Data Snooping Bias in Tests of the Relative Performance of Multiple Forecasting Models

Dan Gabriel Anghel[1,2]

[1] *Department of Money and Banking & CEFIMO, The Bucharest University of Economic Studies*
[2] *Institute for Economic Forecasting, Romanian Academy*

This version: 01/07/2020

Abstract. Tests of the relative performance of multiple forecasting models are sensitive to how the set of alternatives is defined. Evaluating one model against a particular set may show that it has superior predictive ability, while changing the number or type of alternatives in the set may show otherwise. This paper focuses on forecasting models based on technical analysis and shows that data snooping bias occurs in tests that restrict the size and diversity of prediction model universes by ignoring alternatives used by investors and other researchers. Overall, the findings command more caution when interpreting positive results regarding the superior predictive ability of forecasting models based on technical analysis.

JEL c1assification: C12, C18, G11, G14, G17

Keywords: Superior Predictive Ability; Data Snooping; False Discoveries; Reality Check; Technical Analysis; Trading Rule; Efficient Market Hypothesis.

1. Introduction

Searching for better forecasting models is the fundamental objective in many research projects with key theoretical, practical applications. Technological advancements and the rapid growth in computing power have significantly increased the number of investigated alternatives. Analyzing the absolute performance of a model may be desirable in some circumstances, but evaluating relative performance is preferred for obvious reasons. How should the relative predictive ability of a new model be tested given existing alternatives? On the one hand, tests should account for the associated multiple hypotheses by controlling an appropriate compound error rate. The Reality Check (RC) test (White, 2000) or the Positive False Discovery Rate (pFDR) test (Storey, 2002) are two examples that do this. On the other hand, tests should account for the data snooping efforts of others (White, 2000). However, considering models that others use is not an established practice in the financial economics literature. This implies that previously published results may still be biased due to data snooping[1]. Surprisingly, an investigation into if and how neglecting relevant alternative models biases test results has yet to be performed. While many papers have discussed methodologies that handle multiple hypotheses, to the extent of our knowledge no paper has analyzed if the choice for the set of prediction models (tested hypotheses) has a role in shaping results.

This paper fills the gap by investigating how choosing an *unrepresentative* set of alternative models, one that does not account for the data snooping efforts of others, influences the outcomes of tests that evaluate their relative performance. It focuses on forecasting models derived from technical analysis, technical trading rules (TTRs), because the number of investigated alternatives is considerable. Also, evaluating the performance of TTRs is often used as a test for the weak-form Efficient Market Hypothesis (EMH) and accounting for

---

[1] Data snooping refers to the practice of searching for better performing forecasting models on the same data sample. This increases the chances of finding and using lucky models that just fit the noise in the data, have little economic significance, and perform poorly in out-of-sample applications.

existing, relevant alternatives is especially important in such applications (see, e.g., the definition of efficient markets proposed by Timmermann and Granger, 2004). For brevity, the occurrence of data snooping is mainly investigated in the context of the RC test defined by White (2000), which is more commonly used in the literature, while the pFDR test (Storey, 2002) and the Superior Predictive Ability (SPA) test (Hansen, 2005) are discussed in complementary robustness analyses.

The paper contributes to the literature in several important ways. First, it investigates if the sets of models that are typically used by researchers are representative. Second, it defines and performs a Monte Carlo simulation that shows how employing small, unrepresentative sets in seemingly data snooping-free statistical tests can still generate false discoveries. Third, it performs an extensive empirical investigation into the potential impact of this particular type of data snooping bias on test results and conclusions reported in the literature. Overall, we find that prediction model sets ("universes") typically defined and used in the literature are unrepresentative for what investors and other researchers use. Also, not accounting for the data snooping efforts of others biases test outcomes in favor of falsely showing that some TTRs have statistically significant predictive ability. More generally, the results show that controlling for the data snooping efforts of others is very important for obtaining results that can withstand the test of time. Also, they hint that evaluating relative performance becomes problematic when relevant alternatives are not completely observable, such as in the case of tests that evaluate forecasting models derived from technical analysis. In this and other similar circumstances, new testing methodologies that are robust to the choice for the set of tested prediction models–null hypothesis–should be developed and used.

This paper is inspired by and contributes to the recent discussion centered on the impact of test misspecification and cherry-picking results on the robustness of reported findings and associated inferences. The devastating effects of data snooping are discussed in studies such as

Hou et al. (2017), who find that 85% of the 447 investigated "anomalous" variables in asset pricing models are unable to explain stock returns and 93% of the remaining ones do not survive a more stringent $t \geq 3$ threshold. More generally, Kim and Ji (2015) find that the results reported in many surveyed papers become questionable after revised standards for evidence are used instead. They also observe strong evidence of publication bias in favor of statistically significant results. Also, Chang and Li (2017) are unable to replicate the results in more than half of published papers in top economics journals. Harvey (2017) and Harvey and Liu (2014) discuss the importance of increasing the statistical significance threshold and controlling for data snooping in tests that use widely examined data. Harvey (2017) explicitly warns that "*with the combination of unreported tests, lack of adjustment for multiple tests, and direct as well as indirect p-hacking, many of the results being published will fail to hold up in the future.*"

Because of its specific focus, this paper also contributes to the literature examining the excess performance (returns) obtained by TTRs in financial markets, which is one of the most exposed to the risk of data snooping. Thus, the results have important implications for the literature concerned with the theoretical concept of efficient financial markets (Fama, 1970). There is a widely accepted view that financial prices/returns are not completely random[2]. However, given existing market frictions and other limitations, this does not directly imply that stock markets are not weak-form efficient and that investors are able to earn economic profits (Jensen, 1978; Timmermann and Granger, 2004). Park and Irwin (2007) provide a comprehensive review of the early literature examining TTR excess performance and find that some positive evidence exists. For example, 58 out of 92 "modern" studies conclude in favor of technical trading rules having superior predictive ability. Most results show that TTRs can predict price movements to a certain extent and can earn excess returns over the buy-and-hold

---

[2] See Grossman and Stiglitz (1980) for theoretical arguments and Lim and Brooks (2011) for empirical evidence.

benchmark. However, in some cases, excess returns fade after adjusting for trading costs and risk or turn out to be statistically insignificant when accounting for data snooping.

More recent papers, including many that employ modern RC or FDR tests, report that the excess performance of TTRs in developed stock markets has greatly diminished or even disappeared. Examples include Neuhierl and Schlusche (2010), Bajgrowicz and Scaillet (2012), Shynkevich (2012), or Taylor (2014) for the US market, Ratner and Leal (1999), Fifield et al. (2005), or Marshall and Cahan (2005) for others. These results imply that markets have become more efficient over time. However, conflicting findings continue to appear, such as in Urquhart et al. (2015), who observe that moving average trading strategies based on signal anticipation yield superior profits to investors in the US, UK, and Japanese markets. Also, some authors find that TTRs remain profitable in emerging stock markets, examples including Metghalchi et al. (2009) for Asian markets, Sobreiro et al. (2016) for BRICS and other 6 markets in Central and Latin America, Metghalchi et al. (2012) for emerging European markets, or Al-Nassar (2014) for stock markets in the Middle-East. Moreover, TTRs have recently been found to earn some kind of economic profits in tests that reexamine the foreign exchange market (Coakley et al., 2016; Hsu et al., 2016; Zarrabi et al., 2017), the US bond market (Shynkevich, 2016), or the commodity futures market (Han et al., 2016). Are TTRs truly capable of earning significant excess returns after being extensively used by investors and researchers for so many years? Although it is possible that some markets are not efficient or even adaptive (Lo, 2004), test misspecification, methodological limitations, and publication bias may also play a role in shaping the conclusions in the literature. Based on the results reported in this paper, we argue that using unrepresentative rule universes contributes to creating a biased, more favorable image regarding the excess performance of TTRs in financial markets.

The remainder of the paper is organized as follows. Section 2 analyzes if trading rule universes typically used in the literature are representative and presents an alternative universe that should better account for what investors and other researchers use. Section 3 discusses testing methodologies for the relative performance of multiple forecasting models. Section 4 develops and presents the results of a Monte Carlo exercise that investigates if and how using unrepresentative universes biases results. Section 5 presents an extensive empirical investigation that evaluates the potential impact of this particular type of data snooping on the results reported in the TTR literature. Section 6 concludes.

2. Technical Trading Rule Universes

The performance of TTRs has been (and still is) evaluated in many research articles (e.g., Park and Irwin, 2007). Also, many practitioners use technical analysis to make investment decisions in financial markets (e.g., Taylor and Allen, 1992; Menkhoff, 2010; Scott et al., 2016). Even though there is no indication of the exact number and type of rules that have been considered, stakeholders are known to routinely mine financial price data in search of better forecasting models, this resulting in countless investigated alternatives over the years. Thus, we postulate that the number of distinct prediction models in the real, "true" universe is quite large.

Are universes typically used in the literature representative for the data snooping efforts of others? We answer this question by first defining a trading rule universe that better proxies the "true" set. Trading rules that researchers typically use–such as filters, runs, moving averages, the RSI, and the Rate of Change indicators–are incorporated first. This initial selection is supplemented by trading rules inspired by the practitioner literature. In total, we define and use a set that contains 686,304 distinct rules. This is denoted thereafter as *686k.*

Table 1. Independent subsets of trading rules in 686k

No.	Name (Symbol) of Technical Analysis Method (Indicator)	Indicator Type	Number of trading rules
1	Accumulation Swing Index (ASI)	momentum	210
2	Arms Ease of Movement (EMV)	momentum	840
3	Aroon Oscillator (AO)	standardized momentum	10,507
4	Balance of Market Power (BMP)	standardized momentum	39,207
5	Bollinger Oscillator (%b)	momentum	12,402
6	Center of Gravity Oscillator (COG)	momentum	252
7	Chaikin Money Flow (CMF)	standardized money flow	25,258
8	Chaikin Oscillator (CO)	money flow	6,174
9	Chande Momentum Oscillator (CMO)	standardized momentum	27,969
10	Commodity Channel Index (CCI)	momentum	616
11	Demand Index (DI)	standardized money flow	25,258
12	Detrended Price Oscillator (DPO)	momentum	672
13	Dynamic Momentum Index (DYMOI)	standardized momentum	37,584

14	Filter (F)	momentum	51
15	Inertia Indicator (INI)	standardized momentum	22,464
16	Kase Convergence Divergence (KCD)	momentum	43,141
17	Kase Peak Oscillator (KPO)	momentum	8,624
18	Klinger Volume Oscillator (KVO)	money flow	6,174
19	Know Sure Thing (KST)	momentum	5,488
20	Linear Regression Slope (LRS)	momentum	371
21	Market Volume Impact (MVI)	money flow	252
22	Money Flow Index (MFI)	money flow	24,978
23	Moving Average Convergence Divergence (MACD)	momentum	4,704
24	New Relative Volatility Index (NRVI)	standardized momentum	30,331
25	On Balance Volume (OBV)	money flow	210
26	Plus DM vs. Minus DM crossover (DMI)	standardized momentum	441
27	PI Opinion Oscillator (PI)	standardized momentum	7,107
28	Polarized Fractal Efficiency (PFE)	standardized momentum	60,426
29	Random Walk Index for High prices (RWI)	momentum	450
30	Rate of Change (ROC)	momentum	672
31	Relative Momentum Index (RMI)	standardized momentum	48,600
32	Relative Strength Index (RSI)	standardized momentum	10,864
33	Relative Vigor Index (RVig)	standardized momentum	60,426
34	Relative Volatility Index (RVI)	standardized momentum	16,859
35	Runs Indicator (R)	momentum	11
36	Stochastic Momentum Index (SMI)	standardized momentum	33,250
37	Stochastic Oscillator (%k)	standardized momentum	1,769
38	Stochastic RSI Oscillator (SRSI)	standardized momentum	16,859
39	The Quantitative Candlestick (Qstick)	momentum	840
40	Triple Exponential Smoothing (TRIX)	momentum	3,402
41	True Strength Index (TSI)	standardized momentum	60,426
42	Ultimate Oscillator (UO)	standardized momentum	22,842
43	Vortex Oscillator (VX)	standardized momentum	7,114
44	Williams Variable Accumulation Distribution (WVAD)	money flow	210

The search exercise for popular trading rules may be interesting in its own right and may deserve some additional attention. Table 1 lists the 44 technical analysis indicators used to construct *686k*, while further details are provided in Appendix A from the supplementary materials. There, we discuss the terminology, the search procedure, how the choice was made, the ways in which specific trading rules are constructed, and the qualitative improvements in terms of diversity that this new universe brings over the ones previously used in the literature. To the extent of our knowledge, this is the largest trading rule universe considered so far. As a comparison, Sullivan et al. (1999) use 7,846 TTRs, Zarrabi et al. (2017) use 7,650 TTRs, Neuhierl and Schlusche (2010) use 10,256 TTRs, Shynkevich (2012) uses 12,937 TTRs, Hsu et al. (2016) use 21,000 TTRs, Shynkevich (2016) uses 27,000 TTRs, and Coakley et al. (2016) use 113,148 TTRs. Evidently, even *686k* might still not be representative[3], especially because

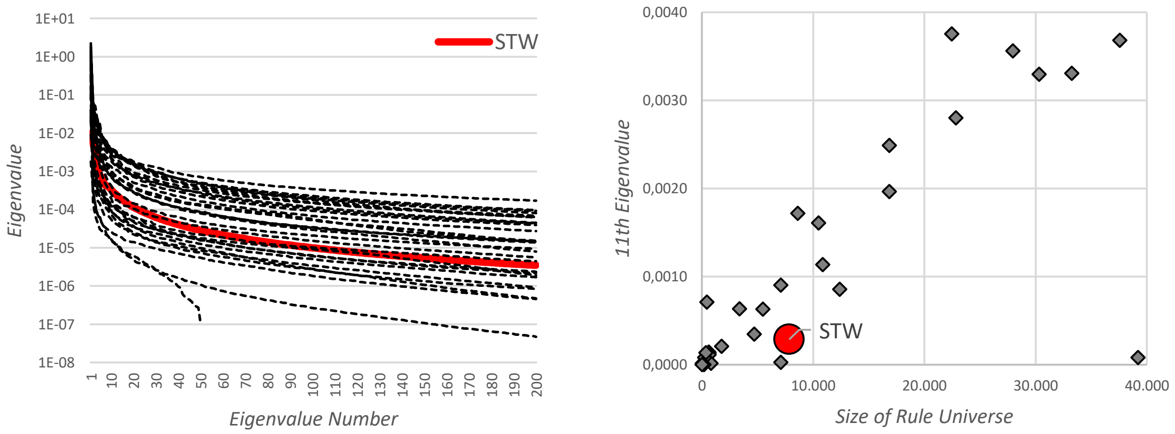
---

[3] Some authors use TTRs derived from artificial intelligence and computer optimization algorithms (e.g., Brabazon el al., 2012). Also, hedge funds and other skilled investors may additionally incorporate more sophisticated mathematical rules, or may use combinations of rules from different areas, such as fundamental analysis, behavioral finance and so on. In this paper, we implement a conservative approach and disregard such alternatives in order to avoid hindsight bias (Timmermann and Granger, 2004) and also because they require more expertise and generate higher implementation costs, which makes them accessible to only a small fraction of investors.

the “true” universe is impossible to observe. However, it should be a better proxy because TTRs are selected based on what others use and are greatly diversified in terms of both type and parameter combinations.

The representativeness hypothesis for rule universes previously used in the literature can now be tested by comparing their effective span[4] to that of *686k*. For brevity, the universe used by Sullivan et al. (1999)–denoted thereafter as STW–is selected as a benchmark, because of its relatively large size (only a handful of previously tested universes contain more rules) and because data on its span are reported by the authors (Figure 1 in Sullivan et al., 1999, p. 1660). Also, in order to ease computational demand, STW is actually compared with 31 of the 44 independent rule universes contained within *686k*[5]. The test is performed using daily closing prices for the Dow Jones Industrial Average (DJIA) index from 1897 to 1986.

Figure 1. The effective span of trading rule universes



Panel A. Eigenvalues 1 to 200

Panel B. Size vs. 11th eigenvalue

NOTE. Panel A reports the first 200 eigenvalues for the covariance matrix of excess returns of TTRs in 31 of the 44 independent rule universes reported in Table 1, alongside the rule universe used by Sullivan et al. (1999), designated as STW. Panel B plots the 11th eigenvalue relative to the size of the rule universe it was estimated on.

The results are reported in Figure 1 and show that STW is dominated in terms of effective span by 19 of the 31 considered universes, including some that have fewer than 7,846 rules. For example, the universe derived from the MACD indicator has 4,704 rules, yet its span

---

[4] This is defined as the total number of non-zero eigenvalues computed for the covariance matrix of excess returns, constructed using all the prediction models in the universe (Sullivan et al., 1999).

[5] The span is not estimated for 13 universes because either they are very large on their own, or because they require traded volume data that is not available for the DJIA index throughout the considered sample.

is on average 2.08 times higher compared to STW[6]. The universe with the largest span is derived from the DYMOI indicator; it contains 37,584 rules and has a span that is on average 34.06 times higher compared to STW. Moreover, the results show that the span of rule universes is positively correlated with their size, implying that adding practitioner-oriented TTRs to account for the data snooping efforts of others helps extracts more information from financial prices. This happens even when the newly added rules are correlated with existing ones, although not perfectly. The span of *686k* is not estimated but should be larger compared to all of the considered universes, which are contained within.

Overall, we find that omitted trading rules that are considered by investors and other researchers do generate payoffs that increase the span of rule universes, implying that the representativeness assumption fails for rule universes typically used in the literature. Section 4 explores if and how this leads to data snooping and biased test results. In preparation, Section 3 discusses testing methodologies used to evaluate TTR excess performance.

3. Tests of the Relative Performance of Multiple Forecasting Models

3.1. The Reality Check (RC) test

A typical test of the relative performance of multiple forecasting models defines the data sample and the universe of models to be evaluated, measures the relative performance of each model, and evaluates the statistical and economic significance of the results. Such a test is biased when not properly handling the errors arising from the associated multiple hypotheses. In the seminal paper for the field, White (2000) defines the Reality Check (RC) test and solves this issue by controlling for the Family-wise Error Rate, defined as $FWER = \mathbb{P}(V \geq 1)$, where V is the total number of Type I errors. The RC delivers asymptotically valid p-values for evaluating the null hypothesis of no excess performance using an empirical distribution estimated via bootstrap simulation. The RC test procedure can be formally defined as follows.

---

[6] This is computed as the average pairwise ratio of the first 200 eigenvalues. This proxies the relative difference in the effective spans of the two universes, assuming the eigenvalue series decay similarly.

TTRs are constructed from technical analysis indicators (denoted *x*), which are functions $f_x\colon \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ that measure certain characteristics of price movements[7]. A TTR (denoted *k*) is defined using a "signal function" $\delta_{k,t}\colon \mathbb{R}^{p_k} \rightarrow \{0, 1\}$, which is based on the values of one or more indicators and is used to make predictions about the expected direction of price movements over a specified interval, typically set to one observation. The predictions are used by investors to mechanically time the market[8], with the aim of earning economic returns. A trading rule universe is a collection of $K \in \mathbb{N}^*$ technical trading rules. The RC test is used to evaluate the null hypothesis that the best performing TTR in a universe (and, thus, any rule) has no superiority over a benchmark prediction model, i.e. that its average excess return is not significantly positive. The test first defines the loss function associated with each TTR as:

$$L(\zeta_t,\, \delta_{k,t-1}) = -\,\delta_{k,t-1}\zeta_t, \qquad t = \overline{1,T}, \qquad k = \overline{1,K} \tag{1}$$

where ζ_t denotes market log-returns. Using *t–1* for the signal function eliminates contemporaneous trading and controls for the look-ahead bias. Considering the buy-and-hold rule as the benchmark ($\delta_{0,t} = 1, t = \overline{1,T}$) and a sample length of T observations, the excess return series ($d_{k,t}$) and the average excess return ($\bar{d}_k$) for each TTR are:

$$d_{k,t} = L(\zeta_t,\, \delta_{0,t-1}) - L(\zeta_t,\, \delta_{k,t-1}), \qquad t = \overline{1,T}, \qquad k = \overline{1,K} \tag{2}$$

$$\bar{d}_k = \frac{1}{T}\sum_{t=1}^{T} d_{k,t}\,, \qquad k = \overline{1,K} \tag{3}$$

---

[7] For example, the Moving Average Convergence/Divergence (MACD) indicator measures price momentum and is defined as $f_{MACD}\colon \mathbb{R}^{n_{MACD}} \rightarrow \mathbb{R}, f_{MACD,t}(m,n,P) = EMA_t(m,P) - EMA_t(n,P)$, where $EMA_t(m,P)$ denotes an exponential moving average of the price series, with smoothing factor $\frac{2}{m+1}$, computed at time $t = \overline{1,T}$. The MACD takes $n_{MACD} = 3$ parameters, namely the price vector P and the integers *m, n* representing the length of the "lookback window" for the two moving averages. In practical applications, the price series is omitted from the definition and the MACD is considered to have $n_{MACD} = 2$ parameters.

[8] For example, the trading rule "$MACD\_1$" can be defined using the signal function $\delta_{MACD\_1,t} = \mathbb{1}_{\{MACD_t(12,26,P)>0\}}$, where $\mathbb{1}_{\{\cdot\}}$ represents the indicator function. A value of 1 predicts that prices will increase and 0 that they will remain constant or decrease. Thus, this rule instructs the investor to go long when the MACD(12,26) takes positive values and to stay out of the market otherwise. Trading rules can be extended to incorporate short positions (in this case, the signal function can take an additional value, $\delta_{k,t}\colon \mathbb{R}^{p_k} \rightarrow \{-1, 0, 1\}$), a flexible money management strategy that can partially open/close positions ($\delta_{k,t}\colon \mathbb{R}^{p_k} \rightarrow [-1, 1]$), or margin trading ($\delta_{k,t}\colon \mathbb{R}^{p_k} \rightarrow [-L, L]$, where L is the leverage defined as the inverse of the margin requirement).

The test statistic is defined as the maximum average excess return[9] (T_n^{RC}) and is evaluated using an empirical distribution ($T_{b,n}^{RC*}$) estimated via bootstrap simulation with B iterations. The asymptotically valid p-value ($\hat{p}_{RC}$) is directly computed to evaluate the null:

$$T_n^{RC} = \max\left(n^{1/2}\bar{d}_1, \dots, n^{1/2}\bar{d}_K\right), \qquad n = T \tag{4}$$

$$T_{b,n}^{RC*} = \max\left(n^{1/2}\bar{d}_{b,1}^*, \dots, n^{1/2}\bar{d}_{b,K}^*\right), \qquad n = T, \qquad b = 1..B \tag{5}$$

$$\hat{p}_{RC} = \frac{1}{B}\sum\nolimits_{b=1}^{B} \mathbb{1}_{\{T_{b,n}^{RC*} > T_n^{RC}\}}, \qquad n = T \tag{6}$$

In its original specification, the RC test does not account for transaction costs. One way to consider them would be to compute the ex-post break-even cost for the best TTR in the universe and to compare it with market costs (e.g., Metghalchi et al., 2012). However, this approach may bias the test in favor of TTRs that trade frequently and have high cost-free performance, but low cost-adjusted performance. To correct for this, we substitute Eq. (1) with an adjusted specification that directly incorporates trading costs into the loss function:

$$L\left(\zeta_t, \delta_{k,t-1}\right) = c_{k,t} - \delta_{k,t-1}\zeta_t, \qquad t = \overline{1,T}, \qquad k = \overline{1,K} \tag{7}$$

Eq. (7) also enables incorporating liquidity and price-impact costs, which are often overlooked, even though they can potentially bias results in favor of showing TTR excess performance more often. Here, the trading cost incurred by rule *k* at time *t* is defined as:

$$c_{k,t} = \mathbb{1}_{\{\delta_{k,t-1} \neq \delta_{k,t-2}\}}(0.5\% + l_t), \qquad t = \overline{1,T}, \qquad k = \overline{1,K} \tag{8}$$

This cost is positive when a trade is executed (if the value of the signal function changes) and zero otherwise. When a trade occurs, a fixed broker fee of 0.5% is added to the liquidity cost (l_t), which is defined based on the daily price range:

$$l_t = \begin{cases} \ln\left(\dfrac{H_t}{C_t}\right), \delta_{k,t-1} > 0 \\ \ln\left(\dfrac{C_t}{L_t}\right), \delta_{k,t-1} = 0 \end{cases}, \qquad t = \overline{1,T}, \qquad k = \overline{1,K} \tag{9}$$

---

[9] Even though we recognize that this definition of excess performance can be a bit too narrow, we decide to implement it for reasons mainly related to the comparability of results to previous work. Future research can extend this investigation to alternative specifications.

where H_t, L_t and C_t are the high, low, and close prices, respectively. Adjusting for liquidity costs in this way is equivalent to simulating trading at the least favorable daily prices: buy trades are executed at the maximum and sell trades are executed at the minimum. This should actually overestimate the liquidity cost incurred by traders, but should additionally account for the price impact cost, which is important especially in thin traded markets.

3.2.Alternative tests for the Relative Performance of Multiple Forecasting Models

Several alternatives to the RC testing framework exist. On the one hand, Hansen (2005), Romano and Wolf (2005), or Hsu et al. (2010) propose modified specifications that improve power or identify all overperforming models, while keeping FWER as the controlled compound error measure. In this paper, we use the Superior Predictive Ability (SPA) test of Hansen (2005) as a robustness check for the results obtained in the empirical analysis reported in Section 5. The main change in the SPA test is the use of a studentized test statistic:

$$T_n^{SPA} = \max\left\{\max_{k=1\ldots m} \frac{n^{1/2}\bar{d}_k}{\widehat{\omega}_k}, 0\right\} \tag{10}$$

where $\widehat{\omega}_k^2$ is a consistent estimator of $\omega_k^2 = VAR(n^{1/2}\bar{d}_k)$. This adjustment essentially changes the way TTR performance is measured by replacing the simple excess return in the RC test with a scaled ratio of the excess return to its estimated variance. Additionally, the SPA test uses the law of the iterated logarithm to eliminate poorly performing models from the analysis (see Hansen, 2005, for more details).

On the other hand, Benjamini and Hochberg (1995), Storey (2002), or Barras et al. (2010) develop alternative tests of relative performance. These are based on Bonferroni bounds and control for the less stringent False Discovery Rate, $FDR = \mathbb{E}(V/R \mid R > 0)\mathbb{P}(R > 1)$, where R is the total number of rejected hypotheses. In this paper, we use the Positive False Discovery Rate (pFDR) test of Storey (2002) as a robustness check for the results obtained in the Monte Carlo exercise reported in Section 4. The test is implemented by fixing a rejection region $[0, \gamma]$, where $\gamma \geq 0$ is the significance level, and using the ordered p-values ($\hat{p}_k \leq \hat{p}_{k+1}$,

$k = \overline{1, K-1}$) obtained from independently testing the K null hypotheses to estimate $pFDR = \mathbb{E}(V/R \mid R > 0)$ using:

$$\widehat{pFDR}_{\lambda}(\gamma) = \frac{\gamma \sum_{k=1}^{K} \mathbb{1}_{\{p_k > \lambda\}}}{(1-\lambda)\max\{1, \sum_{k=1}^{K} \mathbb{1}_{\{p_k < \gamma\}}\}[1-(1-\lambda)^K]} \qquad (11)$$

where λ is selected to minimize the mean squared error of the estimates via bootstrap simulation. Then, inferences are made based on the q-value (q_k), which measures the strength of an observed statistic with respect to pFDR and is estimated recursively by:

$$\begin{cases} \hat{q}_k = \widehat{pFDR}(\hat{p}_k), & k = K \\ \hat{q}_k = \min\{\widehat{pFDR}(\hat{p}_k), \hat{q}_{k+1}\}, & k = \overline{1, K-1} \end{cases} \qquad (12)$$

4. Unrepresentative universes and data snooping bias: a Monte Carlo simulation

Tests for the relative performance of multiple forecasting models seemingly eliminate data snooping by handling for the associated multiple hypotheses. However, because they require fixing the set of prediction models beforehand, they implicitly assume that the model universe is representative for the data snooping efforts of others. This might not always be the case and opens up the possibility of intentional or unintentional human error. In this section, we explicitly test the hypothesis that restricting the size and diversity of trading rule universes increases the number of false discoveries, i.e. that data snooping bias occurs when the relative performance of forecasting models is evaluated using unrepresentative universes. Two distinct Monte Carlo exercises are performed, one using the RC test and the other using the pFDR test for robustness. They both rely on simulated random data (on which TTRs should have no superior predictive ability) and estimate the number of Type I Errors arising in tests performed for trading rule universes of varying sizes, which contain at least one “lucky” rule. The actual bias is measured by the difference in false discovery rates (FDR)[10] between tests that employ smaller, restricted universes and tests that employ a benchmark, which is *686k* in our case.

---

[10] In this context, we use FDR to denote the number of null rejections divided by the total number of tests performed, as opposed to the number of Type I errors divided to the total number of rejections.

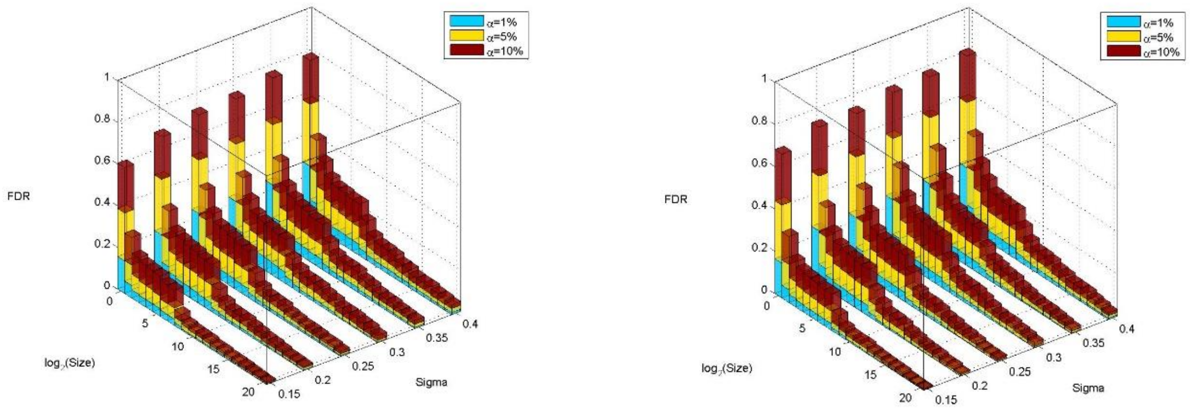
Six random generated data sets are constructed using a discretized zero-drift Geometric Brownian Motion process with volatility parameter $\sigma \in \{0.15; 0.20; 0.25; 0.30; 0.35; 0.40\}$. Each set contains 4,000 calendar years of price and volume data, with each year consisting of roughly $n = 260$ observations (days). Starting with an initial fixed price of $C_0 = 1{,}000$, the next day's closing price is $C_t = C_{t-1}e^{\sigma\epsilon_t\sqrt{\tau}}$, the daily price range is $R_t = H_t - L_t = \sigma C_t \epsilon'_t \sqrt{\tau}$, the high (maximum) price is $H_t = C_t + u_t R_t$, the low (minimum) price is $L_t = C_t - (1 - u_t)R_t$, and the opening price is $O_t = L_t + u'_t R_t$; where ϵ_t and ϵ'_t are independently drawn from a standard normal distribution, u_t and u'_t are independently drawn from a standard uniform distribution and $\tau = n^{-1}$. The daily volume is $V_t = ce^{x_t}$, where x_t is independently drawn from a standard normal distribution and $c = 1{,}000$ is a fixed scale factor.

4.1.Testing procedure and results for the RC test

In the case of the RC methodology, the simulation proceeds in two stages. In the first, a single-rule universe is constructed using the "luckiest" rule in *686k* (the one that generates the highest excess return relative to the buy-and-hold benchmark rule) and its performance is evaluated using the RC test at standard significance levels of 1%, 5%, and 10%. The distribution of the test statistic is estimated using the stationary bootstrap of Politis and Romano (1994), by resampling random blocks with average length $q = 1/\sqrt[4]{n}$ (this is based on the recommendation of Hall et al., 1995) directly from the excess return series. Resampling blocks of data accounts for the autocorrelation in market returns and it is not particularly useful in this exercise; however, it is useful for the empirical investigation in Section 5 and it's also employed here for consistency. The number of bootstrap iterations is set to $B = 1000$.

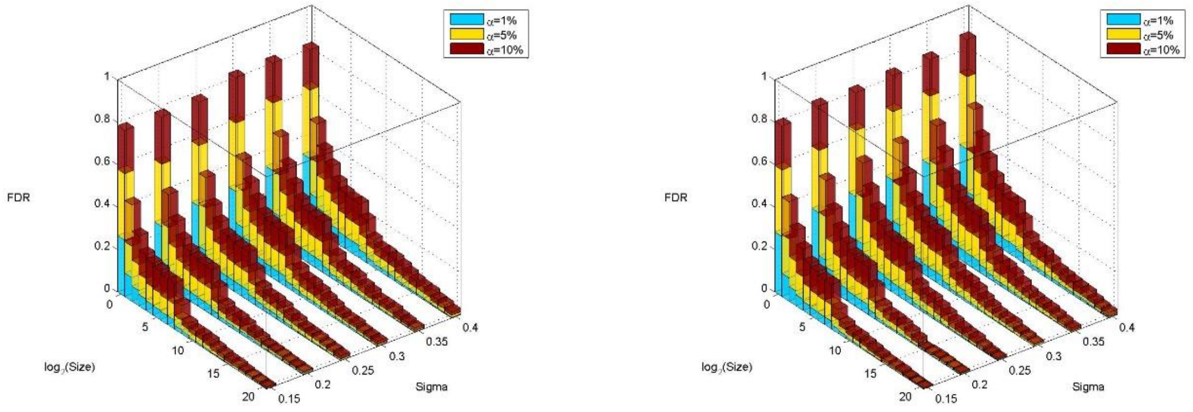
In the second stage, consecutively larger rule universes are constructed and tested on the same sample by adding TTRs to the rule universe until *686k* is tested. New TTRs are added in the order they are listed in Table A4.2 in the supplementary materials. With each additional rule, the distribution of the RC test statistic is re-estimated by resampling an additional 100 times[11] from its excess return series, and the null hypothesis is re-evaluated.

Figure 2. False discovery rates (FDR) and the size of the prediction model universes



Sample length: 1 month

Sample length: 1 quarter



Sample length: 1 year

Sample length: 4 years

This procedure assures that adding new TTRs does not change the trading rule that is evaluated in the RC test, which is always the “luckiest” one in each sample. Instead, the characteristics of the RC distribution used to evaluate its excess performance changes, potentially influencing the result. Because of the random nature of the data, all RC null hypotheses are true and any null rejection constitutes a Type I error, false discovery. To also evaluate potential differences in data snooping bias when the sample length varies, we perform

---

[11] The number of bootstrap iterations is restricted to 100 in the second stage to reduce computational demand, which would otherwise be very large. In a preliminary analysis, 1000 simulations are used to verify the robustness of this choice. The results show that the test outcomes do not materially change.

distinct RC tests on non-overlapping subsamples of *1 month, 1 quarter, 1 year* and *4 years*. In total, 4000 independent tests are performed for each rule universe on each data set[12].

For brevity, we focus on rule universes with a size of $2^m, m = \overline{1,19}$, alongside the benchmark *686k* universe. An overview of estimated false discovery rates is shown in Figure 2 (the size of *686k* is rounded to 2^{20} for illustrative purposes). A pool regression for aggregate false discovery rates is also estimated and the results are reported in Table 2. Detailed results for the largest 11 universes, alongside absolute and relative estimates of the amount of data snooping bias, are shown in Appendix B of the supplementary materials.

Table 2. Regression results for false discovery rates arising in RC tests

This table reports the results of estimating the regression:

$$FDR(\alpha) = \beta_0 + \beta_1 SIZE + \beta_2 VOL + \beta_3 SL + \varepsilon$$

where $FDR(\alpha)$ is the number of false discoveries per 100 tests at the α significance level, $SIZE$ is the base 2 logarithm of the total number of trading rules in the universe, VOL is the volatility of the data generating process, and SL is the sample length expressed in years. In total, 504 observations are included (21 universes, 6 volatility parameters, 4 sample lengths); t-statistics are reported in square parentheses; p-values are reported in round parentheses; ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels respectively.

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
β_0	0.3511	0.2198	0.0898
	[20.60]***	[15.69]***	[11.99]***
β_1	-0.0250	-0.0166	-0.0072
	[-33.95]***	[-27.51]***	[-22.38]***
β_2	0.2169	0.1528	0.0714
	[4.18]***	[3.58]***	[3.13]***
β_3	0.0071	0.0057	0.0025
	[2.56]**	[2.48]**	[2.03]**
Adjusted R^2	*0.7000*	*0.6057*	*0.5045*
F-statistic	*392.34*	*258.64*	*171.71*
	(0.0000)	*(0.0000)*	*(0.0000)*

Several interesting findings are worth noting. First, the number of false discoveries significantly increases when the size of trading rule universes decreases, irrespective of test significance level, sample length, or volatility. This implies that data snooping bias does occur in RC tests when rule universes are small and unrepresentative. The regression reported in Table 2 shows that halving the number of rules in a universe increases Type I errors on average by 1.66 percentage points (pp.) at the 5% significance level (2.50 pp. and 0.72 pp., respectively,

---

[12] 1000 tests are performed for each sample length. When the length is 4 years, all 4,000 years of simulated data are used. When the length is smaller, only the first 1000 periods of that type are used. For example, when the sample length is 1 quarter, only the first 1,000 quarters are used.

when the 10% and 1% significance levels are used instead). It also shows that the size of the rule universe is the most significant determinant of false discoveries. Data snooping bias becomes especially large for universes that contain less than 2^{11} trading rules, but remains significant even for the other restricted universes. Specifically, the average difference in false discoveries between the 10 largest restricted universes and the *686k* benchmark is 0.39 to 2.62 pp. (180% to 975% in relative terms), depending on testing conditions.

Second, the number of false discoveries also increases with volatility, irrespective of the size of the selected universe, test significance level, or sample length. This implies that RC tests perform worse in more volatile markets. For example, when $\sigma = 15\%$, average false discovery rates for the 11 largest universes are 2.69%, 1.28%, and 0.32%, at the 10%, 5%, and 1% levels, respectively. However, when $\sigma = 40\%$ the same averages are 7.08%, 3.82%, and 1.23%. The regression reported in Table 2 shows that an increase of 1 pp. in volatility increases Type I errors on average by 0.21 pp. at the 5% significance level (0.15 pp. and 0.07 pp., respectively, when the 10% and 1% significance levels are used instead). Interestingly, as volatility increases, the amount of data snooping bias caused by varying the size of trading rule universes increases in absolute terms but seems to decrease in relative terms. For example, the average difference in false discovery rates between the 10 largest restricted universes and the *686k* benchmark is between 0.39 and 2.49 pp. (423% to 975% in relative terms) when $\sigma = 15\%$ and between 0.57 and 2.62 pp. (180% to 469% in relative terms) when $\sigma = 40\%$.

Third, the number of false discoveries also increases with the length of the data sample, irrespective of the size of the selected universe, test significance level, or volatility. This implies that RC tests perform worse also when larger samples are used. The regression reported in Table 2 shows that increasing the sample length by one year (roughly 260 daily observations) increases Type I errors on average by 0.71 pp. at the 5% significance level (0.57 pp. and 0.25 pp., respectively, when the 10% and 1% significance levels are used instead). In this case, the data snooping bias caused by varying the size of rule universes increases with the length of the data sample in both absolute and relative terms. For example, the average difference in false discovery rates between the 10 largest restricted universes and the *686k* benchmark is between 0.39 and 2.10 pp. (180% to 452% in relative terms) when a *1 month* length is used but increases to between 0.63 and 2.62 pp. (220% to 975% in relative terms) when the length is *4 years*.

Overall, we find that using unrepresentative universes significantly overstates the excess performance of TTRs. This highlights the need to exercise more caution when interpreting existing evidence in favor of the economic relevance of TTRs. Even though previous findings are not invalidated, we argue that additional tests are required to evaluate their robustness when accounting for the data snooping efforts of others. More generally, the results show that the way researchers choose the size and diversity of prediction model universes has a significant impact on the outcomes of tests that examine the relative performance of multiple forecasting models.

We also find that the potential of RC tests to eliminate data snooping bias negatively correlates with the volatility of the data generating process and the length of the data sample. This is a novel, surprising result that has two important implications. On the one hand, TTRs are able to fit more of the noise in the data and, thus, are "luckier" in extended samples, suggesting that any examination of TTR excess performance should be accompanied by robustness tests performed on shorter time intervals. On the other hand, TTRs are also "luckier" in more volatile markets, suggesting that testing superior predictive ability in such conditions is exposed to additional data snooping risks. As a result, even more caution should be exercised when analyzing evidence in favor of the economical relevance of TTRs in markets associated with high uncertainty, such as small-cap sector stocks (e.g., Shynkevich, 2012), emerging stock markets (e.g., Metghalchi et al., 2012), emerging market currencies (e.g., Hsu et al., 2016), or markets in which prices experience persistent declines (bear markets). Moreover, the concerns can be extended to other research areas, such as “anomalous” asset pricing factors based on technical analysis, derived using portfolios sorted by volatility (e.g. Han et al., 2013). These and other similar findings should be ideally evaluated for robustness when controlling for the correlation between market volatility and data snooping bias.

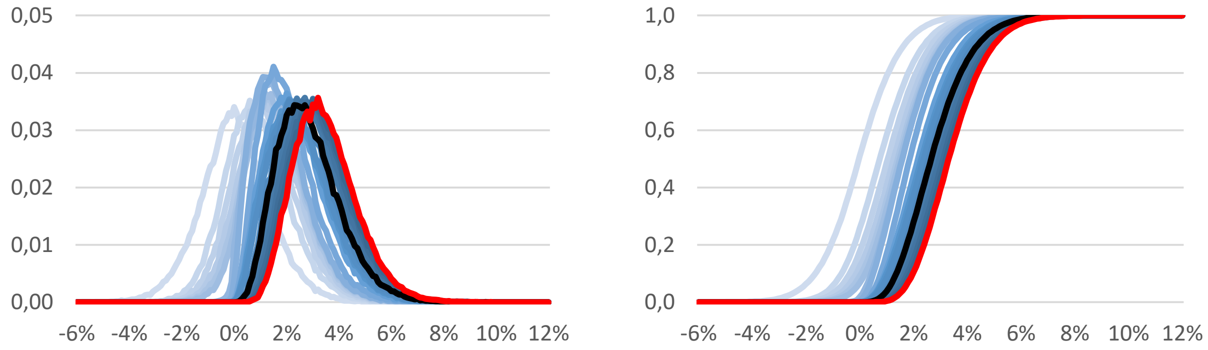
Finally, the simulation results show that false discoveries are not eliminated when the extended *686k* rule universe is used in tests, even though they are significantly reduced compared to all other restricted universes. This finding raises some questions for researchers implementing existing tests of the relative performance of multiple forecasting models. *How does the “true”, representative universe look like? What happens if more prediction models are added? When should one stop adding models?* Situations in which the full set of alternatives used by others is very difficult, if not impossible, to observe, such as in the case of prediction models based on technical analysis, do not allow satisfactory answers to these questions. This exposes the associated statistical tests to ambiguity risk and decreases their scientific relevance. Ultimately, it makes testing for relative model performance subjective and problematic. Because of this, new testing methodologies that are robust to the choice of the prediction model universe should be developed and implemented.

4.2.An illustration of how data snooping bias occurs

To show how data snooping bias arises in tests of the relative performance of multiple forecasting models, we focus on the 1000 simulation results obtained when $\sigma = 0.30$ and the sample length is *1 year*. Figure 3 displays the empirical distribution of the RC test statistic estimated for rule universes of various sizes. For a selection of these, Table 3 provides descriptive statistics, the average p-value for the associated RC test, and the estimated proportion of false discoveries. On the one hand, the results show that the distribution used to evaluate the RC null hypothesis moves to the right as the size of the TTR universe increases, which suggests that newly added rules are not perfectly correlated to existing ones (Arellano-

Valle and Genton, 2008; Hartigan, 2014). On the other hand, the distribution derived from *686k* dominates the other alternatives at every quantile, this showing that it has a significantly higher effective span. Put differently, we find that the distributions shift to the left when TTRs are removed from the benchmark universe, which implies that restricting the size of rule universes by not considering the data snooping efforts of others generates false discoveries and biases test results by decreasing the effective span (informativeness) of tested universes and downward biasing critical values used to evaluate null hypothesis.

Figure 3. Size of rule universes and the empirical distribution of the RC test statistic



Panel A. Probability density

Panel B. Cumulative density

Note. This figure shows the empirical distribution of the test statistic (maximum distributions of excess returns) estimated in RC tests that use consecutive larger rule universes of size $2^m, m = \overline{1,19}$*. The large 686k universe is distinctively depicted using a red line. An intermediate* 2^{13} *universe, comparable in size to what researchers typically use, is depicted using a black line.*

Table 3. The empirical distribution of the test statistic and outcomes of RC tests

Panel A: Characteristics of the maximum distribution of excess returns estimated via bootstrap simulation

$\log_2$(Size)	0	2	4	6	8	10	12	14	16	18	19.39
Average	0.0000	0.0111	0.0150	0.0171	0.0201	0.0243	0.0251	0.0280	0.0297	0.0322	0.0343
Std. Dev.	0.0133	0.0122	0.0118	0.0107	0.0105	0.0119	0.0119	0.0120	0.0118	0.0119	0.0119
Skewness	0.0723	0.3361	0.3736	0.7294	0.8194	0.6509	0.6459	0.6724	0.6365	0.5931	0.5707
Excess Kurtosis	0.4059	0.3360	0.4097	0.6126	0.8198	0.4206	0.4297	0.5201	0.5010	0.4181	0.4052

Panel B: Outcomes of associated RC tests

$\log_2$(Size)	0	2	4	6	8	10	12	14	16	18	19.39
Average p-value	0.0527	0.1816	0.2607	0.2928	0.3602	0.4688	0.4882	0.5580	0.6040	0.6660	0.7143
	[27.99]	[35.32]	[38.38]	[38.88]	[42.17]	[52.33]	[54.03]	[60.21]	[64.76]	[73.52]	[81.22]
FDR(1%)	32.6%	10.7%	6.1%	5.7%	4.2%	2.1%	2.0%	1.4%	1.0%	0.5%	0.2%
FDR(5%)	65.3%	26.2%	18.6%	16.7%	11.4%	6.1%	5.7%	4.2%	3.4%	2.4%	1.6%
FDR(10%)	83.6%	42.8%	29.6%	26.3%	20.8%	11.3%	10.3%	7.1%	6.0%	4.4%	3.2%

Note. Results based on 1000 tests on simulated random data, σ=0.30, sample size=1 year. FDR denotes the proportion of Type I errors (false discoveries). Square parentheses are used to report t-statistics.

4.3.Testing procedure and results for the pFDR test

We perform a complementary Monte Carlo simulation based on the pFDR test of Storey (2002). This can be considered as a robustness check for the results reported so far. For brevity, only a significance level of $\gamma = 5\%$ and a sample length of *1 year* are considered. For each sample, the average cost-adjusted excess return ($\bar{d}_k$) of all trading rules in the *686k* universe is computed. Then, the K null hypotheses $H_k^0: \bar{d}_k \leq 0$ are evaluated using an empirical distribution estimated via a bootstrap procedure similar to the RC test but performed independently for each rule and with $B = 500$ iterations. Next, for the same rule universes of size $2^m, m = \overline{1,19}$ used before, plus the benchmark *686k* universe, the K p-values are ordered and the associated q-values are computed using Eq. (11) and (12) in Section 3.2. For a given sample and rule universe, we consider the composite null hypothesis $H_0: \bar{d}_k \leq 0 \; for \; all \; k = \overline{1,K}$ rejected in the context of the pFDR test if $q_1 \leq \gamma$. The aggregate proportion of samples for which the null is rejected is reported in Panel A of Table 4 and shows that data snooping bias is much more prevalent in pFDR tests. This estimated amount does not significantly change when varying the size of the trading rule universe (detailed results are available at request) but does seem to be positively correlated with the volatility of the data generating process.

Table 4. The empirical distribution of the test statistic and outcomes of RC tests

Panel A: Average number of samples in which the null was rejected, divided by the total number of samples

	σ=0.15	σ=0.20	σ=0.25	σ=0.30	σ=0.35	0.40
	66.03%	64.39%	67.42%	75.90%	75.90%	72.51%

Panel B: Average RNR aggregated by the size of TTR universes and volatility

$\log_2$(Size)	σ=0.15	σ=0.20	σ=0.25	σ=0.30	σ=0.35	0.40
0	65.80%	64.30%	67.30%	75.90%	75.90%	72.50%
1	38.35%	36.80%	37.10%	40.30%	40.30%	40.30%
2	32.05%	28.20%	26.50%	25.83%	25.80%	25.88%
3	35.84%	31.29%	29.69%	28.38%	28.49%	28.98%
4	38.03%	35.94%	34.90%	33.99%	34.40%	33.20%
5	36.33%	33.78%	36.02%	37.02%	37.24%	36.70%
6	26.08%	23.64%	25.87%	27.46%	27.34%	27.09%
7	23.81%	24.35%	25.01%	25.59%	26.04%	24.47%
8	3.03%	3.73%	4.75%	5.25%	5.34%	7.48%
9	3.08%	3.42%	3.96%	4.86%	4.80%	6.40%
10	1.97%	2.52%	3.04%	3.72%	3.72%	5.58%
11	2.70%	3.23%	3.59%	4.26%	4.21%	6.07%
12	3.52%	4.23%	5.45%	5.90%	5.81%	8.53%
13	2.97%	3.57%	4.06%	4.60%	4.66%	6.35%
14	2.97%	3.58%	4.14%	4.78%	4.78%	6.41%
15	2.80%	3.33%	4.09%	4.72%	4.72%	6.54%
16	2.46%	3.00%	3.73%	4.06%	4.03%	5.60%
17	3.87%	4.37%	5.15%	5.97%	5.80%	7.32%
18	6.04%	6.80%	7.46%	8.11%	8.20%	9.60%
19	5.02%	5.46%	6.04%	6.71%	6.71%	8.01%
19.39	4.80%	5.38%	5.93%	6.68%	6.69%	8.13%

We also count the average number of individual null rejections relative to the total number of tested hypotheses, $RNR = K^{-1} \sum_{k=1}^{K} q_k \leq \gamma$, for each sample and each rule universe. The average *RNRs* are reported in Panel B of Table 4 and show that the size of the rule universe and the volatility of the data generating process do influence the total proportion of relative false discoveries. This resembles the earlier results obtained using the RC test and supports the main conclusion that using prediction model universes that do not account for the data snooping efforts of others falsely overstates their relative performance.

Overall, the results show that the pFDR test is significantly more prone to data snooping bias in an absolute sense, this being rather expected because of the less stringent nature of the error measure used. However, when accepting a certain proportion of false discoveries, we find that the negative effects of using restricted prediction model universes can also be found in pFDR tests, even though they are less significant in terms of data snooping bias. Specifically, the relative number of false discoveries significantly increases when the size of the TTR universes is smaller than 2^8 but is fairly constant otherwise. Finally, data snooping bias also increases in pFDR tests with the volatility of the data generating process. Overall, these findings support our earlier conclusions based on the RC methodology and show that our results are robust to the way TTR excess performance is evaluated.

5. Data snooping bias and TTR excess performance: an empirical investigation

In this section, we evaluate if using small, unrepresentative trading rule universes also biases tests that employ real stock market data. We consider daily price and volume data for individual stocks listed in all markets tracked by Thomson Reuters Eikon on November 14, 2013, which have at least 5 listings. In total, there are 81 markets that serve 88 countries. For each, we select up to 40 companies that are part of the main market index. For indices that contain more companies, only 40 of them are randomly select[13]. For indices that contain less, all companies in the index are selected and the list is supplemented using other listings in the descending order of their market capitalization. This results in a sample of 2,579 stocks, for which all available historical trading price and volume data are retrieved up to November 14, 2013. In total 8,667,038 daily stock-price observations are used. A summary of the data sample is presented in Appendix C in the supplementary materials, while additional details can be made available at request.

The empirical exercise first uses the RC test to evaluate the excess performance of TTRs that are part of the benchmark *686k* rule universe. The tests are performed for all stocks on non-overlapping samples that span one calendar year each. Samples that have less than 65 observations are excluded because of insufficient liquidity. To estimate the data snooping bias associated with restricting the size and diversity of the rule universes, the 44 small, unrepresentative universes formed using individual technical analysis indicators (listed in Table 1) are also tested and the results are compared to the benchmark. These universes are similar to the ones typically employed in the literature, in terms of both size and effective span. Higher null rejection rates in tests that use the restricted universes compared to the ones that use the benchmark would indicate the presence of data snooping bias. Also, the difference would provide an estimate of the amount of false discoveries that is due to data snooping.

5.1.Results for the RC test

In total, 34,887 tests are performed for the *686k* universe and 1,534,970 tests are performed for the rule universes constructed using individual indicators. Table 5 provides a summary of the results. In the case of the *686k* universe, which is presented in Panel A, prediction models derived from technical analysis indicators generate positive cost adjusted

---

[13] The companies are ordered by name and then every *[N/40]* in the list is drawn, where N represents the total number of stocks in the index and *[x]* represents the integer part of *x*.

excess returns in 34,678 tests, which amount to 99.4% of the total. However, when considering statistical significance, the RC null hypothesis is rejected only 227 times (0.65% of the total) at the 10% level, 96 times (0.27% of the total) at the 5% level, and 14 times (0.04% of the total) at the 1% level. The results obtained using the restricted rule universes, which are reported in Panel B, show that the RC null hypothesis is rejected 13,525 times (0.88% of the total) at the 10% level, 5,725 times (0.37% of the total) at the 5% level, and 1,132 times (0.07% of the total) at the 1% level. For all significance levels, the average rate of null rejections is approximately two times higher for restricted universes compared to the benchmark: null rejections are inflated 2 times at the 1% level, 1.81 times at the 5% level, and 1.8 times at the 10% level.

Results grouped by both stock and year are reported in Panel C of Table 5. This enables the analysis of instances when at least one of the 44 tests that use unrepresentative rule universes rejects the RC null hypothesis; it evaluates the scenario in which researchers independently test unrepresentative universes on the same data sample and then make inferences based on a meta-analysis of their results. The null hypothesis is rejected at least once for 773 stock-years (2.21% of the total) at the 10% level, 337 stock-years (0.96% of the total) at the 5% level, and 70 stock-years (0.20% of the total) at the 1% level. This shows that considering the positive results of others without accounting for their data snooping efforts widely increases the number of false discoveries. Specifically, the evidence in favor of economically profitable TTRs is inflated on average 7.93 times at the 1% level, 6.61 times at the 5% level, and 6.16 times at the 10% level.

Results grouped by rule universe are reported in Table 6. For the 44 restricted universes, rejection rates at the 10% significance level vary from a minimum of 0.22% to a maximum of 2.99%, with a median (mean) of 1.19% (1.18%). This is significantly higher compared to the null rejection rate obtained when the benchmark *686k* is used, which is 0.65%. This pattern replicates when analyzing test results at the 5% and 1% levels: both median and average rejection rates are 1.8-2 times higher when the restricted universes are used, even though the rules in the benchmark generate positive excess returns more often[14]. The maximum difference is recorded for the smallest rule universe, which is the one derived from the Runs indicator. In this case, rejection rates are 4.6-7.3 higher, depending on the significance level.

Table 5. Summary statistics of RC test results on real stock market data

Panel A. Aggregate results when using 686k

Statistic	Value	Percent of total
i. Number of valid tests	34,887	100.00%
ii. Number of tests in which best TTR obtained positive excess returns	34.678	99.40%
iii. Number of RC null rejections		
10% confidence level	*227*	*0.65%*
5% confidence level	*96*	*0.27%*
1% confidence level	*14*	*0.04%*
iv. Likelihood of TTRs to repeat positive excess returns[⊥]		
*Conditional on TTR indicator class***	*2284*	*6.54%*
*Conditional on TTR indicator class and trading strategy****	*1572*	*4.50%*
*Conditional on TTR indicator class, trading strategy, and parameters*****	*21*	*0.06%*
v. Likelihood of TTRs to repeat economically significant performance (10% significance)[†]		
*Unconditional**	*5*	*0.01%*
*Conditional on TTR indicator class***	*1*	*0.00%*
*Conditional on TTR indicator class and strategy****	*0*	*0.00%*
*Conditional on TTR indicator class, strategy and parameters*****	*0*	*0.00%*

Panel B. Aggregate results when using 44 small, unrepresentative rule universes

Statistic	Value	Percent of total
i. Number of valid tests	1,534,970	100.00%
ii. Number of tests in which best TTR obtained positive excess returns	1,071,904	69.83%
iii. Number of RC null rejections		
10% confidence level	*18,132*	*1.18%*
5% confidence level	*7,867*	*0.51%*
1% confidence level	*1,242*	*0.08%*
iv. Likelihood of TTRs to repeat best performance[⊥]		
*Conditional on TTR indicator class, trading strategy, and parameters*****	*74,502*	*4.85%*
v. Likelihood of TTRs to repeat economically significant performance (10% significance)[†]		
*Unconditional**	*496*	*0.03%*
*Conditional on TTR indicator class, trading strategy, and parameters*****	*28*	*0.00%*

Panel C. Results for tests using the 44 restricted rule universes, aggregated at the stock-year level

Statistic	Value	Percent of total
i. Number of subsamples	34,887	100.00%
ii. Number of tests in which TTRs obtained positive excess returns	25,878	74.17%
iii. Number of subsamples with at least 1 null rejection (out of 44 independent RC tests)		
10% confidence level	*1,399*	*4.01%*
5% confidence level	*635*	*1.82%*
1% confidence level	*111*	*0.31%*
iv. Likelihood of TTRs to repeat best performance[⊥]		
*Conditional on TTR indicator class***	*6,965*	*19.96%*
*Conditional on TTR indicator class and strategy****	*6,866*	*19.68%*
*Conditional on TTR indicator class, strategy and parameters*****	*1,555*	*4.45%*
v. Likelihood of TTRs to repeat economically significant performance (10% significance)[†]		
*Unconditional**	*70*	*0.20%*
*Conditional on TTR indicator class***	*7*	*0.02%*
*Conditional on TTR indicator class and trading strategy*****	*7*	*0.02%*
*Conditional on TTR indicator class, trading strategy, and parameters*****	*5*	*0.01%*

*NOTE: [⊥]The number of times in which TTRs earn positive excess returns in two consecutive years, expressed as a percent of the total number of stock-years. [†]The number of tests in which the RC null hypothesis of no economic profitability is rejected at the 10% level in two consecutive years, expressed as a percent of the total number of stock-years. *Unconditional–estimated for all TTRs in the rule universe. **Conditional on TTR indicator class–estimated only for TTRs that are derived from the same technical analysis indicator (entry/exit strategy and parameter values can vary). ***Conditional on TTR indicator class and trading strategy–estimated for TTRs that are derived from the same technical analysis indicator and the same entry/exit strategy (parameter values can vary). ****Conditional on TTR indicator class, strategy, and parameters– estimated for identical TTRs (are the same in terms of all aspects, including parameter values).*

---

[14] There are 6 restricted rule universes for which RC null rejection rates are lower compared to 686k. However, these occur for the least profitable indicators, which generate positive excess returns less than half of the time and implies that rejection rates are low not because data snooping bias decreased, but because the indicators are not able to consistently predict price movements. All other results show that null rejections in tests that use restricted rule universes are significantly higher compared to those that use the benchmark.

Table 6. Null Rejection Rates aggregated by trading rule universe

Trading rule universe	**rPR***	**NRR, α=0.10**	**NRR, α=0.05**	**NRR, α=0.01**
Accumulation Swing Index	26.14%	1.72%	0.66%	0.07%
Arms Ease of Movement	24.92%	0.78%	0.35%	0.07%
Aroon Oscillator	88.28%	1.41%	0.65%	0.08%
Balance of Market Power	83.45%	1.20%	0.49%	0.06%
Bollinger Oscillator	76.55%	1.36%	0.57%	0.08%
Center of Gravity Oscillator	44.17%	0.25%	0.08%	0.00%
Chaikin Money Flow	89.51%	1.21%	0.54%	0.08%
Chaikin Oscillator	68.94%	1.03%	0.45%	0.07%
Chande Momentum Oscillator	93.03%	1.29%	0.55%	0.09%
Commodity Channel Index	46.03%	1.00%	0.41%	0.08%
Demand Index	87.24%	1.61%	0.74%	0.10%
Detrended Price Oscillator	81.79%	0.79%	0.35%	0.07%
Dynamic Momentum Index	96.58%	0.80%	0.32%	0.04%
Filter	50.70%	2.04%	0.87%	0.16%
Inertia Indicator	92.64%	1.30%	0.57%	0.08%
Kase Convergence Divergence	97.91%	1.43%	0.62%	0.08%
Kase Peak Oscillator	94.66%	0.98%	0.41%	0.06%
Klinger Volume Oscillator	46.99%	0.22%	0.09%	0.01%
Know Sure Thing	80.78%	1.15%	0.47%	0.06%
Linear Regression Slope	66.59%	1.04%	0.47%	0.06%
Market Volume Impact	36.71%	0.40%	0.17%	0.03%
Money Flow Index	87.85%	1.33%	0.57%	0.08%
Moving Average Convergence Divergence	88.01%	1.02%	0.43%	0.06%
New Relative Volatility Index	90.47%	1.10%	0.49%	0.05%
On Balance Volume	8.85%	0.47%	0.22%	0.04%
Plus DM vs. Minus DM crossover	57.29%	1.63%	0.69%	0.12%
PI Opinion Oscillator	84.45%	1.29%	0.55%	0.08%
Polarized Fractal Efficiency	95.40%	1.30%	0.52%	0.08%
Random Walk Index for High prices	50.43%	0.97%	0.42%	0.08%
Rate of Change	70.23%	0.83%	0.37%	0.06%
Relative Momentum Index	92.47%	1.63%	0.72%	0.11%
Relative Strength Index	84.86%	1.33%	0.55%	0.09%
Relative Vigor Index	94.78%	1.73%	0.75%	0.11%
Relative Volatility Index	87.84%	1.13%	0.47%	0.07%
Runs Indicator	33.23%	2.99%	1.42%	0.29%
Stochastic Momentum Index	92.97%	1.77%	0.76%	0.11%
Stochastic Oscillator	57.48%	0.93%	0.43%	0.07%
Stochastic RSI Oscillator	57.10%	0.93%	0.38%	0.05%
The Quantitative Candlestick	20.76%	0.55%	0.23%	0.04%
Triple Exponential Smoothing	75.74%	1.50%	0.70%	0.08%
True Strength Index	89.57%	1.64%	0.72%	0.11%
Ultimate Oscillator	79.49%	1.18%	0.51%	0.07%
Vortex Oscillator	89.43%	1.19%	0.48%	0.06%
Williams Variable Accumulation Distribution	10.20%	0.28%	0.11%	0.01%
SUMMARY RESULTS FOR THE 44 RESTRICTED RULE UNIVERSES				
Minimum	*8.85%*	*0.22%*	*0.08%*	*0.00%*
Maximum	*97.91%*	*2.99%*	*1.42%*	*0.29%*
Median	*81.29%*	*1.19%*	*0.49%*	*0.07%*
Average	*69.83%*	*1.18%*	*0.51%*	*0.08%*
Std. Deviation	*25.88%*	*0.51%*	*0.23%*	*0.04%*
BENCHMARK RESULTS–686k	*99.40%*	*0.65%*	*0.27%*	*0.04%*

NOTE. **rPR is the rate of positive returns, which is defined as the number of tests for which the excess average return of the best rule was higher compared to the buy-and-hold, divided by the total number of tests performed. The Null Rejection Rate (NRR) is the number of tests that reject the null hypothesis of no economic profitability at the α confidence level, expressed as a percentage of the total number of tests performed.*

Overall, we find that the data snooping bias arising from using small, unrepresentative universes also inflates the number of null rejections in RC tests performed on real stock market data. Because false discoveries skew results in favor of showing TTR excess performance more often, this supports our earlier conclusion in favor of exercising more caution when interpreting positive results from existing tests examining the superior predictive ability of TTRs.

The results obtained when testing the extended *686k* rule universe allow us to reconsider the economic profitability of TTRs when controlling for the data snooping efforts of others. Specifically, we find very few instances in which TTRs display economic relevance. Also, null rejection rates (Table 5, Panel A, section iii) are similar and even lower compared to the ones obtained when employing the same universe and the same sample length, but randomly generated data (Table B3 in the supplementary materials). This implies that the observed null rejections fall within the bounds of randomness. Moreover, the likelihood of TTRs replicating significant performance in consecutive years is close to zero (Table 5, Panel A, section v). Overall, the results show that deviations from market efficiency are small, rare, and most likely random. This implies that stock prices are efficient at pricing information obtained using technical analysis indicators and supports the weak-form Efficient Market Hypothesis of Fama (1970). We thus conclude that TTRs do not have any economic relevance and are not able to help investors earn significant, systematic excess returns by timing stock markets around the world. This contradicts some previous conclusions reported in the related literature, which are more favorable to TTRs. Based on the analysis performed in this paper, we argue that the disagreement is at least partially caused by a data snooping bias arising in previous tests from using small, unrepresentative universes.

5.2.Results for the RC test in restricted samples

To get a better understanding of the impact of data snooping bias from using small, unrepresentative universes in limited data samples, we group and analyze the results by year and by stock market. As before, we compare null rejections obtained for the *686k* universe, the 44 restricted universes, and the restricted universes grouped by both stock and year.

The results aggregated by year are reported in Table 7. On the one hand, they show that RC tests using the benchmark *686k* rule universe reject the null hypothesis less often compared to tests that use restricted rule universes, in any time interval. Depending on the year and on the way the analysis is conducted, the excess performance of TTRs in tests that use restricted universes is inflated by up to 24 times. For example, null rejections increase on average 2.26 times in 2008 at the 1% level, 5.75 times in 2012 at the 5% level and 3.44 times in 2010 at the 10% level. If more researchers conduct independent tests using unrepresentative rule universes and then draw conclusions based on a meta-analysis of their results, false discoveries increase by 975% in 2011 at the 1% level, 2275% in 2012 at the 5% level and 1744% in 2010 at the 10% level. Overall, the results show that data snooping biases test results in favor of TTR excess performance, irrespective of the period in which the analysis is performed.

Table 7. Null Rejection Rates aggregated by year

	Number of tests	***686k***			**Restricted rule universes**			**Restricted rule universes (aggregated by stock-year)**		
Year	**using 686k**	**NRR, α=0.10**	**NRR, α=0.05**	**NRR, α=0.01**	**NRR, α=0.10**	**NRR, α=0.05**	**NRR, α=0.01**	**NRR, α=0.10**	**NRR, α=0.05**	**NRR, α=0.01**
1979	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1980	18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1981	86	2.33%	1.16%	1.16%	2.69%	2.16%	1.10%	8.13%	3.48%	1.16%
1982	94	0.00%	0.00%	0.00%	1.28%	0.14%	0.00%	5.31%	2.12%	0.00%
1983	97	0.00%	0.00%	0.00%	0.39%	0.00%	0.00%	1.03%	0.00%	0.00%
1984	133	2.26%	0.75%	0.00%	2.90%	1.65%	0.11%	4.51%	3.00%	1.50%
1985	202	0.50%	0.00%	0.00%	0.50%	0.20%	0.00%	1.48%	0.49%	0.00%
1986	214	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	1.40%	0.00%	0.00%
1987	247	0.81%	0.00%	0.00%	1.02%	0.53%	0.01%	2.83%	0.80%	0.80%
1988	278	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1989	293	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1990	343	2.04%	1.17%	0.00%	3.20%	1.68%	0.29%	9.32%	4.37%	1.16%
1991	406	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	1.97%	0.00%	0.00%
1992	461	0.65%	0.22%	0.00%	1.19%	0.45%	0.00%	4.77%	2.16%	0.00%
1993	560	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.17%	0.00%	0.00%
1994	660	0.45%	0.15%	0.00%	0.79%	0.22%	0.00%	2.72%	0.75%	0.00%
1995	759	1.05%	0.13%	0.00%	1.29%	0.30%	0.00%	3.42%	1.71%	0.00%
1996	907	0.44%	0.44%	0.00%	0.87%	0.48%	0.07%	3.19%	1.32%	0.33%
1997	997	0.60%	0.10%	0.00%	1.35%	0.46%	0.02%	4.91%	2.00%	0.20%
1998	1083	1.29%	0.37%	0.00%	1.95%	0.79%	0.04%	6.18%	3.13%	0.46%
1999	1159	0.26%	0.26%	0.00%	0.42%	0.25%	0.00%	1.63%	0.60%	0.08%
2000	1265	1.26%	0.79%	0.24%	2.33%	1.14%	0.28%	8.14%	3.55%	0.86%
2001	1353	0.81%	0.22%	0.00%	1.19%	0.39%	0.04%	3.47%	2.06%	0.22%
2002	1432	0.28%	0.00%	0.00%	0.94%	0.36%	0.00%	4.60%	1.53%	0.13%
2003	1535	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.58%	0.13%	0.00%
2004	1620	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.30%	0.12%	0.06%
2005	1686	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.83%	0.29%	0.05%
2006	1784	0.22%	0.00%	0.00%	0.33%	0.08%	0.00%	1.56%	0.56%	0.05%
2007	1941	0.10%	0.10%	0.05%	0.22%	0.10%	0.06%	1.08%	0.46%	0.10%
2008	2043	4.89%	2.40%	0.34%	8.53%	4.05%	0.77%	23.25%	12.28%	2.59%
2009	2102	0.10%	0.00%	0.00%	0.06%	0.00%	0.00%	0.57%	0.14%	0.00%
2010	2223	0.09%	0.00%	0.00%	0.31%	0.11%	0.00%	1.66%	0.67%	0.08%
2011	2283	0.74%	0.31%	0.04%	1.97%	0.73%	0.07%	8.49%	3.54%	0.43%
2012	2308	0.35%	0.04%	0.00%	0.57%	0.23%	0.00%	2.38%	0.95%	0.12%
2013	2311	0.22%	0.13%	0.04%	0.34%	0.15%	0.05%	1.29%	0.51%	0.08%
Minimum		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Maximum		4.89%	2.40%	1.16%	8.53%	4.05%	1.10%	23.25%	12.28%	2.59%
Median		*0.26%*	*0.00%*	*0.00%*	*0.50%*	*0.20%*	*0.00%*	*1.97%*	*0.67%*	*0.06%*
Average		*0.62%*	*0.25%*	*0.05%*	*1.06%*	*0.48%*	*0.08%*	*3.46%*	*1.51%*	*0.30%*
Std. Deviation		*0.99%*	*0.49%*	*0.20%*	*1.58%*	*0.82%*	*0.23%*	*4.35%*	*2.27%*	*0.56%*

NOTE. The Null Rejection Rate (NRR) *is the number of tests that reject the null hypothesis of no economic profitability at the* α *confidence level, expressed as a percentage of the total number of tests performed.*

On the other hand, tests that use the *686k* universe show temporal variations in the excess performance of TTRs and can be used to derive implications for the discussion on time-varying market efficiency. Specifically, periods in which prediction models have low success rates (stock markets are more efficient) relate to calm and favorable (positive) market conditions, while periods in which they are able to earn economically significant excess returns (stock markets are less efficient) relate to periods of financial, macroeconomic, social instability. The most successful year for TTRs is by far 2008, the climax of the recent financial crisis. About half of all RC null rejections originate in this year alone. Other periods of financial market instability rank high, such as the European sovereign debt crisis around 2011, the dot-com bubble burst at the beginning of the current millennia, or the Asian financial crisis around 1998. Also, the average excess returns earned by trading strategies and the null rejection rates of RC tests are lower in the first half of the sample, which generally corresponds to a period of more stable (rising) markets.

These findings are consistent with existing evidence that shows rising return predictability when prices decline (e.g., Lim and Brooks, 2011) and seem to support the Adaptive Market Hypothesis of Lo (2004). They also hint that TTRs may have some merit as a risk management aid in timing exit points around the onset of bear markets. However, we argue that the results rather support the Efficient Market Hypothesis and reinforce the conclusion that TTRs lack economic relevance. First, we show in Section 4 that data snooping bias increases with volatility. Thus, the similarities between the null rejection rates obtained in this empirical investigation and the ones obtained in the simulation exercise (Section 4) blur the distinction between true economic relevant results and data snooping. Second, null rejection rates are low even in 2008: they do not exceed 4.89% at the 10% level, 2.4% at the 5% level and 1.16% at the 1% level. Most would agree that investors would not use trading strategies with such low maximum success rates. Third, the randomness observed in excess performance would not encourage an ex-ante decision to use TTRs in any year.

Table 8. Null Rejection Rates aggregated by stock market

	Number of tests using 686k	**Extended rule universe (686k)**			**Restricted rule universes**			**Restricted rule universes (aggregated by stock-year)**		
Market		**α=0.10**	**α=0.05**	**α=0.01**	**α=0.10**	**α=0.05**	**α=0.01**	**α=0.10**	**α=0.05**	**α=0.01**
AE	295	0.34%	0.00%	0.00%	1.64%	0.56%	0.01%	7.79%	2.37%	0.33%
AR	600	1.50%	0.17%	0.00%	2.10%	0.65%	0.00%	5.50%	3.00%	0.16%
AT	301	0.00%	0.00%	0.00%	0.27%	0.06%	0.00%	3.32%	0.66%	0.33%
AU	865	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	1.27%	0.23%	0.00%
BA	75	9.33%	6.67%	5.33%	10.46%	7.74%	5.43%	16.00%	12.00%	8.00%
BE	352	0.28%	0.00%	0.00%	0.85%	0.47%	0.01%	2.27%	1.13%	0.56%
BG	143	2.10%	0.70%	0.00%	4.96%	2.93%	0.15%	7.69%	6.29%	1.39%
BH	260	3.85%	2.31%	0.38%	4.89%	2.62%	0.81%	12.30%	6.92%	1.92%
BR	574	0.17%	0.17%	0.00%	0.34%	0.19%	0.00%	1.56%	0.52%	0.00%
BRVM	120	0.83%	0.00%	0.00%	0.45%	0.01%	0.00%	3.33%	0.83%	0.00%
CA	1060	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.75%	0.00%	0.00%
CH	342	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.29%	0.00%	0.00%
CL	532	1.13%	0.75%	0.00%	2.16%	0.85%	0.29%	7.14%	3.00%	0.93%
CN	349	0.00%	0.00%	0.00%	0.52%	0.08%	0.00%	6.01%	2.29%	0.00%
CO	168	1.19%	0.00%	0.00%	0.96%	0.64%	0.00%	2.38%	1.19%	0.00%
CY	222	0.00%	0.00%	0.00%	3.53%	0.92%	0.04%	11.26%	4.95%	0.00%
CZ	136	0.00%	0.00%	0.00%	0.18%	0.06%	0.01%	0.73%	0.00%	0.00%
DE	662	0.00%	0.00%	0.00%	0.04%	0.01%	0.00%	0.90%	0.30%	0.00%
DK	419	0.00%	0.00%	0.00%	0.19%	0.02%	0.00%	2.62%	0.71%	0.00%
EE	154	2.60%	1.95%	0.00%	4.06%	2.72%	0.42%	6.49%	4.54%	1.94%
EG	325	2.77%	1.54%	0.00%	3.99%	1.79%	0.17%	12.92%	5.53%	0.61%
ES	578	0.00%	0.00%	0.00%	0.15%	0.03%	0.00%	2.59%	0.86%	0.00%
FI	436	0.23%	0.00%	0.00%	0.39%	0.05%	0.00%	2.98%	1.60%	0.00%
FR	926	0.11%	0.00%	0.00%	0.20%	0.09%	0.00%	1.29%	0.43%	0.10%
GR	624	1.12%	0.00%	0.00%	2.06%	0.78%	0.03%	8.17%	3.04%	0.16%
HK	758	0.00%	0.00%	0.00%	0.22%	0.02%	0.00%	0.65%	0.52%	0.00%
HR	214	0.47%	0.00%	0.00%	2.12%	0.75%	0.07%	7.94%	3.73%	0.00%
HU	183	0.00%	0.00%	0.00%	0.44%	0.11%	0.00%	2.73%	0.54%	0.00%
ID	567	0.71%	0.35%	0.00%	1.08%	0.56%	0.02%	3.35%	1.05%	0.35%
IE	412	0.73%	0.49%	0.00%	1.95%	0.94%	0.04%	5.58%	2.91%	0.48%
IL	664	0.30%	0.00%	0.00%	1.14%	0.27%	0.00%	3.61%	1.50%	0.00%
IN	470	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	1.70%	0.42%	0.00%
IQ	153	2.61%	0.00%	0.00%	2.80%	1.27%	0.00%	7.84%	3.92%	0.00%
IS	59	1.69%	1.69%	0.00%	1.57%	1.27%	0.00%	3.38%	3.38%	0.00%
IT	566	0.00%	0.00%	0.00%	0.21%	0.03%	0.00%	3.71%	1.23%	0.00%
JO	487	0.21%	0.00%	0.00%	1.32%	0.22%	0.00%	5.95%	1.64%	0.20%
JP	981	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	1.01%	0.20%	0.00%
KE	560	0.00%	0.00%	0.00%	0.75%	0.10%	0.00%	6.78%	2.14%	0.00%
KR	705	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.85%	0.14%	0.00%
KW	572	0.17%	0.00%	0.00%	1.52%	0.31%	0.00%	5.41%	2.62%	0.17%
KZ	45	4.44%	2.22%	2.22%	4.74%	2.52%	1.91%	11.11%	8.88%	2.22%
LB	115	0.00%	0.00%	0.00%	1.87%	0.98%	0.11%	6.95%	4.34%	2.60%
LK	609	1.64%	0.82%	0.00%	2.36%	1.29%	0.17%	6.23%	3.77%	0.98%
LT	257	3.11%	2.33%	0.00%	4.16%	2.50%	0.60%	8.17%	4.66%	1.55%
LV	182	3.85%	2.75%	1.10%	4.08%	3.12%	1.36%	6.04%	4.39%	2.74%
MA	400	0.50%	0.25%	0.25%	1.34%	0.61%	0.18%	4.00%	2.25%	0.50%
MU	147	0.68%	0.00%	0.00%	2.21%	0.54%	0.00%	10.20%	2.72%	0.00%
MX	485	0.62%	0.21%	0.00%	0.93%	0.56%	0.00%	3.09%	1.44%	0.20%
MY	963	0.52%	0.00%	0.00%	0.83%	0.32%	0.00%	3.32%	1.34%	0.20%
NA	193	1.04%	0.00%	0.00%	0.83%	0.29%	0.00%	4.14%	2.59%	1.55%
NG	217	2.30%	0.00%	0.00%	2.11%	0.47%	0.00%	5.99%	2.76%	0.00%
NL	833	0.84%	0.12%	0.00%	1.00%	0.52%	0.09%	2.52%	1.44%	0.36%
NO	622	0.32%	0.16%	0.00%	0.65%	0.12%	0.00%	3.37%	1.28%	0.00%
NZ	547	1.10%	0.73%	0.00%	1.32%	0.89%	0.12%	3.10%	1.64%	0.91%
OM	381	3.15%	1.84%	0.00%	3.85%	1.95%	0.11%	10.49%	5.24%	0.78%
PE	492	3.05%	1.42%	0.41%	4.44%	1.83%	0.43%	10.56%	5.28%	0.81%
PH	505	0.79%	0.40%	0.00%	1.75%	0.62%	0.07%	5.34%	2.57%	0.19%
PK	538	0.56%	0.19%	0.00%	1.97%	0.60%	0.02%	5.94%	2.60%	0.18%
PL	364	0.55%	0.27%	0.00%	0.55%	0.37%	0.01%	1.64%	0.54%	0.27%
PT	545	0.37%	0.00%	0.00%	2.09%	0.66%	0.00%	8.44%	4.22%	0.00%
QA	407	0.25%	0.00%	0.00%	0.64%	0.25%	0.01%	2.94%	0.98%	0.24%
RO	458	1.53%	0.87%	0.00%	2.93%	1.50%	0.15%	6.11%	3.93%	1.09%
RS	102	1.96%	0.00%	0.00%	5.35%	1.50%	0.04%	13.72%	7.84%	0.00%
RU	149	3.33%	2.67%	0.00%	4.24%	2.34%	0.12%	8.72%	5.36%	1.34%
SA	415	0.00%	0.00%	0.00%	0.44%	0.12%	0.00%	3.37%	2.16%	0.00%

SE	666	0.00%	0.00%	0.00%	0.26%	0.07%	0.00%	1.95%	0.30%	0.00%
SG	686	0.15%	0.00%	0.00%	0.53%	0.19%	0.01%	2.18%	0.87%	0.14%
SI	82	2.44%	1.22%	0.00%	5.70%	3.70%	1.24%	13.41%	9.75%	3.65%
SK	72	0.00%	0.00%	0.00%	0.85%	0.00%	0.00%	4.16%	0.00%	0.00%
TH	701	0.71%	0.29%	0.14%	1.36%	0.58%	0.14%	5.13%	2.28%	0.28%
TN	313	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%	2.55%	0.31%	0.00%
TR	636	0.31%	0.00%	0.00%	0.42%	0.14%	0.00%	1.41%	0.62%	0.00%
TW	729	0.00%	0.00%	0.00%	0.18%	0.01%	0.00%	2.19%	0.54%	0.00%
TZ	69	1.45%	0.00%	0.00%	3.22%	0.79%	0.09%	14.49%	5.79%	4.34%
UA	153	7.19%	1.96%	0.00%	8.49%	3.59%	0.11%	26.79%	14.37%	1.30%
UK	821	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.48%	0.00%	0.00%
US	1072	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.65%	0.18%	0.00%
VE	138	0.00%	0.00%	0.00%	1.79%	1.18%	0.08%	2.89%	2.89%	1.44%
VN	227	4.85%	3.52%	0.88%	6.89%	4.09%	1.31%	13.65%	8.81%	3.96%
ZA	682	0.15%	0.00%	0.00%	0.06%	0.00%	0.00%	0.73%	0.14%	0.00%
Minimum		*0.00%*	*0.00%*	*0.00%*	*0.01%*	*0.00%*	*0.00%*	*0.29%*	*0.00%*	*0.00%*
Maximum		*9.33%*	*6.67%*	*5.33%*	*10.46%*	*7.74%*	*5.43%*	*26.79%*	*14.37%*	*8.00%*
Median		*0.42%*	*0.00%*	*0.00%*	*1.11%*	*0.50%*	*0.00%*	*3.86%*	*2.16%*	*0.15%*
Average		*1.10%*	*0.51%*	*0.13%*	*1.83%*	*0.87%*	*0.20%*	*5.45%*	*2.77%*	*0.64%*
(Std. Deviation)		*1.67%*	*1.07%*	*0.66%*	*2.05%*	*1.26%*	*0.68%*	*4.56%*	*2.83%*	*1.26%*

NOTE. The Null Rejection Rate (NRR) *is the number of tests that reject the null hypothesis of no economic profitability at the* α *confidence level, expressed as a percentage of the total number of tests performed.*

Results aggregated by stock market are reported in Table 8. One the one hand, the excess performance recorded by TTRs in the *686k* universe shows that some asymmetries also exist between the different markets in our sample. At the 10% confidence level, no RC null rejections occur for 26 stock markets (mainly developed ones), while TTRs earn excess returns 9.33% of the time in Bosnia and Herzegovina, 7.19% of the time in Ukraine, 4.85% of the time in Vietnam, 4.44% of the time in Kazakhstan, and at rates between 1% and 4% in other 50 (mainly emerging) stock markets. As we lower the confidence level towards 1%, TTRs become unprofitable in all but 8 stock markets. Overall, these results are consistent with the literature showing that TTRs are not relevant in developed stock markets and are more informative and more profitable in smaller, less developed ones.

On the other hand, our analysis shows that the positive results should be treated with caution, as they may be impacted by data snooping bias and may not necessarily imply that some markets are inefficient. First, given that volatility tends to be higher in emerging and frontier markets, TTRs are "luckier" and a higher rate of false discoveries is expected. Second, the null rejection rates obtained in tests using *686k* are at most similar compared to the ones obtained in the simulation exercise presented in Section 4. Only 8 markets can be realistically considered as exceptions but these are very small and have important trading barriers for investors. This decreases the likelihood that superior information gained using prediction models based on technical analysis is used for making actual investment decisions. Third, the rates of null rejections do not surpass 10% even in these markets and are generally low from an economic perspective. Because of this, we argue that these results are within the bounds of the Efficient Market Hypothesis (Fama, 1970; Jensen, 1978; Timmermann and Granger, 2004). Our conclusion thus departs from the related literature that supports the excess performance of TTRs in less developed stock markets.

Comparing the tests conducted using restricted rule universes to the ones that use the benchmark shows that data snooping bias is the factor that causes this disagreement. Specifically, RC tests that use the *686k* universe reject the null hypothesis less often for all markets in the sample. For example, compared to *686k*, positive discoveries increase 2.13 times at the 1% level in tests that use the restricted rule universes for Bahrain. Similarly, positive discoveries for The Netherlands increase 4.33 times at the 5% level, while positive discoveries for Kuwait increase 8.94 times at the 10% level. In the extreme, but not entirely unrealistic scenario in which researchers conduct independent tests using unrepresentative universes and then draw conclusions based on a meta-analysis of the reported results, null rejections are inflated for Bahrain 5.05 times at the 1% level, for Argentina 17.65 times at the 5% level, and for Kuwait 31.82 times at the 10% level.

5.3.Results for the SPA test

We evaluate the robustness of the empirical results to changes in the testing methodology. Particularly, the data snooping bias is reevaluated using the SPA test of Hansen (2005). The rate of null rejections is estimated and compared for tests that use the benchmark *686k* rule universe and tests that use 43 smaller, unrepresentative universes. The same as before are used, except that the universes generated by the Filter and Runs indicators (which are very small) are merged. For this exercise, only data from 18 emerging stock markets in Central and

Eastern Europe (listed in Table 9, Panel C) is considered, because the results in Section 5.2 show that this is where RC test null rejections predominantly occur. A total of 4,208 tests are performed using the benchmark universe and 180,944 tests using the 43 restricted alternatives.

Table 9. Null rejection rates and data snooping bias in SPA tests

Panel A: NRR aggregated by rule universe

Rule universe	**α=0.1**	**α=0.05**	**α=0.01**
%b	3.35%	1.64%	0.38%
%k	2.99%	1.52%	0.38%
AO	3.49%	1.64%	0.36%
ASI	23.08%	21.22%	11.86%
BMP	3.49%	2.21%	1.00%
CCI	4.25%	2.28%	0.57%
CMF	2.85%	1.66%	0.76%
CMO	2.88%	1.45%	0.36%
CO	2.50%	1.50%	0.76%
COG	6.63%	5.30%	1.21%
DI	3.54%	2.00%	0.48%
DMI	2.26%	1.07%	0.24%
DPO	0.95%	0.48%	0.14%
DYMOI	2.14%	0.97%	0.21%
EMV	19.77%	18.23%	9.91%
F	5.44%	3.49%	0.88%
INI	2.66%	1.40%	0.45%
KCD	1.28%	0.55%	0.12%
KPO	1.73%	0.67%	0.14%
KST	1.52%	0.67%	0.12%
KVO	4.56%	2.54%	0.45%
LRS	2.57%	1.05%	0.21%
MACD	1.05%	0.40%	0.05%
MFI	3.07%	1.45%	0.33%
MVI	4.06%	2.76%	0.67%
NRVI	2.92%	1.52%	0.31%
OBV	43.96%	43.13%	31.44%
PFE	3.16%	1.45%	0.33%
PI	2.73%	1.28%	0.29%
Qstick	22.24%	21.06%	11.64%
RMI	4.33%	2.26%	0.45%
ROC	1.90%	0.88%	0.19%
RSI	3.42%	1.64%	0.48%
RVI	2.71%	1.31%	0.33%
RVig	3.26%	1.78%	0.88%
RWI	5.58%	4.61%	2.14%
SMI	3.33%	1.50%	0.29%
SRSI	4.23%	2.19%	0.40%
TRIX	3.11%	1.45%	0.21%
TSI	3.64%	2.04%	0.38%
UO	2.57%	1.43%	0.36%
VX	2.61%	1.33%	0.40%
WVAD	40.54%	39.42%	26.19%
Min	*0.95%*	*0.40%*	*0.05%*
Max	*43.96%*	*43.13%*	*31.44%*
Median	*3.16%*	*1.52%*	*0.38%*
Average	*6.24%*	*4.85%*	*2.53%*
St.Dev.	*9.44%*	*9.49%*	*6.52%*
Benchmark–*686k*	0.97%	0.40%	0.07%

Panel B: NRR aggregated by year

686k				**Restricted rule universes**			
Year	**α=0.1**	**α=0.05**	**α=0.01**	**Year**	**α=0.1**	**α=0.05**	**α=0.01**
1991	0.00%	0.00%	0.00%	1991	3.15%	3.15%	1.37%
1992	0.00%	0.00%	0.00%	1992	1.95%	1.67%	0.37%
1993	0.00%	0.00%	0.00%	1993	13.24%	12.34%	9.48%
1994	2.70%	0.00%	0.00%	1994	4.90%	4.53%	2.14%
1995	0.00%	0.00%	0.00%	1995	3.98%	3.67%	2.28%
1996	0.00%	0.00%	0.00%	1996	8.34%	6.58%	3.51%
1997	0.00%	0.00%	0.00%	1997	5.59%	5.41%	3.61%
1998	0.00%	0.00%	0.00%	1998	5.13%	3.62%	1.47%
1999	0.00%	0.00%	0.00%	1999	7.35%	6.98%	4.67%
2000	1.41%	0.70%	0.70%	2000	5.26%	3.32%	1.72%
2001	1.84%	0.61%	0.00%	2001	3.65%	2.74%	1.46%
2002	2.25%	0.56%	0.00%	2002	5.62%	3.78%	1.20%
2003	0.00%	0.00%	0.00%	2003	4.72%	4.45%	2.63%
2004	0.00%	0.00%	0.00%	2004	6.14%	5.80%	3.87%
2005	0.41%	0.00%	0.00%	2005	5.74%	5.29%	3.38%
2006	0.00%	0.00%	0.00%	2006	5.15%	4.67%	2.77%
2007	0.66%	0.33%	0.33%	2007	5.72%	5.32%	3.27%
2008	7.59%	4.11%	0.32%	2008	20.54%	10.93%	2.66%
2009	0.00%	0.00%	0.00%	2009	4.40%	4.14%	2.52%
2010	0.00%	0.00%	0.00%	2010	4.32%	4.02%	2.48%
2011	1.13%	0.00%	0.00%	2011	4.61%	2.71%	1.14%
2012	0.00%	0.00%	0.00%	2012	5.17%	4.50%	2.61%
2013	0.00%	0.00%	0.00%	2013	3.96%	3.60%	1.95%
Min	*0.00%*	*0.00%*	*0.00%*	*Min*	*1.95%*	*1.67%*	*0.37%*
Max	*7.59%*	*4.11%*	*0.70%*	*Max*	*20.54%*	*12.34%*	*9.48%*
Median	*0.00%*	*0.00%*	*0.00%*	*Median*	*5.15%*	*4.45%*	*2.52%*
Average	*0.78%*	*0.27%*	*0.06%*	*Average*	*6.03%*	*4.92%*	*2.72%*
St.Dev.	*1.70%*	*0.87%*	*0.17%*	*St.Dev.*	*3.83%*	*2.47%*	*1.79%*

Panel C: NRR aggregated by stock market

Extended rule universe (686k)				**Restricted rule universes**			
Market	**α=0.1**	**α=0.05**	**α=0.01**	**Market**	**α=0.1**	**α=0.05**	**α=0.01**
BA	9.33%	5.33%	0.00%	BA	13.95%	10.45%	6.48%
BG	1.40%	0.70%	0.00%	BG	7.63%	4.91%	2.18%
CY	1.80%	0.45%	0.00%	CY	4.92%	2.67%	1.04%
CZ	0.00%	0.00%	0.00%	CZ	5.35%	4.79%	2.77%
EE	1.30%	1.30%	0.00%	EE	5.68%	4.20%	1.43%
GR	0.80%	0.32%	0.00%	GR	6.28%	4.99%	2.66%
HR	0.47%	0.00%	0.00%	HR	5.17%	3.61%	1.66%
HU	1.09%	0.00%	0.00%	HU	6.40%	5.71%	3.43%
LT	1.56%	0.39%	0.00%	LT	5.86%	4.04%	1.49%
LV	2.75%	1.10%	1.10%	LV	5.67%	4.36%	1.97%
PL	0.27%	0.00%	0.00%	PL	6.54%	6.06%	3.60%
RO	0.22%	0.22%	0.00%	RO	5.09%	3.53%	1.51%
RS	0.00%	0.00%	0.00%	RS	6.36%	3.28%	1.44%
RU	0.66%	0.66%	0.66%	RU	5.62%	4.14%	2.08%
SI	3.66%	1.22%	0.00%	SI	11.17%	9.36%	4.23%
SK	0.00%	0.00%	0.00%	SK	2.07%	1.65%	0.90%
TR	0.00%	0.00%	0.00%	TR	6.80%	6.34%	4.12%
UA	1.96%	0.65%	0.00%	UA	7.14%	3.62%	0.87%
Min	*0.00%*	*0.00%*	*0.00%*	*Min*	*2.07%*	*1.65%*	*0.87%*
Max	*9.33%*	*5.33%*	*1.10%*	*Max*	*13.95%*	*10.45%*	*6.48%*
Median	*0.95%*	*0.35%*	*0.00%*	*Median*	*6.07%*	*4.28%*	*2.02%*
Average	*1.52%*	*0.69%*	*0.10%*	*Average*	*6.54%*	*4.87%*	*2.44%*
St.Dev.	*2.20%*	*1.24%*	*0.29%*	*St.Dev.*	*2.53%*	*2.17%*	*1.46%*

NOTE. The Null Rejection Rate (NRR) *is the number of tests that reject the null hypothesis of no economic profitability at the* α *confidence level, expressed as a percentage of the total number of tests performed.*

Table 9 reports the results, showing that data snooping bias also has a significant impact on tests that use the SPA methodology. Particularly, compared to tests that use the *686k* universe, null rejection rates for unrepresentative universes occur on average (median) 6.4 to 35.4 (3.2 to 5.3) times more often, depending on the selected confidence level. When aggregating the data by year and by stock market, the relative differences in null rejection rates are similar. Moreover, many years and markets exist for which tests that use restricted rule universes reject the null hypothesis, while tests that use the benchmark do not. Overall, the results support earlier conclusions regarding the impact of data snooping bias on tests that analyze the relative performance of multiple forecasting models. In particular, they show that the excess performance of TTRs is overstated when using small, unrepresentative universes also when the SPA test is employed.

Regarding market efficiency, even though the SPA test is more powerful and rejects the null of no economic profitability more often compared to the RC test, analyzing the excess performance of TTRs in this context yields that null rejection rates remain scarce. Particularly, when employing the *686k* universe, only 41 tests (0.97%) reject the null at the 10% confidence level, 17 tests (0.40%) reject the null at the 5% confidence level, and 3 tests (0.07%) reject the null at the 1% confidence level. Because these results are obtained for some of the smallest stock markets in our sample, they provide additional support for the Efficient Market Hypothesis and our earlier conclusion that TTRs are not relevant from an economic perspective when used by investors for timing stock markets around the world.

6. Conclusions

This paper performs a novel investigation into how selecting and using small universes of prediction models, which do not account for what investors and other researchers use, biases statistical tests that evaluate their relative performance. The paper focuses on the literature concerned with the performance of models derived from technical analysis, technical trading rules (TTRs). A preliminary analysis shows that the effective span of trading rule universes is positively correlated with their size. Even though universes employed in the literature are becoming larger, they do not typically account for the data snooping efforts of others. In particular, their size and effective span are lower compared to a more representative universe of 686,304 models, which we define and use as a benchmark.

In a simulation exercise performed on randomly generated data, we find that restricting the size and diversity of trading rule universes increases false discoveries and biases RC test results in favor of showing that some TTRs have superior predictive ability. False discoveries increase with the volatility of the underlying data generating process and the length of the data sample. Changing the testing methodology from the RC test of White (2000) to the pFDR test of Storey (2002) does not significantly alter these findings. In a complementary empirical exercise performed on historical stock market data, we find that using unrepresentative universes (comparable in size and information span to the ones that are typically employed in the literature) overestimates the economic relevance of TTRs by 1.8-2 times on average. In the extreme, but not entirely unrealistic case in which independent researchers perform tests using unrepresentative universes and then draw conclusions based on a meta-analysis of their results, the excess performance of TTRs is inflated by 6.16-7.93 times on average, and even by as much as 32 times. We obtain almost identical results when changing the testing methodology from the RC test of White (2000) to the SPA test of Hansen (2005).

Overall, our findings have several important implications. First, they highlight the need to thoroughly investigate and mitigate data snooping as a way to increase the relevance and robustness of tests that evaluate the relative performance of multiple forecasting models. This contributes to the important debate on the quality of scientific output in the financial economics literature (Harvey, 2017). In particular, we argue that previous findings showing TTRs to be economically relevant should be treated with more caution. Trading rules can appear relevant even in tests that use randomly generated data and their "luckiness" increases in some setups (e.g., higher volatility, longer data samples). As a consequence, existing evidence should be reexamined using tests that control for the data snooping efforts of others. New tests should also consider and control for this problem. Here, when reevaluating the economic relevance of TTRs by accounting for transaction costs and adjusting for data snooping bias using a more representative prediction model universe, we find no significant evidence to support that they are able to earn excess returns when used for timing stock markets around the world. This implies that prices in stock markets incorporate information efficiently relative to a broad set of technical trading rules, which supports the weak-form Efficient Market Hypothesis of Fama (1970), as opposed to the Adaptive Market Hypothesis of Lo (2004) that has recently gained some visibility (Lim and Brooks, 2011).

Second and more generally, our results imply that data snooping bias occurs and is significant in statistical tests that evaluate the relative performance of multiple forecasting models without accounting for relevant alternatives. Moreover, when all alternatives are not observable, such as in the case of TTRs, then testing for relative performance becomes problematic. Testing for absolute performance avoids this problem and should provide more robust results, but evaluating relative performance remains important for answering important scientific questions. This implies the need to develop new testing methodologies that examine the relative performance of multiple forecasting models while handling for the data snooping bias caused by subjectively fixing the set of alternatives.

References

Al-Nassar, N. S., 2014. The profitability of trading rules in stock markets: Evidence from GCC countries. PhD Thesis, RMIT University.

Arellano-Valle, R. B., Genton, M. G., 2008. On the exact distribution of the maximum of absolutely continuous dependent random variables. Statistics and Probability Letters 78(1), 27-35.

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics 106(3), 473-491.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. The Journal of Finance 65(1), 179-216.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological), 289-300.

Brabazon, A., Dang, J., Dempsey, I., O'Neill, M., & Edelman, D. (2012). Natural computing in finance: a review. In Rozenberg, Grzegorz; Bäck, Thomas; Kok, Joost N.(eds.). Handbook of Natural Computing: Theory, Experiments and Applications, 1707-1735.

Coakley, J., Marzano, M., Nankervis, J., 2016. How profitable are FX technical trading rules?. International Review of Financial Analysis 45, 273-282.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25(2), 383-417.

Fifield, S.G., Power, D.M., Donald Sinclair, C., 2005. An analysis of trading strategies in eleven European stock markets. The European Journal of Finance 11(6), 531-548.

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. The American Economic Review 70(3), 393-408.

Hall, P., Horowitz, J.L., Jing, B.Y., 1995. On blocking rules for the bootstrap with dependent data. Biometrika 82(3), 561-574.

Han, Y., Yang, K. and Zhou, G., 2013. A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis, 48(5), pp.1433-1461.

Han, Y., Hu, T., Yang, J., 2016. Are there exploitable trends in commodity futures prices?. Journal of Banking and Finance 70, 214-234.

Hansen, P.R., 2005. A Test for Superior Predictive Ability. Journal of Business and Economic Statistics 23(4), 365-380.

Hartigan, J. A., 2014. Bounding the maximum of dependent random variables. Electronic Journal of Statistics 8(2), 3126-3140.

Harvey, C. R., 2017. Presidential address: The scientific outlook in financial economics. The Journal of Finance 72(4), 1399-1440.

Harvey, C. R., Liu, Y., 2014. Evaluating Trading Strategies. The Journal of Portfolio Management 40(5), 108-118.

Hou, K., Xue, C., Zhang, L., 2017. Replicating anomalies. NBER Working Paper No. 23394.

Hsu, P. H., Hsu, Y. C., Kuan, C. M., 2010. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. Journal of Empirical Finance 17(3), 471-484.

Hsu, P. H., Taylor, M. P., Wang, Z., 2016. Technical trading: Is it still beating the foreign exchange market?. Journal of International Economics 102, 188-208.

Jensen, M.C., 1978. Some Anomalous Evidence Regarding Market Efficiency. Journal of Financial Economics, 6(2/3), 95-101.

Kim, J. H., Ji, P. I., 2015. Significance testing in empirical finance: A critical review and assessment. Journal of Empirical Finance 34, 1-14.

Lim, K.P., Brooks, R., 2011. The evolution of stock market efficiency over time: A survey of the empirical literature. Journal of Economic Surveys 25(1), 69-108.

Lo, A.W., 2004. The Adaptive Markets Hypothesis. The Journal of Portfolio Management 30(5), 15-29.

Marshall, B. R., Cahan, R. M., 2005. Is the 52-week high momentum strategy profitable outside the US?. Applied Financial Economics 15(18), 1259-1267.

Menkhoff, L., 2010. The use of technical analysis by fund managers: International evidence. Journal of Banking and Finance 34(11), 2573-2586.

Metghalchi, M., Du, J., Ning, Y., 2009. Validation of moving average trading rules: Evidence from Hong Kong, Singapore, South Korea, Taiwan. Multinational Business Review 17(3), 101-122.

Metghalchi, M., Marcucci, J., Chang, Y. H., 2012. Are moving average trading rules profitable? Evidence from the European stock markets. Applied Economics 44(12), 1539-1559.

Neuhierl, A., Schlusche, B., 2010. Data snooping and market-timing rule performance. Journal of Financial Econometrics, 9(3) 550-587.

Park, C.H., Irwin, S.H., 2007. What do we know about the profitability of technical analysis?. Journal of Economic Surveys 21(4), 786-826.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89(428), 1303-1313.

Ratner, M., Leal, R. P., 1999. Tests of technical trading strategies in the emerging equity markets of Latin America and Asia. Journal of Banking and Finance 23(12), 1887-1905.

Romano, J. P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73(4), 1237-1282.

Scott, G., Carr, M., Cremonie, M., 2016. Technical Analysis: Modern Perspectives. CFA Research Foundation Reviews 11(1), 1-36.

Shynkevich, A., 2012. Performance of technical analysis in growth and small cap segments of the US equity market. Journal of Banking and Finance 36(1), 193-208.

Shynkevich, A., 2016. Predictability in bond returns using technical trading rules. Journal of Banking and Finance 70, 55-69.

Sobreiro, V. A., da Costa, T. R. C. C., Nazário, R. T. F., e Silva, J. L., Moreira, E. A., Lima Filho, M. C., Kimura, H., Zambrano, J. C. A., 2016. The profitability of moving average trading rules in BRICS and emerging stock markets. The North American Journal of Economics and Finance 38, 86-101.

Storey, J.D., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(3), 479-498.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. The Journal of Finance 54(5), 1647-1691.

Taylor, M. P., Allen, H., 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance 11(3), 304-314.

Taylor, N., 2014. The rise and fall of technical trading rule success. Journal of Banking and Finance 40, 286-302.

Timmermann, A., Granger, C. W., 2004. Efficient market hypothesis and forecasting. International Journal of Forecasting 20(1), 15-27.

Urquhart, A., Gebka, B., Hudson, R., 2015. How exactly do markets adapt? Evidence from the moving average rule in three developed markets. Journal of International Financial Markets, Institutions and Money 38, 127-147.

White, H., 2000. A reality check for data snooping. Econometrica 68(5), 1097-1126.

Zarrabi, N., Snaith, S., Coakley, J., 2017. FX technical trading rules can be profitable sometimes!. International Review of Financial Analysis 49, 113-127.

Data Snooping Bias in Tests of the Relative Performance of Multiple Forecasting Models

Supplementary Materials

Appendix A. Constructing the benchmark *686k* trading rule universe

A1. Motivation

How large is the "true" universe of relevant technical trading rules? Researchers typically test around 10^3 to 10^4 of the most popular rules and a few recent papers marginally surpass 10^5. However, anecdotal evidence and a review of the practitioner-orientated literature suggests that the universe from which traders select rules is much larger and diverse. More specifically: (1) technical analysis is relatively easy and cheap to learn and implement compared to other forecasting methods; (2) technological advances (such as increased computing power) and wider data availability have decreased the cost of designing, testing and implementing mechanical trading systems; (3) there is a wide market for technical analysis services that are offered by brokers and financial consultants; (4) there has been rapid development of many dedicated software packages that contain hundreds of build-in indicators and also give users the possibility to define custom trading rules; (5) the typical procedure for backtesting TTRs involves some kind of optimization algorithm that iterates through many possible parameter combinations. Also, the attractiveness of technical analysis to investors may have increased due to findings that support its predictive power in financial markets, even compared with other forecasting methods. For example, Avramov et al. (2018) find that technicians are able to predict individual stock returns to economically significant degrees up to a one-year horizon, while fundamental analysis has little ability to predict future returns. Also, Lin (2018) finds that a technical analysis aligned index exhibits statistically and economically significant in-sample and out-of-sample predictive power and outperforms well-known macroeconomic variables. The rule universe that we define, denoted as *686k*, takes account of these developments and should be more representative for the "true" universe that investors and other researchers use.

A2. The search exercise for popular trading rules

On the one hand, *686k* contains rules that have been extensively used in key related papers, such as Brock et al. (1992), Sullivan et al. (1999) or Bajgrowicz and Scaillet (2012). Park and Irwin (2007) review the relevant literature on which we base our choice. On the other hand, we supplement *686k* with other popular trading rules developed by practitioners and not used before in statistical tests. In general, the practitioner orientated literature is large and consists of articles published in non-scientific journals, books, or various online materials. For constructing the trading rules in this paper, we search the Journal of Technical Analysis, published by the Chartered Market Technician Association; the IFTA Journal, published by the International Federation of Technical Analysts; and the Technical Analysis of Stocks & Commodities: The Traders' Magazine (see www.traders.com/). Technical analysis books, such as Wilder (1978), Achelis (2001), or Colby (2002), are also examined. Proposals for trading strategies that allegedly "beat the market" are frequently published. Most of the time, the arguments are based on valid economic reasoning. However, these are only accompanied by anecdotal evidence, while statistical tests that thoroughly evaluate their performance, especially in a way that controls for data snooping, are not implemented.

We focus on unambiguous trading rules that can be mechanically implemented, technical analysis indicators. Depending on the characteristics of their output series, indicators can be classified as *simple*, *oscillators*, or *standardized oscillators*. Oscillators are indicators that display a mean-reverting behavior and "oscillate" around an "equilibrium" level. Standardized oscillators are additionally bounded within a fixed interval, usually [-100, 100] or [0, 100]. Depending on their functionality, several types of indicators exist. *Trend indicators* measure the intensity of price movements without measuring their direction. Examples are the Average Directional Movement Index or the Commodity Selection Index (Wilder, 1978). *Momentum indicators* measure both the intensity and the direction of price movements. They constitute the largest and the most popular (frequently published or analyzed) group of indicators. Examples are the Stochastic Oscillator (Lane, 1984), the Relative Strength Index (Wilder, 1978), or the Moving Average Convergence Divergence indicator (Appel, 1999). *Volume indicators* measure the amplitude or direction of traded volume. Dormeier (2011) provides several examples. *Money Flow indicators* are volume-adjusted momentum indicators, meaning that they measure the momentum of price movements and modify this by incorporating information regarding the amount of volume supporting it. Some examples are the On Balance Volume (Sweeney, 1997), the Williams Variable Accumulation Distribution indicator (Williams, 1986) or the Money Flow Index (Quong and Soudack, 1989). *Volatility indicators* measure the variability of prices or returns. Practitioners use standard measures of volatility, such as the standard deviation or the semivariance, alongside specialized measures that also use the information in the open, high, and low prices. Examples are the Average True Range indicator (Wilder, 1978), the Bollinger Band Width Index (Bollinger, 2001), or the Ulcer Index (Martin and McCann, 1989). *Market Breadth indicators* gauge overall investor attitude (optimistic, pessimistic, or neutral) towards the state and future direction of the market. They are similar to market sentiment indicators, except they are calculated based on backward-looking historical market data, instead of forward-looking investor surveys. Colby (2002) gives several examples of such indicators, which include the Advance/Decline Ratio and the Breadth Advance/Decline Indicator.

Pure trend, pure volume, and pure volatility indicators are seldom used on their own to construct trading rules because they do not provide information regarding the direction of expected price movements. Also, market breadth indicators are more appropriate for tactical allocation decisions and are difficult to use for predicting expected prices at the individual asset level. Thus, given their stand-alone usage, intuitive interpretation, versatility, and popularity, we decide to only use momentum and money flow indicators.

Table A2.1. Summary of technical analysis indicators

No.	Indicator Name (Symbol)	Indicator Type	References
1	Accumulation Swing Index (ASI)	momentum	Wilder (1978)
2	Arms Ease of Movement (EMV)	momentum	Arms (1983)
3	Aroon Oscillator (AO)	standardized momentum	Chande (1995)
4	Balance of Market Power (BMP)	standardized momentum	Livshin (2001)
5	Bollinger Oscillator (%b)	momentum	Bollinger (2001)
6	Center of Gravity Oscillator (COG)	momentum	Ehlers (2002b)
7	Chaikin Money Flow (CMF)	standardized money flow	Chaikin (1994)
8	Chaikin Oscillator (CO)	money flow	Ehrlich(2000), Achelis (2001)
9	Chande Momentum Oscillator (CMO)	standardized momentum	Colby (2002)
10	Commodity Channel Index (CCI)	momentum	Lambert (1982)
11	Consecutive Runs (CR)	momentum	
12	Demand Index (DI)	standardized money flow	Aspray (1986)
13	Detrended Price Oscillator (DPO)	momentum	Achelis (2001)
14	Dynamic Momentum Index (DYMOI)	standardized momentum	Chande and Kroll (1994)
15	Filter (F)	momentum	Alexander (1961, 1964)
16	Inertia Indicator (INI)	standardized momentum	Dorsey (1995)
17	Kase Convergence Divergence (KCD)	momentum	Kase (1996)
18	Kase Peak Oscillator (KPO)	momentum	Kase (1996)
19	Klinger Volume Oscillator (KVO)	money flow	Klinger (1994), Klinger (1995)
20	Know Sure Thing (KST)	momentum	Pring (1997)
21	Linear Regression Slope (LRS)	momentum	Hayashi (2000)
22	Market Volume Impact (MVI)	money flow	Macek (1993)
23	Money Flow Index (MFI)	money flow	Quong and Soudack (1989)
24	Moving Average Convergence Divergence (MACD)	momentum	Appel (1999)
25	New Relative Volatility Index (NRVI)	standardized momentum	Dorsey (1995)
26	On Balance Volume (OBV)	money flow	Sweeney (1997)
27	PDM (+DM) vs MDM (-DM) crossover rule (DMI)	standardized momentum	Wilder (1978)
28	PI Opinion Oscillator (PI)	standardized momentum	Colby (2002)
29	Polarized Fractal Efficiency (PFE)	standardized momentum	Hannula (1994)
30	Random Walk Index for High prices (RWI)	momentum	Poulos (1991)
31	Rate of Change (ROC)	momentum	Pring (1992), Faber (1994)
32	Relative Momentum Index (RMI)	standardized momentum	Altman (1993)
33	Relative Strength Index (RSI)	standardized momentum	Wilder (1978)
34	Relative Vigor Index (RVig)	standardized momentum	Ehlers (2002a)
35	Relative Volatility Index (RVI)	standardized momentum	Dorsey (1993)
36	Stochastic Momentum Index (SMI)	standardized momentum	Blau (1993)
37	Stochastic Oscillator (%k)	standardized momentum	Lane (1984), Schirding (1984)
38	Stochastic RSI Oscillator (SRSI)	standardized momentum	Chande and Kroll (1994)
39	The Quantitative Candlestick (Qstick)	momentum	Chande and Kroll (1994)
40	Triple Exponential Smoothing (TRIX)	momentum	Hutson (1983)
41	True Strength Index (TSI)	standardized momentum	Blau (1991)
42	Ultimate Oscillator (UO)	standardized momentum	Williams (1985)
43	Vortex Oscillator (VX)	standardized momentum	Botes and Siepman (2010)
44	Williams Variable Accumulation Distribution (WVAD)	money flow	Williams (1986)

We select the indicators based on their popularity and distinctiveness. Popularity is subjectively evaluated using the number of references we find for each indicator. Distinctiveness is also subjectively applied as a criterion by avoiding as much as possible indicators that have similar characteristics and, thus, produce highly correlated predictions. In the end, we decide on a sample of 44 indicators, which are summarized in Table A2.1. For each, we provide at least one reference that explains how to implement and interpret its values. For brevity, other details are not presented. Many are also described in Colby (2002), while additional explanations can be provided at request.

A3. Trading strategies

Besides incorporating a large and diverse selection of indicators, the new universe distinguishes itself by also diversifying the type of entry and exit strategies used to define TTRs. This is important, as recent results show that seldom considered strategies may extract valuable information from financial price data. For example, Hudson et al. (2017) find that contrarian rules are able to better predict prices and are more profitable compared to trend-following rules.

Three types of entry strategies are implemented. The *"standard trend-seeking strategy"* goes long when the value of an indicator is larger than a specified threshold[1]. This type of strategy can be used for almost any type of indicator, as it does not depend on its characteristics. The *"trend anticipation strategy"* goes long when the value of the indicator increases. This is usually implemented using *"signal bands (lines)"*, which are typically a short-term moving average or a delayed series of the indicator[2]. This type of strategy can mainly be used for oscillators, as it specifically relies on their mean-reverting characteristics. Finally, the *"trend reversal"* strategy goes long when unusual low values are reached by the indicator, these signaling that the market is "oversold"[3]. This type of strategy can only be implemented for standardized oscillators, which take values in a fixed interval.

To close positions, exit strategies that mirror the entry ones are used: (1) exit when the value of the indicator reaches a threshold that signals that a declining trend is about to begin (applicable for all indicators), (2) exit when the value of the indicator crosses below the signal band (applicable for oscillators), and (3) exit when a high extreme value is reached for an indicator, which signals that the market is "overbought" (applicable for standardized

---

[1] For example, an investor would go long when the MACD indicator takes positive values and would stay out of the market otherwise. In this example, the threshold value of zero distinguishes between rising and declining expected trends. This threshold is a parameter and other values can be used instead.

[2] For example, when the MACD indicator crosses over its "signal line", it suggests that a new rising trend is about to form and instructs the investor to go long.

[3] The classic example is based on the RSI indicator. Investors go long when the RSI falls below and then rises above the threshold value of 20, which is yet another parameter that can be modified as considered.

oscillators). Some authors test TTRs by closing positions after a designated number of observations. This exit strategy is not implemented here because it does not generally correspond to how traders make decisions and also because the considered time interval is just another parameter that exposes the analysis to an extra layer of data snooping risk,

The open and exit strategies are applied in all possible combinations to the 44 indicators depending on their type. This results in 167 template trading rules that depend on a set of parameters, such as the lookback windows for indicators and threshold values for strategies. The actual trading rules are then defined by setting values for all parameters of each template rule. Note that, because of the known limitations on short selling, we only consider trading rules that go long and earn the next day's market return, while exiting the market is equivalent to going into cash and earning a zero return. Section A3 in this supplementary materials presents the template trading rules and provides a description of the parameterization procedure, which produces universes of between 11 and 60,426 trading rules for each indicator, depending on its characteristics, such as the number of parameters and the type of trading strategies it supports.

A4. Assigning parameters to template trading rules

Here, we summarize the procedure used to construct the 686,304 trading rules in the *686k* benchmark rule universe. Table A4.1 presents a legend of symbols, while Table A4.2 presents the template trading rules and detailed instructions and restrictions used to set parameters. The parameterization procedure goes as follows. For all trading rule templates, the first rule is defined by assigning the minimum value to all parameters. All other rules are defined by consecutively incrementing the value of all parameters (in the order they appear in the template) until all possible combinations within the defined bounds have been reached. For example, in the case of the Balance of Market Power (BMP) indicator, there are five types of template trading rules. The first one, *"BMP(n) > x"*, represents a pure momentum strategy. The first set of assigned parameters are $n = 2$ and $x = -90$ and make the rule "BMP(2) > -90", which instructs the investor to go long when the value of the 2-day BMP is higher than -90 and stay out of the market otherwise. The second rule is “BMP(4) > -90”; the third one is “BMP(6) > -90” and so on until “BMP(42) > -90” is defined. Then, x is incremented once and “BMP(2) > -85”, “BMP(4) > -85”, …,“BMP(42) > -85”, “BMP(2) > -80”, “BMP(4) > -85” are defined. This continues until the rule “BMP(42) > 90” is defined. The procedure then moves to the second template, *“BMP(n1) > S(n2)”*, and assigns parameters in a similar fashion until all possible parameter combinations within the specified bounds have been used.

Table A4.1. Legend

Symbol	**Significance**
#	Number of.
>	[Operator] “Greater than.”
<	[Operator] “Lower than.”
|↗	[Operator] “The value (of the series) on the left falls below and then rises above the value (of the series) on the right.” E.g. evaluating x_t |↗ y_t: *if* $x_{t-1} \leq y_{t-1}$ *and* $x_t > y_t$ *then true, else false*.
|↘	[Operator] “The value (of the series) on the left rises above and then falls back below the value (of the series) on the right.” E.g. evaluating x_t |↘ y_t: *if* $x_{t-1} \geq y_{t-1}$ *and* $x_t < y_t$ *then true, else false*.
||	A separator between entry rule(s) and exit rule(s) for a trading strategy. E.g. usage: “entry rule(s) || exit rule(s)”. When missing, the exit rule is defined as the negated entry rule.
n, n1, n2, …	Parameters associated with the length of the lookback period (window lengths).
x, x1, x2, …	Parameters associated with threshold values for trading strategies based on technical analysis indicators.
S(n)	Signal line (band) based on a simple moving average of historical indicator values, where the time window is *n* observations.
D(n)	Signal line (band) based on a delayed series of historical indicator values, where the delay is *n*.

Table A4.2. Indicators, template trading rules, and parameter iteration rules/restrictions

Indicator (Symbol)	**Trading rule template**	**Window length (n) parameters**				**Threshold (x) parameters**				**Additional restrictions**	**# trading rules**
		#	**Min**	**Max**	**Incr.**	**#**	**Min**	**Max**	**Incr.**		
Filter (F)											***50***
	F > x					1	0.01	0.5	0.01		50
Runs indicator (R)											***11***
	R > x					1	0	10	1		11
PDM (+DM) vs MDM (-DM) crossover											***441***
	DM_PLUS(n1) > DM_MINUS(n2)	2	2	43	2						441
Accumulation Swing Index (ASI)											***210***
	ASI(n1) > S(n2)	2	2	43	2					$n_2 \leq 14$	147
	ASI(n1) > D(n2)	2	2	43	2					$n_2 \leq 7$	63
Arms Ease of Movement (EMV)											***840***
	EMV(n) > x	1	2	43	1					x=0	42
	EMV(n1) > S(n2)	2	2	43	1					$n_2 \leq 14$	546
	EMV(n1) > D(n2)	2	2	43	1					$n_2 \leq 7$	252
Aroon Oscillator (AO)											***10,507***
	AO(n) > x	1	2	43	2	1	-90	90	10		399
	AO(n) |↗ x1 || *AO*(n) > x2	1	2	43	3	2	-90	90	10		5054
	AO(n) |↗ x1 || *AO*(n) |↘ x2	1	2	43	3	2	-90	90	10		5054
Balance of Market Power (BMP)											***39,207***
	BMP(n) > x	1	2	43	2	1	-90	90	5		777
	BMP(n1) > S(n2)	2	2	43	3					$n_2 \leq 14$	70
	BMP(n1) > D(n2)	2	2	43	3					$n_2 \leq 7$	28
	BMP(n) |↗ x1 || BMP(n) > x2	1	2	43	3	2	-90	90	5		19,166
	BMP(n) |↗ x1 || BMP(n) |↘ x2	1	2	43	3	2	-90	90	5		19,166
Bollinger Oscillator (%b)											***12,402***
	%B(n) > x	1	2	43	5	1	-100	100	10		189
	%B(n1) > S(n2)	2	2	43	5					$n_2 \leq 14$	27
	%B(n1) > D(n2)	2	2	43	5					$n_2 \leq 7$	18
	%B(n) |↗ x1 || %B(n) > x2	1	2	43	5	2	-150	100	10		6,084
	%B(n) |↗ x1 || %B(n) |↘ x2	1	2	43	5	2	-150	100	10		6,084
Center of Gravity Oscillator (COG)											***252***
	COG(n) > x	1	2	43	1					x=0	42
	COG(n1) > S(n2)	2	2	43	2					$n_2 \leq 14$	147
	COG(n1) > D(n2)	2	2	43	2					$n_2 \leq 7$	63
Chaikin Money Flow (CMF)											***25,258***
	CMF(n) > x	1	2	43	3	1	-90	90	5		518
	CMF(n1) > S(n2)	2	2	43	3					$n_2 \leq 14$	70
	CMF(n1) > D(n2)	2	2	43	3					$n_2 \leq 7$	28
	CMF(n) |↗ x1 || CMF(n) > x2	1	2	43	5	2	-90	90	5		12,321
	CMF(n) |↗ x1 || CMF(n) |↘ x2	1	2	43	5	2	-90	90	5		12,321
Chaikin Oscillator (CO)											***6,174***
	CO(n1,n2) > x	2	2	43	1					x=0	1,764
	CO(n1,n2) > S(n3)	3	2	43	2					$n_2 \leq 14$	3,087
	CO(n1,n2) > D(n3)	3	2	43	2					$n_2 \leq 7$	1,323
Chande Momentum Oscillator (CMO)											***27,969***
	CMO(n) > x	1	2	43	3	1	-95	95	5		546
	CMO(n1) > S(n2)	2	2	43	5					$n_2 \leq 14$	27

CMO(n1) > D(n2)	2	2	43	5					$n_2 \le 7$	18
CMO(n) |↗ x1 || CMO(n) > x2	1	2	43	5	2	-95	95	5		13,689
CMO(n) |↗ x1 || CMO(n) |↘ x2	1	2	43	5	2	-95	95	5		13,689
Commodity Channel Index (CCI)										***616***
CCI(n) > x	1	2	43	3	1	-90	90	5		518
CCI(n1) > S(n2)	2	2	43	3					$n_2 \le 14$	70
CCI(n1) > D(n2)	2	2	43	3					$n_2 \le 7$	28
Demand Index (DI)										***25,258***
DI(n) > x	1	2	43	3	1	-90	90	5		518
DI(n1) > S(n2)	2	2	43	3					$n_2 \le 14$	70
DI(n1) > D(n2)	2	2	43	3					$n_2 \le 7$	28
DI(n) |↗ x1 || DI(n) > x2	1	2	43	5	2	-90	90	5		12,321
DI(n) |↗ x1 || DI(n) |↘ x2	1	2	43	5	2	-90	90	5		12,321
Detrended Price Oscillator (DPO)										***672***
DPO(n) > x	1	2	43	3	1	-20	20	1		574
DPO(n1) > S(n2)	2	2	43	3					$n_2 \le 14$	70
DPO(n1) > D(n2)	2	2	43	3					$n_2 \le 7$	28
Dynamic Momentum Index (DYMOI)										***37,584***
DYMOI(n1,n2,n3) > x	3	2	43	7	1	10	90	10		1,944
DYMOI(n1,n2,n3) > S(n4)	4	2	43	7					$n_2 \le 14$	432
DYMOI(n1,n2,n3) > D(n4)	4	2	43	7					$n_2 \le 7$	216
DYMOI(n1,n2,n3) |↗ x1 || DYMOI(n1,n2,n3) > x2	3	2	43	7	2	10	90	10		17,496
DYMOI(n1,n2,n3) |↗ x1 || DYMOI(n1,n2,n3) |↘ x2	3	2	43	7	2	10	90	10		17.496
Inertia Indicator (INI)										***22,464***
INI(n1,n2,n3) > x	3	2	43	5	1	30	70	4		8,019
INI(n1,n2,n3) > S(n4)	4	2	43	5					$n_2 \le 14$	2,187
INI(n1,n2,n3) > D(n4)	4	2	43	5					$n_2 \le 7$	1,458
INI(n1,n2,n3) |↗ x1 || INI(n1,n2,n3) > x2	3	2	43	7	2	30	70	10		5,400
INI(n1,n2,n3) |↗ x1 || INI(n1,n2,n3) |↘ x2	3	2	43	7	2	30	70	10		5,400
Kase Convergence Divergence (KCD)										***43,141***
KCD(n1,n2,n3,n4) > x	4	2	43	10	1	-90	90	10		11,875
KCD(n1,n2,n3,n4) > S(n5)	5	2	43	7					$n_2 \le 14$	2,592
KCD(n1,n2,n3,n4) > D(n5)	5	2	43	7					$n_2 \le 7$	1,296
KCD(n1,n2,n3,n4) |↗ x1 || KCD(n1,n2,n3,n4) > x2	4	2	43	15	2	-90	90	15		13,689
KCD(n1,n2,n3,n4) |↗ x1 || KCD(n1,n2,n3,n4) |↘ x2	4	2	43	15	2	-90	90	15		13,689
Kase Peak Oscillator (KPO)										***8,624***
KPO(n1,n2) > x	2	2	43	3	1	-180	180	10		7,252
KPO(n1,n2) > S(n3)	3	2	43	3					$n_2 \le 14$	980
KPO(n1,n2) > D(n3)	3	2	43	3					$n_2 \le 7$	392
Klinger Volume Oscillator (KVO)										***6,174***
KVO(n1,n2) > x	2	2	43	1					x=0	1,764
KVO(n1,n2) > S(n3)	3	2	43	2					$n_2 \le 14$	3,087
KVO(n1,n2) > D(n3)	3	2	43	2					$n_2 \le 7$	1,323
Know Sure Thing (KST)										***5,488***
KST(n1,n2) > x	2	2	43	3	1	-50	50	5		4,116
KST(n1,n2) > S(n3)	3	2	43	3					$n_2 \le 14$	980
KST(n1,n2) > D(n3)	3	2	43	3					$n_2 \le 7$	392
Linear Regression Slope (LRS)										***371***
LRS(n) > x	1	2	43	2	1	-30	30	5		273
LRS(n1) > S(n2)	2	2	43	3					$n_2 \le 14$	70

LRS(n1) > D(n2)	2	2	43	3					*$n_2 \le 7$*	28
Market Volume Impact (MVI)										***252***
MVI(n) > x	1	2	43	1					x=0	42
MVI(n1) > S(n2)	2	2	43	2					*$n_2 \le 14$*	147
MVI(n1) > D(n2)	2	2	43	2					*$n_2 \le 7$*	63
Money Flow Index (MFI)										***24,978***
MFI(n) > x	1	2	43	3	1	10	90	5		238
MFI(n1) > S(n2)	2	2	43	3					*$n_2 \le 14$*	70
MFI(n1) > D(n2)	2	2	43	3					*$n_2 \le 7$*	28
MFI(n) |↗ x1 || MFI(n) > x2	1	2	43	5	2	-90	90	5		12,321
MFI(n) |↗ x1 || MFI(n) |↘ x2	1	2	43	5	2	-90	90	5		12,321
Moving Average Convergence Divergence (MACD)										***4,704***
MACDp*(n1,n2) > x	2	2	43	3	1	-16	16	2		3,332
MACD(n1,n2) > S(n3)	3	2	43	3					n3≤14	980
MACD(n1,n2) > D(n3)	3	2	43	3					n3≤7	392
New Relative Volatility Index (NRVI)										***30,331***
NRVI(n1,n2) > x	2	2	43	3	1	20	80	5		2,548
NRVI(n1,n2) > S(n3)	3	2	43	5					*$n_2 \le 14$*	243
NRVI(n1,n2) > D(n3)	3	2	43	5					*$n_2 \le 7$*	162
NRVI(n1,n2) |↗ x1 || NRVI(n1,n2) > x2	2	2	43	5	2	20	80	5		13,689
NRVI(n1,n2) |↗ x1 || NRVI(n1,n2) |↘ x2	2	2	43	5	2	20	80	5		13,689
On Balance Volume (OBV)										***210***
OBV(n1) > S(n2)	2	2	43	2						147
OBV(n1) > D(n2)	2	2	43	2						63
PI Opinion Oscillator (PI)										***7,107***
PI(n) > x	1	2	43	2	1	5	95	5		399
PI(n1) > S(n2)	2	2	43	2					*$n_2 \le 14$*	147
PI(n1) > D(n2)	2	2	43	2					*$n_2 \le 7$*	63
PI(n) |↗ x1 || PI(n) > x2	1	2	43	5	2	5	95	5		3,249
PI(n) |↗ x1 || PI(n) |↘ x2	1	2	43	5	2	5	95	5		3,249
Polarized Fractal Efficiency (PFE)										***60,426***
PFE(n1,n2) > x	2	2	43	5	1	-90	90	10		1,539
PFE(n1,n2) > S(n3)	3	2	43	5					*$n_2 \le 14$*	243
PFE(n1,n2) > D(n3)	3	2	43	5					*$n_2 \le 7$*	162
PFE(n1,n2) |↗ x1 || PFE(n1,n2) > x2	2	2	43	5	2	-90	90	10		29,241
PFE(n1,n2) |↗ x1 || PFE(n1,n2) |↘ x2	2	2	43	5	2	-90	90	10		29,241
Random Walk Index (RWI) for High prices										***450***
RWI_High(n1) > RWI_Low(n2)	2	2	19	1						324
RWI_High(n1) > S(n2)	2	2	19	1					*$n_2 \le 6$*	90
RWI_High(n1) > D(n2)	2	2	19	1					*$n_2 \le 3$*	36
Rate of Change (ROC)										***672***
ROC(n) > x	1	2	43	3	1	-0.2	0.2	0.01		574
ROC(n1) > S(n2)	2	2	43	3					*$n_2 \le 14$*	70
ROC(n1) > D(n2)	2	2	43	3					*$n_2 \le 7$*	28
Relative Momentum Index (RMI)										***48,600***
RMI(n1,n2) > x	2	2	43	5	1	10	90	5		1,377
RMI(n1,n2) > S(n3)	3	2	43	5					*$n_2 \le 14$*	243
RMI(n1,n2) > D(n3)	3	2	43	5					*$n_2 \le 7$*	162
RMI(n1,n2) |↗ x1 || RMI(n1,n2) > x2	2	2	43	5	2	10	90	5		23,409
RMI(n1,n2) |↗ x1 || RMI(n1,n2) |↘ x2	2	2	43	5	2	10	90	5		23,409

Relative Strength Index (RSI)										***10,864***
RSI(n) > x	1	2	43	1	1	4	96	2		1,974
RSI(n1) > S(n2)	2	2	43	1					$n_2 \le 14$	546
RSI(n1) > D(n2)	2	2	43	1					$n_2 \le 7$	252
RSI(n) |↗ x1 || RSI(n) > x2	1	2	43	3	2	10	90	5		4,046
RSI(n) |↗ x1 || RSI(n) |↘ x2	1	2	43	3	2	10	90	5		4,046
Relative Vigor Index (RVig)										***60,426***
Rvig(n1,n2) > x	2	2	43	5	1	-90	90	10		1,539
Rvig(n1,n2) > S(n3)	3	2	43	5					$n_2 \le 14$	243
Rvig(n1,n2) > D(n3)	3	2	43	5					$n_2 \le 7$	162
Rvig(n1,n2) |↗ x1 || Rvig(n1,n2) > x2	2	2	43	5	2	-90	90	10		29,241
Rvig(n1,n2) |↗ x1 || Rvig(n1,n2) |↘ x2	2	2	43	5	2	-90	90	10		29,241
Relative Volatility Index (RVI)										***16,859***
RVI(n1,n2) > x	2	2	43	3	1	10	90	5		3,332
RVI(n1,n2) > S(n3)	3	2	43	5					$n_2 \le 14$	243
RVI(n1,n2) > D(n3)	3	2	43	5					$n_2 \le 7$	162
RVI(n1,n2) |↗ x1 || RVI(n1,n2) > x2	2	2	43	5	2	10	90	10		6,561
RVI(n1,n2) |↗ x1 || RVI(n1,n2) |↘ x2	2	2	43	5	2	10	90	10		6,561
Stochastic Momentum Index (SMI)										***33,250***
SMI(n1,n2,n3) > x	3	2	43	10	1	-50	50	5		2,625
SMI(n1,n2,n3) > S(n4)	4	2	43	10					$n_2 \le 14$	250
SMI(n1,n2,n3) > D(n4)	4	2	43	10					$n_2 \le 7$	125
SMI(n1,n2,n3) |↗ x1 || SMI(n1,n2,n3) > x2	3	2	43	10	2	-50	50	10		15,125
SMI(n1,n2,n3) |↗ x1 || SMI(n1,n2,n3) |↘ x2	3	2	43	10	2	-50	50	10		15,125
Stochastic Oscillator (%K)										***1,769***
%k(n) > x	1	2	43	3	1	5	95	5		266
%k(n1) > S(n2)	2	2	43	5					$n_2 \le 14$	27
%k(n1) > D(n2)	2	2	43	5					$n_2 \le 7$	18
%k(n) |↗ x1 || %k(n) > x2	1	2	43	5	2	10	90	10		729
%k(n) |↗ x1 || %k(n) |↘ x2	1	2	43	5	2	10	90	10		729
Stochastic RSI Oscillator (SRSI)										***16,859***
SRSI(n1,n2) > x	2	2	43	3	1	10	90	5		3,332
SRSI(n1,n2) > S(n3)	3	2	43	5					$n_2 \le 14$	243
SRSI(n1,n2) > D(n3)	3	2	43	5					$n_2 \le 7$	162
SRSI(n1,n2) |↗ x1 || SRSI(n1,n2) > x2	2	2	43	5	2	10	90	10		6,561
SRSI(n1,n2) |↗ x1 || SRSI(n1,n2) |↘ x2	2	2	43	5	2	10	90	10		6,561
The Quantitative Candlestick (Qstick)										***840***
Qstick(n) > x	1	2	43	1					x=0	42
Qstick(n1) > S(n2)	2	2	43	1					$n_2 \le 14$	546
Qstick(n1) > D(n2)	2	2	43	1					$n_2 \le 7$	252
Triple Exponential Smoothing (TRIX)										***3,402***
TRIX(n1,n2) > x	2	2	43	5	1	-90	90	5		2,997
TRIX(n1,n2) > S(n3)	3	2	43	5					$n_2 \le 14$	243
TRIX(n1,n2) > D(n3)	3	2	43	5					$n_2 \le 7$	162
True Strength Index (TSI)										***60,426***
TSI(n1,n2) > x	2	2	43	5	1	-90	90	10		1,539
TSI(n1,n2) > S(n3)	3	2	43	5					$n_2 \le 14$	243
TSI(n1,n2) > D(n3)	3	2	43	5					$n_2 \le 7$	162
TSI(n1,n2) |↗ x1 || TSI(n1,n2) > x2	2	2	43	5	2	-90	90	10		29,241
TSI(n1,n2) |↗ x1 || TSI(n1,n2) |↘ x2	2	2	43	5	2	-90	90	10		29,241

Ultimate Oscillator (UO)											***22,842***
	UO(n1,n2,n3) > x	3	2	43	7	1	10	90	10		1,944
	UO(n1,n2,n3) > S(n4)	4	2	43	7					$n_2 \leq 14$	432
	UO(n1,n2,n3) > D(n4)	4	2	43	7					$n_2 \leq 7$	216
	UO(n1,n2,n3) |↗ x1 ‖ UO(n1,n2,n3) > x2	3	2	43	10	2	10	90	10		10,125
	UO(n1,n2,n3) |↗ x1 ‖ UO(n1,n2,n3) |↘ x2	3	2	43	10	2	10	90	10		10,125
Vortex Oscillator (VX)											***7,114***
	VX(n) > x	1	2	43	3	1	-90	90	5		518
	VX(n1) > S(n2)	2	2	43	3					$n_2 \leq 14$	70
	VX(n1) > D(n2)	2	2	43	3					$n_2 \leq 7$	28
	VX(n) |↗ x1 ‖ VX(n) > x2	1	2	43	5	2	-90	90	10		3,249
	VX(n) |↗ x1 ‖ VX(n) |↘ x2	1	2	43	5	2	-90	90	10		3,249
Williams Variable Accumulation Distribution (WVAD)											***210***
	WVAD(n1) > S(n2)	2	2	43	2					$n_2 \leq 14$	147
	WVAD(n1) > D(n2)	2	2	43	2					$n_2 \leq 7$	63

NOTE: The columns in this table present the following information on the construction of TTRs:

Column 1: Indicator name and symbol.

Column 2: Trading rule templates.

Columns 3-6: Parameterization rules for window length parameters:

Column 3: the number of window-length parameters associated with each the trading rule template;

Column 4: the minimum value for this type of parameter.

Column 5: the maximum value for this type of parameter.

Column 6: the increment value for this type of parameter.

Columns 7-10: Parameterization rules for threshold parameters (similar to the window-length parameters in columns 3-6).

Column 11: Additional restrictions.

Column 12: The total number (sum) of trading rules constructed after applying the parameterization procedure for each template trading rule or indicator. Aggregate values at the indicator level are additionally underlined.

**MACDp represents the Moving Average Convergence Divergence (MACD) indicator expressed as a relative difference between the two moving average components. Specifically, if STMA designates the short-term moving average and LTMA designates the long-term one, then MACDp = (STMA – LTMA) / LTMA. For comparison, in its standard form, MACD = STMA – LTMA.*

References

Achelis, S. B., 2001. Technical Analysis from A to Z. New York, McGraw Hill. Available at: http://freetradingdownloads.com/Technical%20Analysis%20from%20A%20to%20Z.pdf (accessed January 24, 2017).

Alexander, S. S., 1961. Price movements in speculative markets: Trends or random walks. Industrial Management Review, 2, 7-27.

Alexander, S. S., 1964. Price Movements in Speculative Markets: Trends or Random Walks. Industrial Management Review, 5, 25-46.

Altman, R., 1993). Relative Momentum Index: Modifying Rsi. Technical Analysis of Stocks & Commodities, 11(2), 57-60.

Appel, G., 1999. Technical Analysis Power Tools for Active Investors. Financial Times Prentice Hall.

Arms, R. W., 1983. Volume cycles in the stock market: market timing through equivolume charting. Dow Jones-Irwin.

Aspray, T.E., 1986. Fine-tuning the demand index. Technical Analysis of Stocks & Commodities, 4(4), 141-143.

Avramov, D., Kaplanski, G., Levy, H., 2018. Talking Numbers: Technical versus fundamental investment recommendations. Journal of Banking and Finance, 92, 100-114.

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106(3), 473-491.

Blau, W., 1991. True Strength Index. Technical Analysis of Stocks & Commodities, 9(11), 438-446.

Blau, W., 1993. Stochastic Momentum. Technical Analysis of Stocks & Commodities, 11(1), 11-18.

Bollinger, J., 2001. Bollinger on Bollinger bands. McGraw Hill Professional.

Botes, E., Siepman, D., 2010. The Vortex Indicator. Technical Analysis of Stocks & Commodities, 28(1), 20-30.

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance, 47(5), 1731-1764.

Chaikin, M., 1994. Chatting With Marc Chaikin. Technical Analysis of Stocks & Commodities, 12(1), 30-37.

Chande, T., 1995. The Time Price Oscillator. Technical Analysis of Stocks & Commodities, 13(9), 369-374.

Chande, T. S., Kroll, S., 1994. The New Technical Trader. John Wiley & Sons.

Colby, R.W., 2002. The Encyclopedia Of Technical Market Indicators, Second Edition. McGraw-Hill.

Dormeier, B. P., 2011. Investing with Volume Analysis. Identify, Follow, and Profit from Trends. FT Press.

Dorsey, D.G., 1993. The Relative Volatility Index. Technical Analysis of Stocks & Commodities, 11(6), 253-256.

Dorsey, D.G., 1995. Refining the Relative Volatility Index. Technical Analysis of Stocks & Commodities, 13(9), 388-391.

Ehlers, J.F., 2002a. Relative Vigor Index. Technical Analysis of Stocks & Commodities, 20(1), 16-20.

Ehlers, J.F., 2002b. The Center Of Gravity Oscillator. Technical Analysis of Stocks & Commodities, 20(5), 20-24.

Ehrlich, C.F., 2000. Using Oscillators With On-Balance Volume. Technical Analysis of Stocks & Commodities, 18(9), 22-29.

Faber, B.R., 1994. The Rate Of Change Indicator. Technical Analysis of Stocks & Commodities, 12(10), 403-405.

Hannula, H., 1994. Polarized Fractal Efficiency. Technical Analysis of Stocks & Commodities, 12(1), 38-41.

Hayashi, F., 2000. Econometrics. Princeton University Press.

Hudson, R., McGroarty, F., Urquhart, A., 2017. Sampling frequency and the performance of different types of technical trading rules. Finance Research Letters, 22, 136-139.

Hutson, J.K., 1983. Good Trix. Technical Analysis of Stocks & Commodities, 1(5), 105-108.

Kase, C. A., 1996. Trading with the odds: using the power of probability to profit in the futures market. Irwin Professional Publishing.

Klinger, S.J., 1994. The Klinger Volume Oscillator (KVO): A Theoretical Model. Journal of Technical Analysis, 44, 45-52.

Klinger, S.J., 1995. Identifying Trends with Volume Analysis. Technical Analysis of Stocks & Commodities, 15(12), 556-560.

Lambert, R.D., 1982. Commodity Channel Index: Tool for Trading Cyclic Trends. Volume. Technical Analysis of Stocks & Commodities, 1(5), 120-122.

Lane, G.C., 1984. Lane's Stochastics. Technical Analysis of Stocks & Commodities, 2(3), 87-90.

Lin, Q., 2018. Technical analysis and stock return predictability: An aligned approach. Journal of Financial Markets, 38, 103-123.

Livshin, I., 2001. Balance Of Market Power. Technical Analysis of Stocks & Commodities, 19(8), 18-32.

Macek, A.J., 1993. Combining Volume And Market Change. Technical Analysis of Stocks & Commodities, 11(4), 162-165.

Martin, P. G., McCann, B. B., 1989. The Investor's Guide to Fidelity Funds. John Wiley & Sons.

Park, C.H., Irwin, S.H., 2007. What do we know about the profitability of technical analysis?. Journal of Economic Surveys, 21(4), 786-826.

Poulos, E.M., 1991. Of Trends And Random Walks. Technical Analysis of Stocks & Commodities, 9(2), 49-52.

Pring, M. J., 1997. Martin Pring on market momentum. McGraw-Hill.

Pring, M.J., 1992. Rate of Change. Technical Analysis of Stocks & Commodities, 10(8), 325-327.

Quong, G., Soudack, A., 1989. Volume-weighted RSI: money flow. Technical Analysis of Stocks & Commodities, 7(3), 76-77.

Schirding, H., 1984. Stochastic Oscillator. Technical Analysis of Stocks & Commodities, 2(3), 94-97.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. The Journal of Finance, 54(5), 1647-1691.

Sweeney, J., 1997. On-Balance Volume. Technical Analysis of Stocks & Commodities, 15(10), 468-471.

Wilder, J. W., 1978. New concepts in technical trading systems. Trend Research.

Williams, L., 1985. The Ultimate Oscillator. Technical Analysis of Stocks & Commodities, 3(4), 140-141.

Williams, L.R., 1986. The Secrets of Selecting Stocks for Immediate and Substantial Gains. Windsor Books; 2 edition.

Appendix B. False discoveries and data snooping bias in RC tests estimated via Monte Carlo Simulation

Table B1. Sample length of 1 month

Significance level:	α = 10%						α = 5%						α = 1%					
σ No. TTRs	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40
1024	6.03	6.84	7.22	7.32	7.65	7.82	3.86	4.80	5.05	5.15	5.41	5.72	1.56	2.13	2.49	2.41	2.52	2.77
2048	2.90	3.49	3.82	3.75	3.84	4.20	1.58	2.15	2.41	2.41	2.34	2.69	0.60	0.83	0.84	1.11	0.98	1.17
4096	2.14	2.68	2.91	2.88	2.87	3.35	1.14	1.61	1.61	1.68	1.67	2.01	0.40	0.59	0.58	0.85	0.68	0.78
8192	2.01	2.54	2.53	2.41	2.53	2.78	1.05	1.40	1.44	1.48	1.34	1.66	0.36	0.50	0.49	0.68	0.56	0.57
16384	1.96	2.45	2.48	2.27	2.36	2.57	1.03	1.37	1.37	1.41	1.24	1.52	0.36	0.50	0.48	0.64	0.53	0.55
32768	1.92	2.38	2.38	2.21	2.30	2.54	0.97	1.36	1.37	1.35	1.22	1.44	0.36	0.47	0.47	0.62	0.51	0.55
65536	1.71	2.26	2.24	2.13	2.16	2.47	0.83	1.28	1.25	1.28	1.16	1.37	0.31	0.43	0.43	0.57	0.46	0.53
131072	1.68	2.24	2.20	2.10	2.15	2.47	0.83	1.27	1.23	1.27	1.16	1.35	0.30	0.40	0.42	0.54	0.46	0.52
262144	0.74	1.25	1.36	1.39	1.49	1.84	0.39	0.62	0.66	0.82	0.84	0.89	0.14	0.20	0.23	0.34	0.30	0.35
524288	0.45	0.85	0.89	1.09	1.10	1.31	0.27	0.45	0.45	0.61	0.64	0.66	0.06	0.15	0.18	0.22	0.23	0.25
688896	0.39	0.68	0.70	0.92	0.98	1.12	0.22	0.35	0.41	0.51	0.53	0.57	0.06	0.12	0.14	0.15	0.22	0.22
Absolute data snooping bias*																		
Minimum	0.06	0.17	0.19	0.17	0.12	0.19	0.05	0.10	0.04	0.10	0.11	0.09	0.00	0.03	0.04	0.07	0.01	0.03
Average	1.76	2.02	2.10	1.84	1.87	2.02	0.98	1.28	1.27	1.24	1.17	1.36	0.39	0.50	0.52	0.65	0.50	0.58
Maximum	5.64	6.16	6.52	6.40	6.67	6.70	3.64	4.45	4.64	4.64	4.88	5.15	1.50	2.01	2.35	2.26	2.30	2.55
Relative data snooping bias**																		
Minimum	15%	25%	27%	18%	12%	17%	23%	29%	10%	20%	21%	16%	0%	25%	29%	47%	5%	14%
Average	452%	297%	300%	199%	190%	180%	443%	366%	311%	242%	221%	239%	642%	417%	372%	432%	229%	265%
Maximum	1446%	906%	931%	696%	681%	598%	1655%	1271%	1132%	910%	921%	904%	2500%	1675%	1679%	1507%	1045%	1159%

Note: This table reports the number of false discoveries per 100 tests, aggregated by the size of rule the universe, the volatility of the data generating process and significance level for the RC test.

**Absolute data snooping bias denotes the average difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe; it is expressed as a number per 100 simulations, i.e. percentage points.*

***Relative data snooping bias denotes the average percentage difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe. It is calculated as the absolute bias divided by the number of false discoveries obtained for the benchmark universe.*

Table B2. Sample length of 1 quarter

Significance level:	α = 10%						α = 5%						α = 1%					
σ / No. TTRs	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40
1024	6.80	7.40	7.37	7.70	7.78	8.09	4.39	5.09	5.32	5.53	5.84	5.89	1.65	2.51	2.46	2.63	2.71	2.81
2048	3.12	3.81	3.98	4.31	4.48	4.48	1.80	2.44	2.47	2.67	2.58	2.86	0.61	0.94	1.08	1.02	1.07	1.07
4096	2.15	2.85	3.12	3.20	3.34	3.42	1.19	1.70	1.73	1.93	1.85	2.08	0.40	0.62	0.62	0.68	0.72	0.73
8192	1.91	2.46	2.46	2.68	2.58	2.83	1.04	1.43	1.44	1.47	1.46	1.56	0.34	0.54	0.49	0.55	0.54	0.57
16384	1.84	2.34	2.29	2.57	2.33	2.42	0.99	1.38	1.32	1.32	1.30	1.40	0.32	0.53	0.44	0.50	0.49	0.51
32768	1.80	2.29	2.27	2.49	2.18	2.36	0.96	1.34	1.28	1.28	1.26	1.39	0.32	0.50	0.44	0.47	0.47	0.50
65536	1.71	2.23	2.19	2.40	2.11	2.30	0.91	1.29	1.22	1.25	1.20	1.36	0.31	0.48	0.43	0.46	0.46	0.49
131072	1.71	2.21	2.19	2.38	2.10	2.29	0.90	1.29	1.22	1.24	1.20	1.36	0.31	0.48	0.43	0.46	0.46	0.49
262144	0.92	1.42	1.60	1.59	1.47	1.72	0.40	0.74	0.77	0.89	0.77	0.95	0.14	0.20	0.22	0.31	0.34	0.33
524288	0.53	1.00	0.92	1.12	1.06	1.26	0.30	0.47	0.51	0.63	0.55	0.69	0.11	0.10	0.14	0.19	0.21	0.27
688896	0.39	0.75	0.76	0.94	0.96	1.03	0.22	0.36	0.35	0.47	0.51	0.58	0.06	0.08	0.12	0.16	0.17	0.21
Absolute data snooping bias*																		
Minimum	0.14	0.25	0.16	0.18	0.10	0.23	0.08	0.11	0.16	0.16	0.04	0.11	0.05	0.02	0.02	0.03	0.04	0.06
Average	1.86	2.05	2.08	2.10	1.98	2.09	1.07	1.36	1.38	1.35	1.29	1.37	0.39	0.61	0.56	0.57	0.58	0.57
Maximum	6.41	6.65	6.61	6.76	6.82	7.06	4.17	4.73	4.97	5.06	5.33	5.31	1.59	2.43	2.34	2.47	2.54	2.60
Relative data snooping bias**																		
Minimum	36%	33%	21%	19%	10%	22%	36%	31%	46%	34%	8%	19%	83%	25%	17%	19%	24%	29%
Average	477%	273%	274%	224%	207%	203%	485%	377%	394%	287%	253%	237%	652%	763%	463%	354%	339%	270%
Maximum	1644%	887%	870%	719%	710%	685%	1895%	1314%	1420%	1077%	1045%	916%	2650%	3038%	1950%	1544%	1494%	1238%

Note: This table reports the number of false discoveries per 100 tests, aggregated by the size of rule the universe, the volatility of the data generating process and significance level for the RC test.

**Absolute data snooping bias denotes the average difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe; it is expressed as a number per 100 simulations, i.e. percentage points.*

***Relative data snooping bias denotes the average percentage difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe. It is calculated as the absolute bias divided by the number of false discoveries obtained for the benchmark universe.*

Table B3. Sample length of 1 year

Significance level:	α = 10%						α = 5%						α = 1%					
σ / No. TTRs	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40
1024	7.83	7.80	7.82	8.30	8.32	8.27	5.79	5.56	5.77	6.16	6.44	6.38	2.66	2.64	2.91	2.94	3.25	3.16
2048	4.53	4.35	4.47	4.64	5.04	5.00	2.59	2.79	3.05	3.19	3.26	3.19	1.11	0.99	1.04	1.47	1.32	1.52
4096	3.29	3.28	3.37	3.76	3.94	3.98	1.94	2.06	2.21	2.41	2.52	2.60	0.72	0.67	0.79	1.04	0.91	1.07
8192	2.62	2.66	2.80	3.17	3.24	3.44	1.49	1.56	1.75	1.94	1.93	2.16	0.54	0.49	0.57	0.73	0.68	0.88
16384	2.43	2.35	2.41	2.72	2.72	2.89	1.30	1.36	1.45	1.67	1.57	1.84	0.47	0.44	0.49	0.56	0.52	0.69
32768	2.30	2.20	2.33	2.56	2.42	2.61	1.25	1.25	1.34	1.47	1.47	1.61	0.45	0.42	0.48	0.49	0.50	0.60
65536	2.15	2.12	2.27	2.48	2.36	2.54	1.17	1.20	1.30	1.42	1.41	1.50	0.42	0.40	0.48	0.47	0.48	0.57
131072	2.15	2.12	2.27	2.48	2.36	2.54	1.17	1.20	1.30	1.42	1.41	1.50	0.42	0.40	0.48	0.47	0.48	0.57
262144	1.32	1.48	1.64	1.87	1.74	1.99	0.63	0.74	0.90	0.93	0.94	1.13	0.18	0.26	0.31	0.28	0.32	0.39
524288	0.69	0.87	1.04	1.24	1.21	1.29	0.42	0.45	0.58	0.64	0.61	0.85	0.13	0.17	0.24	0.20	0.20	0.25
688896	0.56	0.70	0.87	0.98	0.88	1.08	0.33	0.30	0.49	0.47	0.51	0.65	0.07	0.14	0.19	0.14	0.13	0.18
Absolute data snooping bias*																		
Minimum	0.13	0.17	0.17	0.26	0.33	0.21	0.09	0.15	0.09	0.17	0.10	0.20	0.06	0.03	0.05	0.06	0.07	0.07
Average	2.37	2.22	2.17	2.34	2.46	2.38	1.45	1.52	1.48	1.66	1.65	1.63	0.64	0.55	0.59	0.73	0.74	0.79
Maximum	7.27	7.10	6.95	7.32	7.44	7.19	5.46	5.26	5.28	5.69	5.93	5.73	2.59	2.50	2.72	2.80	3.12	2.98
Relative data snooping bias**																		
Minimum	23%	24%	20%	27%	38%	19%	27%	50%	18%	36%	20%	31%	86%	21%	26%	43%	54%	39%
Average	423%	318%	250%	239%	279%	220%	438%	506%	301%	352%	323%	250%	914%	391%	310%	518%	566%	439%
Maximum	1298%	1014%	799%	747%	845%	666%	1655%	1753%	1078%	1211%	1163%	882%	3700%	1786%	1432%	2000%	2400%	1656%

Note: This table reports the number of false discoveries per 100 tests, aggregated by the size of rule the universe, the volatility of the data generating process and significance level for the RC test.

**Absolute data snooping bias denotes the average difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe; it is expressed as a number per 100 simulations, i.e. percentage points.*

***Relative data snooping bias denotes the average percentage difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe. It is calculated as the absolute bias divided by the number of false discoveries obtained for the benchmark universe.*

Table B4. Sample length of 4 years

Significance level:	α = 10%						α = 5%						α = 1%					
σ No. TTRs	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40	0.15	0.20	0.25	0.30	0.35	0.40
1024	8.04	8.31	8.26	8.40	8.49	8.76	5.96	6.21	6.48	6.70	6.76	7.00	2.83	3.28	3.37	3.47	3.55	3.61
2048	4.68	5.01	5.17	5.39	5.57	5.67	2.93	3.31	3.54	3.42	3.74	3.60	1.10	1.26	1.35	1.56	1.72	1.61
4096	3.54	3.83	4.12	4.17	4.55	4.59	2.10	2.30	2.57	2.62	2.84	2.87	0.70	0.95	0.96	1.13	1.30	1.11
8192	2.89	3.16	3.42	3.28	3.69	3.79	1.58	1.81	1.94	2.19	2.30	2.30	0.52	0.74	0.71	0.84	0.99	0.89
16384	2.41	2.69	2.83	2.86	3.14	3.12	1.28	1.47	1.67	1.79	1.97	1.93	0.43	0.57	0.56	0.61	0.72	0.72
32768	2.26	2.37	2.52	2.55	2.80	2.73	1.18	1.30	1.42	1.48	1.69	1.58	0.38	0.48	0.48	0.51	0.55	0.57
65536	2.09	2.25	2.43	2.39	2.62	2.55	1.13	1.23	1.36	1.39	1.59	1.49	0.36	0.46	0.45	0.47	0.52	0.53
131072	2.09	2.25	2.43	2.39	2.62	2.55	1.13	1.23	1.36	1.39	1.59	1.49	0.36	0.46	0.45	0.47	0.52	0.53
262144	1.13	1.52	1.73	1.83	2.15	2.01	0.63	0.77	0.89	1.04	1.16	1.13	0.21	0.26	0.30	0.28	0.37	0.41
524288	0.66	0.85	1.15	1.27	1.38	1.36	0.39	0.53	0.66	0.69	0.76	0.76	0.14	0.14	0.17	0.14	0.23	0.27
688896	0.49	0.67	0.93	1.08	1.13	1.09	0.27	0.36	0.49	0.51	0.51	0.63	0.07	0.08	0.11	0.11	0.15	0.18
Absolute data snooping bias*																		
Minimum	0.17	0.18	0.22	0.19	0.25	0.27	0.12	0.17	0.17	0.18	0.25	0.13	0.07	0.06	0.06	0.03	0.08	0.09
Average	2.49	2.55	2.48	2.37	2.57	2.62	1.56	1.66	1.70	1.76	1.93	1.79	0.63	0.78	0.77	0.84	0.90	0.85
Maximum	7.55	7.64	7.33	7.32	7.36	7.67	5.69	5.85	5.99	6.19	6.25	6.37	2.76	3.20	3.26	3.36	3.40	3.43
Relative data snooping bias**																		
Minimum	35%	27%	24%	18%	22%	25%	44%	47%	35%	35%	49%	21%	100%	75%	55%	27%	53%	50%
Average	508%	381%	266%	220%	228%	241%	578%	460%	347%	345%	378%	283%	904%	975%	700%	762%	598%	469%
Maximum	1541%	1140%	788%	678%	651%	704%	2107%	1625%	1222%	1214%	1225%	1011%	3943%	4000%	2964%	3055%	2267%	1906%

Note: This table reports the number of false discoveries per 100 tests, aggregated by the size of rule the universe, the volatility of the data generating process and significance level for the RC test.

**Absolute data snooping bias denotes the average difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe; it is expressed as a number per 100 simulations, i.e. percentage points.*

***Relative data snooping bias denotes the average percentage difference in the number of false discoveries between the 10 restricted (smaller) rule universes and the benchmark rule universe. It is calculated as the absolute bias divided by the number of false discoveries obtained for the benchmark universe.*

Appendix C. Overview of data sample used in the empirical analysis

Table C1. Summary

Market symbol	Country (Countries)	No. selected companies	First day	Last day	Total days (observations)	Average days per company
AE	United Arab Emirates	40	15.11.2000	14.11.2013	63,041	1,576
AR	Argentina	40	11.04.1990	13.11.2013	139,421	3,486
AT	Austria	20	02.01.1987	13.11.2013	84,396	4,220
AU	Australia	40	02.01.1981	14.11.2013	218,022	5,451
BA	Bosnia And Herzegovina	20	09.07.2002	13.11.2013	14,522	726
BE	Belgium	20	02.01.1987	13.11.2013	90,858	4,543
BG	Bulgaria	15	27.05.1998	14.11.2013	31,327	2,088
BH	Bahrain	40	06.06.1995	12.11.2013	47,333	1,183
BR	Brazil	40	19.03.1993	13.11.2013	143,404	3,585
BRVM	Benin, Burkina Faso, Guinea Bissau, Côte d'Ivoire, Mali, Niger, Senegal, Togo	36	29.12.2006	13.11.2013	21,394	594
CA	Canada	40	17.03.1980	13.11.2013	269,234	6,731
CH	Switzerland	20	05.01.1987	14.11.2013	99,041	4,952
CL	Chile	40	27.09.1993	13.11.2013	134,396	3,360
CN	China	40	02.01.1991	14.11.2013	84,580	2,115
CO	Colombia	20	09.02.1995	13.11.2013	38,934	1,947
CY	Cyprus	19	15.12.1994	13.11.2013	46,014	2,422
CZ	Czech Republic	13	10.01.1996	14.11.2013	34,075	2,621
DE	Germany	40	05.04.1991	14.11.2013	169,714	4,243
DK	Denmark	20	02.01.1985	14.11.2013	103,631	5,182
EE	Estonia	16	06.09.1996	14.11.2013	33,176	2,074
EG	Egypt	30	06.07.1993	13.11.2013	76,014	2,534
ES	Spain	35	02.01.1990	14.11.2013	148,755	4,250
FI	Finland	25	04.01.1988	14.11.2013	109,865	4,395
FR	France	40	07.01.1985	14.11.2013	235,149	5,879
GR	Greece	40	02.01.1991	13.11.2013	157,728	3,943
HK	Hong Kong	40	02.01.1980	14.11.2013	186,807	4,670
HR	Croatia	23	30.06.1994	13.11.2013	49,235	2,141
HU	Hungary	13	23.12.1992	14.11.2013	45,871	3,529
ID	Indonesia	40	04.01.1982	14.11.2013	128,820	3,221
IE	Ireland	40	09.11.1994	14.11.2013	109,985	2,750
IL	Israel	40	08.10.1992	14.11.2013	163,151	4,079
IN	India	30	01.01.1990	14.11.2013	134,296	4,477
IQ	Iraq	37	04.06.2008	13.11.2013	22,398	605
IS	Iceland	13	29.07.1992	13.11.2013	12,980	998
IT	Italy	40	02.01.1987	14.11.2013	166,499	4,162
JO	Jordan	40	22.11.1993	14.11.2013	109,470	2,737
JP	Japan	40	04.01.1984	14.11.2013	241,640	6,041
KE	Kenya	40	24.04.1995	14.11.2013	122,025	3,051
KR	South Korea	40	03.05.1983	14.11.2013	187,222	4,681
KW	Kuwait	40	03.01.1993	14.11.2013	124,355	3,109
KZ	Kazakhstan	7	27.10.1998	14.11.2013	8,844	1,263
LB	Lebanon	17	03.01.1993	13.11.2013	23,549	1,385
LK	Sri Lanka	40	01.06.1987	14.11.2013	126,926	3,173
LT	Lithuania	25	07.08.1997	14.11.2013	53,474	2,139
LV	Latvia	31	09.02.1998	14.11.2013	37,270	1,202
MA	Morocco	40	23.01.1998	13.11.2013	83,847	2,096
MU	Mauritius	40	25.10.1996	14.11.2013	170,310	4,258
MW	Malawi	8	10.07.2008	13.11.2013	3,160	395
MX	Mexico	35	02.01.1990	13.11.2013	117,463	3,356
MY	Malaysia	40	02.03.1984	14.11.2013	235,309	5,883
NA	Namibia	29	18.10.1996	13.11.2013	52,818	1,821
NG	Nigeria	40	20.04.1998	13.11.2013	51,206	1,280
NL	Netherlands	40	02.01.1985	14.11.2013	210,676	5,267
NO	Norway	40	03.01.1980	14.11.2013	162,526	4,063
NZ	New Zealand	40	16.06.1986	14.11.2013	133,777	3,344
OM	Oman	30	24.06.1995	14.11.2013	80,319	2,677
PE	Peru	38	02.03.1992	13.11.2013	105,446	2,775
PH	Philippines	30	02.01.1986	14.11.2013	120,718	4,024
PK	Pakistan	40	25.05.1994	13.11.2013	123,991	3,100
PL	Poland	40	13.08.1992	14.11.2013	92,865	2,322

PT	Portugal	40	05.01.1994	14.11.2013	130,597	3,265
QA	Qatar	40	12.12.1999	14.11.2013	93,066	2,327
RO	Romania	39	20.11.1995	14.11.2013	100,047	2,565
RS	Serbia	14	26.11.2003	14.11.2013	21,956	1,568
RU	Russian Federation	40	01.09.1995	14.11.2013	30,959	774
SA	Saudi Arabia	40	10.01.1999	14.11.2013	108,614	2,715
SE	Sweden	30	02.01.1987	14.11.2013	165,922	5,531
SG	Singapore	40	02.01.1979	14.11.2013	172,560	4,314
SI	Slovenia	7	09.04.1996	14.11.2013	20,862	2,980
SK	Slovakia	11	05.01.1994	13.11.2013	12,305	1,119
TH	Thailand	40	02.01.1981	14.11.2013	184,656	4,616
TN	Tunisia	40	24.06.1993	14.11.2013	104,468	2,612
TR	Turkey	40	02.01.1991	14.11.2013	160,842	4,021
TW	Taiwan	40	05.01.1981	14.11.2013	187,881	4,697
TZ	Tanzania	13	06.06.2000	13.11.2013	11,596	892
UA	Ukraine	20	15.10.1998	14.11.2013	32,461	1,623
UK	United Kingdom	40	02.01.1985	14.11.2013	211,146	5,279
US	United States	39	17.03.1980	13.11.2013	275,298	7,059
VE	Venezuela	11	12.08.1993	13.11.2013	24,658	2,242
VN	Vietnam	40	28.07.2000	14.11.2013	57,676	1,442
ZA	South Africa	40	07.04.1988	14.11.2013	166,196	4,155
	TOTAL	**2,579**	**02.01.1979**	**14.11.2013**	**8,667,038**	**3,361**