Guest Lecture, Stat 207, Fall 2020

Bootstrap Reality Check & Technical Trading Rules

David Donoho

2020-11-11

Outline

Financial Reading

- White, Halbert. "A reality check for data snooping." Econometrica 68.5 (2000): 1097-1126.
- **Sullivan, Ryan, Allan Timmermann, and Halbert White. "Data-snooping, technical trading rule performance, and the bootstrap." The Journal of Finance 54.5 (1999): 1647-1691.**
- **Sullivan, Ryan, Allan Timmermann, and Halbert White. "Dangers of data mining: The case of Calendar effects in stock returns." Journal of Econometrics 105.1 (2001): 249-286.**
- Romano, Joseph P., and Michael Wolf. "Stepwise multiple testing as formalized data snooping." Econometrica 73.4 (2005): 1237-1282.
- Bailey, David H., et al. "Pseudo mathematics and financial charlatanism: the effects of backtest overfitting on out-of-sample performance." Notices of the AMS, May 2014

Empirical Stock Market Analysis Terms

- Fixed Historical Database of stock returns
  n stocks, T days
  $y_t(s) = \frac{P_t - P_{t-1}}{P_{t-1}}$ return of stock s over period $(t-1, t]$.
- Portfolio: allocation of capital to stocks $\pi_t(s)$ at time t
  Long: $\pi_t(s) > 0$
  Short: $\pi_t(s) < 0$
  Cap: $Cap = 1'\pi_t$
- Return on investment:

$$r_{t+1}(\pi_t) = y'_{t+1}\pi_t$$

Signal Functions in White et al.

- $S(\mathcal{Y}_t; \beta) \in \{1, 0, -1\}$
  eg: Invest in Recent Winners
  $$S_+(\mathcal{Y}_t) = sgn(Ave_{[t-4,t]} y_u)$$
  eg: Invest in Recent Losers
  $$S_-(\mathcal{Y}_t) = -sgn(Ave_{[t-4,t]} y_u)$$
- Universe of Signal functions
  $$S_k, k = 1, \ldots, K$$
- Comparison of Signal Predictive Ability to a baseline
  $$f_{k,t+1} = \log[1 + y_{t+1} S_k(\mathcal{Y}_t; \beta_{k,t})] - \log[1 + y_{t+1} S_0(\mathcal{Y}_t; \beta_{k,t})]$$
  Example Benchmarks:
  $S_0 = 1$. buy and hold;
  $S_0 \sim \{\pm 1\}$ equiprobable

White 1999, Sullivan, Timmerman, White (200)

Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance

David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu

> Another thing I must point out is that you cannot prove a vague theory wrong. [...] Also, if the process of computing the consequences is indefinite, then with a little skill any experimental result can be made to look like the expected consequences.
>
> —Richard Feynman [1964]

"training set" in the machine-learning literature). The OOS performance is simulated over a sample not used in the design of the strategy (a.k.a. "testing set"). A backtest is *realistic* when the IS performance is consistent with the OOS performance.

When an investor receives a promising backtest

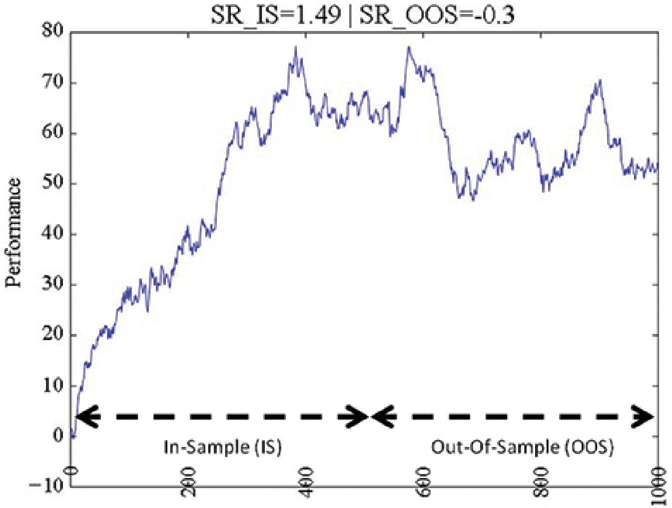


Figure 4. Performance IS vs. performance OOS for one path after introducing strategy selection.

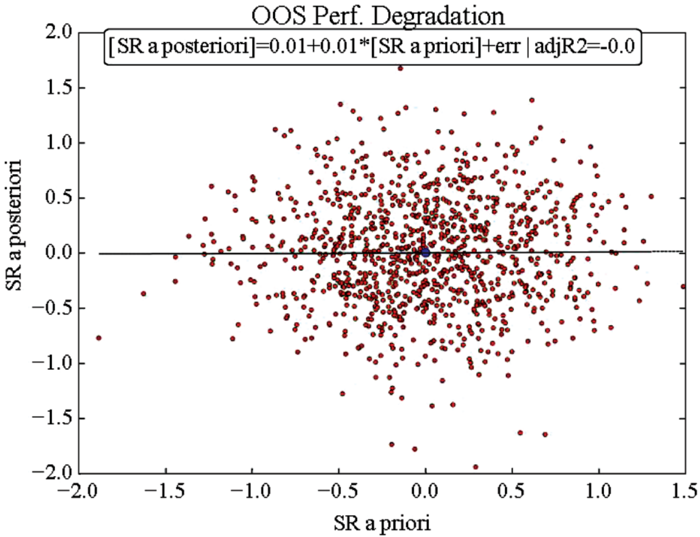


Figure 3. Performance IS vs. OOS before introducing strategy selection.

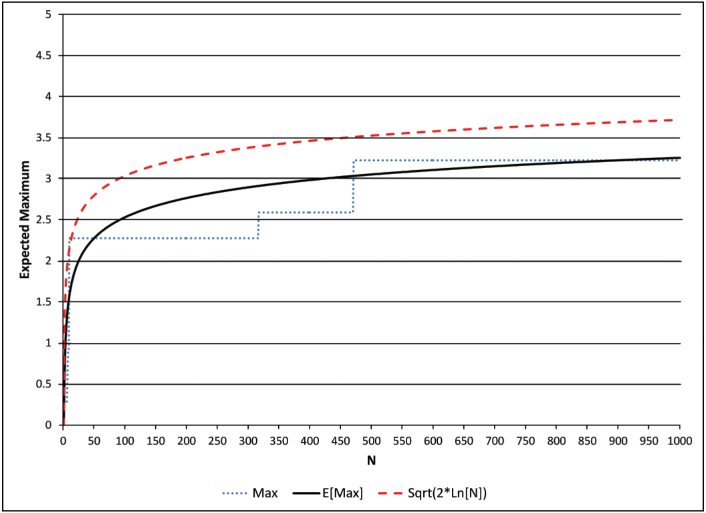


Figure 1. Overfitting a backtest's results as the number of trials grows.

Figure 1 provides a graphical representation of

Calendar Rules in Sullivan, Timmerman, White (1998)

October. This is one of the peculiarly dangerous months to speculate in stocks. The others are July, January September, April, November, May, March, June, December, August and February.
Mark Twain (1894)

- January Effect
  Haugen, Robert A., and Josef Lakonishok. The incredible January effect: the stock market's unsolved mystery. Dow Jones-Irwin, 1987.
- Monday Effect
  Wang, Ko, Yuming Li, and John Erickson. "A new look at the Monday effect." The Journal of Finance 52.5 (1997): 2171-2186.
  Mehdian, Seyed, and Mark J. Perry. "The reversal of the Monday effect: new evidence from US equity markets." Journal of Business Finance & Accounting 28.7?8 (2001): 1043-1065.

Calendar Rules in Sullivan, Timmerman, White (1998), 2

Full Universe: 9452 Different Calendar Rules
Reduced Universe: 244 rules

- Day of the week. $\{-1, 0, 1\}^5$ excluding fixed (e.g. 1^5, -1^5, 0^5)
- Week of the Month: $\{-1, 0, 1\}^4$ excluding fixed
- Month of the Year: $\{-1, 0, 1\}^{12}$ excluding fixed
- SemiMonth: $\{-1, 0, 1\}^2$ every month; also $\{-1, 1\}$ in month i
- Holidays: pre-, post-, normal; $\{-1, 0, 1\}$ on pre- $\times$ $\{-1, 0, 1\}$ on post- ; exclude fixed rules
- End of December
- Turn-of-the-month

Hypothesis to be tested

No Calendar Rule better than baseline

$$H_0 : \max_{k=1,\ldots,K} E(f_k) \leq 0$$

Empirical Returns

$$\bar{f}_k = Ave_{1 \leq t \leq T} f_{k,t}$$

Stationary Bootstrap: Romano and Politis 1994

$$V^{*b} = \sqrt{T} \max_{k=1,\ldots,K} \bar{f}_k^{*b}$$

Empirical Sharpe Ratio

$$\bar{SR}_k = \frac{Ave_{1 \leq t \leq T} f_{k,t}}{SD_{1 \leq t \leq T} f_{k,t}}$$

$$U^{*b} = \max_{k=1,\ldots,K} \bar{SR}_k^{*b}$$

Best Calendar Rule Performance, Full Universe

Performance of the Best Calendar Rules under the Mean Return Criterion

This table presents the performance results of the best calendar rule, chosen with respect to the mean return criterion, in each of the sample periods, for the full universe of 9,452 calendar rules. The table reports the annualized mean return for the benchmark model and the best performing model, along with White's Reality Check *P*-value and the nominal *P*-value (*i.e.*, that which results from applying the Reality Check methodology to the best trading rule *only*, thereby ignoring the effects of the data-snooping).

Full Universe Best Model: Mean Return Criterion				
Sample	Benchmark Return	Model Return	Nominal *P*-value	White's *P*-value
Jan 1897 - Dec 1910	4.17	9.40	0.028	0.617
Jan 1911 - Dec 1924	2.43	6.05	0.089	0.739
Jan 1925 - Dec 1938	1.51	13.22	0.065	0.367
Jan 1939 - May 1952	3.42	8.01	0.048	0.481
June 1952 - Dec 1963	9.21	17.38	0.000	0.216
Jan 1964 - Dec 1975	0.93	10.88	0.000	0.241
Jan 1976 - May 1986	7.56	10.61	0.090	0.915
Jan 1897 - May 1986	3.88	8.50	0.000	0.196
Jan 1897 - Dec 1996	4.63	8.66	0.000	0.243

Best Calendar Rules, Full Universe

Table I

Best Calendar Rules under the Mean Return Criterion

This table reports the historically best-performing calendar rule, chosen with respect to the mean return criterion, in each sample period for the full universe of 9,452 calendar rules.

Sample	Full Universe Best Model: Mean Return Criterion
Jan 1897 - Dec 1910	Month of Year -- j, f, m, a, m, j, j, a, s, o, n, d = 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1
Jan 1911 - Dec 1924	Month of Year -- j, f, m, a, m, j, j, a, s, o, n, d = 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1
Jan 1925 - Dec 1938	Day of the Week -- m, t, w, th, f = -1, 0, 0, 0, 0
Jan 1939 - May 1952	Turn of Month -- -4, -3, -2, -1, 1, 2, 3, 4, otherwise = 0, 0, 0, 0, 0, 0, 0, 1, 1
June 1952 - Dec 1963	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1
Jan 1964 - Dec 1975	Day of the Week -- m, t, w, th, f = 0, 0, 1, 1, 1
Jan 1976 - May 1986	Month of Year -- j, f, m, a, m, j, j, a, s, o, n, d = 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1
Jan 1897 - May 1986	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1
Jan 1897 - Dec 1996	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1

Best Calendar Rule Performance, Reduced Universe

Table VI

Performance of the Best Calendar Rules, in the Reduced Universe, under the Mean Return Criterion

This table presents the performance results of the best calendar rule, chosen with respect to the mean return criterion, in each of the sample periods, for the reduced universe of 244 calendar rules. The table reports the annualized mean return for the benchmark model and the best performing model, along with White's Reality Check *P*-value and the nominal *P*-value (*i.e.*, that which results from applying the Reality Check methodology to the best trading rule *only*, thereby ignoring the effects of the data-snooping).

	Reduced Universe Best Model: Mean Return Criterion			
Sample	Benchmark Return	Model Return	Nominal *P*-value	White's *P*-value
Jan 1897 - Dec 1910	4.17	7.60	0.000	0.553
Jan 1911 - Dec 1924	2.43	4.86	0.115	0.687
Jan 1925 - Dec 1938	1.51	13.22	0.065	0.270
Jan 1939 - May 1952	3.42	7.60	0.062	0.349
June 1952 - Dec 1963	9.21	17.38	0.000	0.170
Jan 1964 - Dec 1975	0.93	9.79	0.000	0.211
Jan 1976 - May 1986	7.56	9.32	0.048	0.906
Jan 1897 - May 1986	3.88	8.50	0.000	0.119
Jan 1897 - Dec 1996	4.63	8.66	0.000	0.167

Best Calendar Rules, Reduced Universe

Best Calendar Rules, in the Reduced Universe, under the Mean Return Criterion

This table reports the historically best-performing calendar rule, chosen with respect to the mean return criterion, in each sample period for the reduced universe of 244 calendar rules.

Sample	Reduced Universe Best Model: Mean Return Criterion
Jan 1897 - Dec 1910	Month of Year -- j, f, m, a, m, j, j, a, s, o, n, d = 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1
Jan 1911 - Dec 1924	Month of Year -- j, f, m, a, m, j, j, a, s, o, n, d = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1
Jan 1925 - Dec 1938	Day of the Week -- m, t, w, th, f = -1, 0, 0, 0, 0
Jan 1939 - May 1952	Turn of Month -- -4, -3, -2, -1, 1, 2, 3, 4, otherwise = 1, 1, 1, 1, 1, 1, 1, 1, 0
June 1952 - Dec 1963	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1
Jan 1964 - Dec 1975	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1
Jan 1976 - May 1986	Semi Month -- second half of October, otherwise = 0, 1
Jan 1897 - May 1986	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1
Jan 1897 - Dec 1996	Day of the Week -- m, t, w, th, f = 0, 1, 1, 1, 1

Sullivan, Timmerman, White (1999)

THE JOURNAL OF FINANCE • VOL. LIV, NO. 5 • OCTOBER 1999

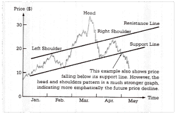
Data-Snooping, Technical Trading Rule Performance, and the Bootstrap

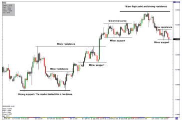
RYAN SULLIVAN, ALLAN TIMMERMANN,
and HALBERT WHITE*

ABSTRACT

In this paper we utilize White's Reality Check bootstrap methodology (White (1999)) to evaluate simple technical trading rules while quantifying the data-snooping bias and fully adjusting for its effect in the context of the full universe from which the trading rules were drawn. Hence, for the first time, the paper presents a comprehensive test of performance across all technical trading rules examined. We consider the study of Brock, Lakonishok, and LeBaron (1992), expand their universe of 26 trading rules, apply the rules to 100 years of daily data on the Dow Jones Industrial Average, and determine the effects of data-snooping.

Technical Analysis





Technical Rules in Sullivan, Timmerman, White (1999), A

A. Filter Rules

Filter rules are used in Alexander (1961) to assess the efficiency of stock price movements. Fama and Blume (1966) explain the standard filter rule:

> An x per cent filter is defined as follows: If the daily closing price of a particular security moves up at least x per cent, buy and hold the security until its price moves down at least x per cent from a subsequent high, at which time simultaneously sell and go short. The short position is maintained until the daily closing price rises at least x per cent above a subsequent low at which time one covers and buys. Moves less than x per cent in either direction are ignored. (p. 227)

Technical Rules in Sullivan, Timmerman, White (1999), B1

B. Moving Averages

Moving average cross-over rules, highlighted in BLL, are among the most popular and common trading rules discussed in the technical analysis literature. The standard moving average rule, which utilizes the price line and the moving average of price, generates signals as explained in Gartley (1935):

> In an uptrend, long commitments are retained as long as the price trend remains above the moving average. Thus, when the price trend reaches a top, and turns downward, the downside penetration of the moving average is regarded as a sell signal. . . . Similarly, in a downtrend, short positions are held as long as the price trend remains below the moving average. Thus, when the price trend reaches a bottom, and turns upward, the upside penetration of the moving average is regarded as a buy signal. (p. 256)

Technical Rules in Sullivan, Timmerman, White (1999), B2

There are numerous variations and modifications of this rule. We examine several of these. For example, more than one moving average (MA) can be used to generate trading signals. Buy and sell signals can be generated by crossovers of a slow moving average by a fast moving average, where a slow MA is calculated over a greater number of days than the fast MA.[10]

There are two types of "filters" we impose on the moving average rules. The filters are said to assist in filtering out false trading signals (i.e., those signals that would result in losses). The fixed percentage band filter requires the buy or sell signal to exceed the moving average by a fixed multiplicative amount, b. The time delay filter requires the buy or sell signal to remain valid for a prespecified number of days, d, before action is taken. Note that only one filter is imposed at a given time. Once again, we consider holding a given long or short position for a prespecified number of days, c.

C. Support and Resistance

The notion of support and resistance is discussed as early as in Wyckoff (1910) and is tested in BLL under the title of "trading range break." A simple trading rule based on the notion of support and resistance (S&R) is to buy

when the closing price exceeds the maximum price over the previous n days, and sell when the closing price is less than the minimum price over the previous n days. Rather than base the rules on the maximum (minimum) over a prespecified range of days, the S&R trading rules can also be based on an alternate definition of local extrema. That is, define a minimum (maximum) to be the most recent closing price that is less (greater) than the e previous closing prices. As with the moving average rules, a fixed percentage band filter, b, and a time delay filter, d, can be included. Also, positions can be held for a prespecified number of days, c.

D. Channel Breakouts

A channel (sometimes referred to as a trading range) can be said to occur when the high over the previous n days is within x percent of the low over the previous n days, not including the current price. Channels have their origin in the "line" of Dow Theory which was set forth by Charles Dow around the turn of the century.[11] The rules we develop for testing the channel breakout are to buy when the closing price exceeds the channel, and to sell when the price moves below the channel. Long and short positions are held for a fixed number of days, c. Additionally, a fixed percentage band, b, can be applied to the channel as a filter.

Technical Rules in Sullivan, Timmerman, White (1999), E

E. On-Balance Volume Averages

Technical analysts often rely on volume of transactions data to assist in their market-timing efforts. Although volume is generally used as a secondary tool, we include a volume-based indicator trading rule in our universe of rules. The on-balance volume (OBV) indicator, popularized in Granville (1963), is calculated by keeping a running total of the indicator each day and adding the entire amount of daily volume when the closing price increases, and subtracting the daily volume when the closing price decreases. We then apply a moving average of n days to the OBV indicator, as suggested in Gartley (1935). The OBV trading rules employed are the same as for the moving average trading rules, except in this case the value of interest is the OBV indicator rather than price.

Eigenanalysis of Technical Rules

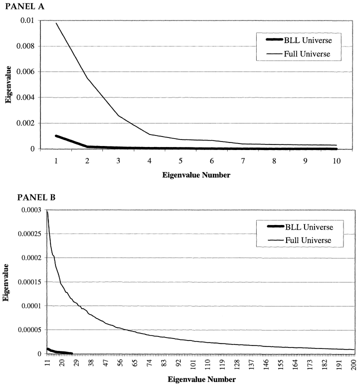


Figure 1. Span of the Brock, Lakonishok, and LeBaron (1992) universe of trading rules versus the full universe of trading rules: Eigenvalues 1 to 200 of the covariance matrix of returns. The eigenvalues of the covariance matrix of returns are sorted in descending order for the Brock, Lakonishok, and LeBaron (BLL) universe of trading rules (i.e., a 26 × 26 matrix), and for 500 randomly chosen rules from the full universe of trading rules (i.e., a 500 × 500 covariance matrix), including the 26 BLL rules. Panel A plots the 10 largest values in sorted descending order along the x-axis, where the y-axis measures the eigenvalue. Panel B plots eigenvalues 11 to 200, again sorted in descending order.

Hypothesis to be tested

No Technical Rule better than baseline

$$H_0 : \max_{k=1,\ldots,K} E(f_k) \leq 0$$

Empirical Returns

$$\bar{f}_k = Ave_{1 \leq t \leq T} f_{k,t}$$

Stationary Bootstrap: Romano and Politis 1994

$$V^{*b} = \sqrt{T} \max_{k=1,\ldots,K} \bar{f}_k^{*b}$$

Empirical Sharpe Ratio

$$\bar{SR}_k = \frac{Ave_{1 \leq t \leq T} f_{k,t}}{SD_{1 \leq t \leq T} f_{k,t}}$$

$$U^{*b} = \max_{k=1,\ldots,K} \bar{SR}_k^{*b}$$

Best Technical Rule Performance, by Period

Table III

Performance of the Best Technical Trading Rules under the Mean Return Criterion

This table presents the performance results of the best technical trading rule, chosen with respect to the mean return criterion, in each of the sample periods. Results are provided for both the Brock, Lakonishok, and LeBaron (BLL) universe of technical trading rules and our full universe of rules. The table reports the performance measure (i.e., the annualized mean return) along with White's Reality Check *p*-value and the nominal *p*-value. The nominal *p*-value results from applying the Reality Check methodology to the best trading rule *only*, thereby ignoring the effects of the data-snooping.

	BLL Universe of Trading Rules			Full Universe of Trading Rules		
Sample	Mean Return	White's *p*-Value	Nominal *p*-Value	Mean Return	White's *p*-Value	Nominal *p*-Value
In-sample						
Subperiod 1 (1897–1914)	9.52	0.021	0.000	16.48	0.000	0.000
Subperiod 2 (1915–1938)	13.90	0.000	0.000	20.12	0.000	0.000
Subperiod 3 (1939–1962)	9.46	0.000	0.000	25.51	0.000	0.000
Subperiod 4 (1962–1986)	7.87	0.004	0.000	23.82	0.000	0.000
90 years (1897–1986)	10.11	0.000	0.000	18.65	0.000	0.000
100 years (1897–1996)	9.39	0.000	0.000	17.17	0.000	0.000
Out-of-sample						
Subperiod 5 (1987–1996)	8.63	0.154	0.055	14.41	0.341	0.004
S&P 500 Futures (1984–1996)	4.25	0.421	0.204	9.43	0.908	0.042

Best Technical Rules, by Period

Table I

Best Technical Trading Rules under the Mean Return Criterion

This table reports the historically best-performing trading rule, chosen with respect to the mean return criterion, in each sample period for both of the trading rule universes: the Brock, Lakonishok, and LeBaron (1992) (BLL) universe with 26 rules and our full universe with 7,846 rules.

Sample	BLL Universe of Trading Rules	Full Universe of Trading Rules
In-sample		
Subperiod 1 (1897–1914)	50-day variable moving average, 0.01 band	5-day support & resistance, 0.005 band, 5-day holding period
Subperiod 2 (1915–1938)	50-day variable moving average, 0.01 band	5-day moving average
Subperiod 3 (1939–1962)	50-day variable moving average, 0.01 band	2-day on-balance volume
Subperiod 4 (1962–1986)	150-day variable moving average	2-day on-balance volume
90 years (1897–1986)	50-day variable moving average, 0.01 band	5-day moving average
100 years (1897–1996)	50-day variable moving average, 0.01 band	5-day moving average
Out-of-sample		
Subperiod 5 (1987–1996)	200-day variable moving average, 0.01 band	Filter rule, 0.12 position initiation, 0.10 position liquidation
S&P 500 Futures (1984–1996)	200-day variable moving average	30- and 75-day on-balance volume

Best Technical Rule Performance, 100 years

Table IV

Technical Trading Rule Summary Statistics: 100-Year Dow Jones Industrial Average Sample (1897–1996) with the Mean Return Criterion

This table provides summary statistics, White's Reality Check p-value, and the nominal p-value for the best-performing rule (the simple five-day moving average), chosen with respect to the mean return criterion, and the recursive cumulative wealth rule, over the full 100-year sample of the Dow Jones Industrial Average. The nominal p-value results from applying the Reality Check methodology to the best trading rule *only*, thereby ignoring the effects of the data-snooping. The cumulative wealth trading rule bases today's signal on the best trading rule as of yesterday, according to total accumulated wealth. The recursive cumulative wealth rule is not the best trading rule ex post, thus the Reality Check p-value does not apply.

Summary Statistics	Best Rule	Cumulative Wealth Rule
Annualized average return	17.2%	14.9%
Nominal p-value	0.000	0.000
White's Reality Check p-value	0.000	n/a
Total number of trades	6,310	6,160
Number of winning trades	2,501	2,476
Number of losing trades	3,809	3,684
Average number of days per trade	4.3	4.2
Average return per trade	0.29%	0.26%
Number of long trades	3,155	3,103
Number of long winning trades	1,389	1,372
Number of long losing trades	1,766	1,731
Average number of days per long trade	4.7	4.6
Average return per long trade	0.39%	0.35%
Number of short trades	3,155	3,057
Number of short winning trades	1,112	1,104
Number of short losing trades	2,043	1,953
Average number of days per short trade	3.9	3.8
Average return per short trade	0.19%	0.16%