820
	SC	1	1	2	-26.49	11.81	-1.84	-25.36	109.95	0.7652
	Excess Return	2	2	1	-11.67	5.59	-1.26	-7.14	29.71	0.0601
1974-80	AIC	4	3	2	-14.23	-8.02	-0.75	-11.50	38.00	0.4999
	SC	2	1	1	-58.92	0.18	-5.45	-56.17	224.89	0.4951
	Excess Return	4	4	1	-14.06	-8.28	-0.74	-11.34	37.07	0.5015

Notes: Column 2 shows the in-sample model selection criterion. Columns 3 and 4 show the chosen orders of the autoregressive and moving average components. Column 5 summarizes the deterministic component of the model: 1 indicates a simple constant, 2 indicates a weekend dummy on returns while 3 indicates that a full set of day-of-the-week dummies was used. For the other columns, see the notes to Table 2.

Table 6: Cumby-Modest tests of market timing

	Benchmark case			Sharpe ratio rules			X* rules		
	beta	s.e.	p-value	beta	s.e.	p-value	beta	s.e.	p-value
1929-35	0.41	19.07	0.49	NA	NA	NA	16.11	7.53	0.02
1934-40	13.81	6.74	0.02	21.06	15.28	0.08	11.07	4.88	0.01
1939-45	-1.96	6.13	0.63	-14.41	15.84	0.82	-7.15	8.53	0.80
1944-50	23.92	6.04	0.00	83.74	19.24	0.00	23.46	6.68	0.00
1949-55	NA	NA	NA	12.05	9.34	0.10	3.41	4.26	0.21
1954-60	47.55	18.27	0.00	8.43	5.32	0.06	8.16	5.44	0.07
1959-65	66.34	9.54	0.00	21.10	7.70	0.00	36.51	10.83	0.00
1964-70	82.27	13.16	0.00	60.75	18.49	0.00	95.61	17.28	0.00
1969-75	69.28	28.83	0.01	NA	NA	NA	15.12	24.30	0.27
1974-80	NA	NA	NA	NA	NA	NA	26.08	26.16	0.16
	7		6	6		3	9		5

Notes: The three panels show the results of Cumby-Modest tests (see equation (5)) on the benchmark case of excess returns, the case of rules trained on the Sharpe ratio and the X* statistic. The columns of each subpanel show the coefficient, its standard error and its p-value. The final row displays the number of positive betas and the number of p-values less than 0.05.