Lucky trading rules

Dan-Gabriel ANGHEL[1,2]

[1] *Institute for Economic Forecasting, Romanian Academy*

[2] *Department of Money and Banking, The Bucharest University of Economic Studies*

dan.anghel@fin.ase.ro

Current version: 12/21/2020

Abstract. We estimate the proportion of false discoveries caused by methodological choices and finite-sample limitations in multiple testing procedures concerned with the overperformance of speculative trading rules. Our main contribution is to show that a bearish tendency of lucky rules in downward trending markets explains the vast majority of false discoveries for a wide variety of testing conditions. This *bearish preference effect* dominates spurious correlation, which has previously been thought to be the main driver behind data snooping. We also find that methodological adjustments designed to control for this effect do not significantly decease false discoveries, showing that the associated data snooping bias should be very persistent in empirical tests. Among others, our results hint that evidence showing time-varying market inefficiencies, which center on important downtrends in the data, may in fact be biased. Also, they imply that the previously documented merit of speculative trading rules as a risk management aid in timing exit points around the onset of bear markets could be a product of luck rather than true economic relevance. The results obtained in a data-snooping-free investigation on the empirical performance of speculative trading rules in the 'less efficient' cryptocurrency market support this conclusion.

JEL Classification: C12, G11, G14

Keywords: Lucky Trading Strategies, Spurious Correlation, Data Snooping, False Discoveries, Efficient Market Hypothesis, Cryptocurrency Market

1. Introduction

The discussion on data snooping in the financial economics literature is not new (see, e.g., the seminal work of Cowles, 1933; or, more recently, Brock, Lakonishok, and LeBaron, 1992), yet the inner workings of how data snooping bias occurs are still unclear. For example, data snooping is typically associated with the spurious ability of trading strategies to forecast future market returns, but the actual correspondence between this property of lucky strategies and the amount of Type I Errors, false discoveries arising in empirical tests has not been tested so far. Also, besides the obvious effect of ignoring trading costs and the data snooping efforts of others, alternative explanations regarding how false discoveries arise have not been explored. In this paper, we test for alternative drivers of data snooping bias by investigating the factors–data sample properties, methodological choices, and properties of lucky trading rules–that cause or increase false discoveries in statistical tests that should in theory (asymptotically) handle data snooping. Additionally, we quantify the relative contribution of each factor to the aggregate data snooping bias. Our main contribution is to show that a bearish tendency of lucky trading rules in downward trending markets–which we call the *directional preference effect*–is the main driver of false discoveries, while spurious correlation is less important. Among others, this implies that the sources of data snooping are more diverse than previously recognized and that novel methodological approaches are required to adequately control for them in empirical tests that can only use finite data samples.

Although the discussion is not new, it has only recently gained significant traction (see, e.g., the presidential address of Harvey, 2017). Also, evidence showing how much data snooping we could expect when studying important topics in financial economics only recently emerged. For example, Dichtl et al. (2020) find that almost all equity premium forecasting strategies fail to beat the historical mean in out-of-sample tests after controlling for data snooping and accounting for transaction costs, supporting earlier evidence that point in the same direction (Welch and Goyal, 2008). Harvey and Liu (2020) use a new data-snooping-free methodology for investigating the forecasting ability of trading strategies based on stock fundamentals over future stock returns and obtain results that point toward a similar conclusion. McLean and Pontiff (2016), Linnainmaa and Roberts (2018), or Chordia, Goyal and Saretto (2020) further estimate that the proportion of false discoveries published in the related literature may be between 26% and 58%, showing that data snooping may severely bias our perception about equity risk factors or fund manager skill. Harvey, Liu, and Zhu (2016) argue that most claimed research findings in financial economics are likely false and suggest that a newly discovered factor needs to clear a much higher hurdle (t-ratio ≥ 3).

Such findings are enlightening, but are limited to only a few research questions. For example, they are not being applicable to other financial markets where factor models that describe price behavior are less clear, such as the emerging cryptocurrency market[1]. Also, the discussion regarding lucky trading strategies based solely on historical market prices lags behind, even though many strategies have been proposed and tested by applied researchers over the years (Park and Irwin, 2007; Nazário et al., 2017; Henrique, Sobreiro, and Kimura, 2019). Although some papers have examined the data-snooping-free performance of speculative trading rules (e.g., Sullivan, Timmermann, and White, 1999; Bajgrowicz and Scaillet, 2012; Taylor, 2014), these constitute a relative small proportion of the total number of manuscripts that have been published on this topic. Recently, Anghel (2020a) estimates that at least 50% of all tests concerned with trading rule overperformance may be biased when ignoring the data snooping efforts of others. However, while the results hint that test outcomes depend on data properties and methodological choices, a more in-depth analysis has not been performed. To the extent of our knowledge, no paper has performed a quantitative investigation on the exact contribution of each methodological choice or data sample property to false discoveries in the literature concerned with speculative trading rules.

Here, we fill this gap by investigating the factors that make technical trading rules[2] appear more profitable than they actually are and also search for adjustments to multiple testing procedures that would account for their luck. We show that spurious overperformance depends on trading rules properties such as (bearish) directional preference, the correlation between trading signals and market returns, and trading frequency, which are all functions of methodological choices and data sample properties. Three independent

[1] Some pricing models have been proposed for the cryptocurrency market (e.g., Liu and Tsyvinski, 2019; or Liu, Liang, and Cui, 2020) but the literature–similar to the market it follows–is in its infancy.
[2] Technical trading rules are speculative trading strategies that make decisions based solely on historical price information, with the ones derived from Technical Analysis (see, e.g., Colby, 2002) being the most common.

sources of data snooping are identified. First, the *spurious correlation effect (SCE)* can be defined as the tendency of trading rules to time the market purely by luck and is traditionally associated with data snooping. Among others, the *SCE* depends on the volatility of the data generating process and on the market participation of trading rules, which in itself is a non-trivial function of sample properties and testing conditions. Second, the *directional preference effect (DPE)* is a novel effect that can be defined as the tendency of trading rules to anticipate a market's overall tendency and perfectly balance the proportion between long and short positions, again purely by luck. The *DPE* is mainly a function of the average market return. Both the *SCE* and the *DPE* get stronger with the number of trading rules being tested, i.e. the amount of data snooping. Despite this, the *SCE* can be eliminated in multiple testing procedures by simply accounting for the trading rule set from which lucky rules were extracted, while the *DPE* is more persistent and difficult to eliminate. Third, the *trading costs effect (TCE)*, which can be defined as the tendency of trading rules to appear more profitable when trading costs and other market frictions are mischaracterized.

We further perform a decomposition of the false discoveries that arise in multiple testing procedures. These procedures are designed to control for data snooping, but the asymptotically-valid theory may fail when faced with finite-sample limitations and/or unreasonable implementation strategies. Here, we use a simulation exercise to investigate how different testing conditions relate to false discoveries. On the one hand, testing conditions are altered by specific researcher choices and we investigate (i) (not) incorporating trading costs, (ii) (not) enabling short positions[3], (iii) (not) standardizing the test statistic, and (iv) (not) accounting for the data snooping efforts of others. On the other hand, testing conditions also change with the properties of the data samples being considered, and we investigate how false discoveries relate to average returns, volatility, skewness, and kurtosis. The results show that 21% to 78% (48% on average) of null rejections in single hypothesis tests–i.e., tests that only use one luck trading rule and do not account for alternatives used by applied researchers–and 1% to 12% (5% on average) of null rejections in multiple hypothesis tests can be false, depending on testing conditions and sample properties. Neglecting

[3] In empirical analyses, this depends on possible short selling restrictions imposed by regulators and/or on the availability of borrowable assets.

trading fees and/or liquidity costs account for approximately 42% and 35% of false discoveries, respectively. Considering short trades when they are not functional or standardizing the test statistic mostly add to the bias, but their contribution depends on other testing conditions. Most importantly, we find that 60% to 100% of false discoveries arise in downward trading markets, implying that the *DPE* is the main determinant of data snooping bias.

Interestingly, the results also show that accounting for the data snooping efforts of others is effective at eliminating the *spurious correlation effect*, but less so with respect to the *directional preference effect*. In particular, we find that false discovery proportions obtained after fully accounting for alternative trading rules reach at most 0.7% in data samples exhibiting positive average returns, but vary between 2% and 24.2% in samples exhibiting negative average returns. We observe a similar result when testing naïve trading rules that randomly enter and exit the market, implying that the seemingly superior ability of speculative trading rules in downward trending markets arises purely by luck. Moreover, the *directional preference effect* is resilient to methodological adjustments such as bootstrapping from the excess return series of additional, suitably-constructed trading rules. Finally, we estimate that a t-statistic of 3.24 is required on average to adequately control for data snooping in tests performed at a 5% significance level (2.63 and 4.77 for the 1% and 10% levels, respectively). Because most applied researchers usually consider the more traditional 1.96 threshold (1.64 and 2.57 for the 1% and 10% levels, respectively), this result complements recent evidence (e.g., Harvey and Liu, 2020; Chordia, Goyal and Saretto, 2020) showing that false discoveries may abound in the literature. Overall, our findings add to the conclusion that accounting for lucky trading strategies requires raising the statistical significance threshold to well above $t=3$.

Third, we perform an empirical investigation of the overperformance of technical trading rules in the cryptocurrency market. Besides the recent surge in the number of papers that analyze this relatively young market (Jiang, Li, and Wang, 2020), the study of cryptocurrencies is relevant in our context because they are seemingly "less efficient" compared to more traditional assets. In particular, published results show that cryptocurrency returns are more predictable compared to returns in more traditional financial markets (e.g., Zhang et al., 2018; Al-Yahyaee, Mensi, and Yoon, 2018; Sensoy, 2019) and that technical trading rules may help investors earn statistically and economically significant, "abnormal" returns (Corbet et al., 2019; Grobys, Ahmed, and Sapkota, 2020; Fischer, Krauss, and Deinert, 2019). However, some evidence seem to indicate otherwise. For example, Anghel (2020b) shows that the no abnormal return null hypothesis can be hardly rejected after controlling for market frictions and data snooping. Our results complement and extend such findings by showing very little evidence of trading rule overperformance in the cryptocurrency market. Most null rejections can be associated with samples exhibiting downward price trends, thus validating our previous result in an empirical setup. However, after adjusting for luck, we find that test null rejections are well within the bounds of randomness, with some negligible exceptions.

The remainder of the paper is structured as follows. In Section 2, we discuss the statistical properties of lucky trading rules and identify the possible sources of data snooping bias. In Section 3, we report on a simulation exercise designed to investigate the factors that drive false discoveries. Section 4 performs an in depth investigation on the *directional preference effect*, which is the main determinant of data snooping bias. Section 5 investigates trading rule overperformance and data snooping in the cryptocurrency market. Section 6 concludes.

2. Sources of data snooping bias

In an empirical analysis, various testing procedures can be used to assess the statistical and economic performance of speculative trading rules. Using tests that account for multiple hypotheses is a crucial requirement[4], while choosing the right test and accounting for the data snooping efforts of others is at least as important. Nevertheless, many tests center on evaluating the null hypothesis $H_0\colon \max \mathbb{E}[d_k] \leq 0$, where d_k is the negative loss function associated with the overperformance of a trading rule or, more generally, of a prediction model $k = 1..K$, where K is the number of models being considered simultaneously. In the context of speculative rules used by traders in financial markets, the loss function is defined as the excess return over a benchmark trading strategy, which is typically the buy-and-hold rule

[4] Examples are the False Discovery Rate test proposed by Benjamini and Hochberg (1995), the Reality Check test proposed by White (2000), and their extensions. Harvey, Liu, and Saretto (2020) provide a comprehensive review of relevant multiple testing methods that can be used in finance applications.

because it is the optimal strategy when future prices are not predictable (e.g., follow a random walk). In addition, trading rules are designed to forecast the direction of future price movements and not the actual prices. Put differently, they are designed to instruct investors what positions they should take in the market. Specifically, trading rules generate predictions via a "signal" function δ_k that uses past, observed prices (or returns) and outputs either 1 when the price is expected to increase (the expected return is positive and the investor should go long), 0 when the price is expected to remain constant (the expected return is zero and the investor should stay out of the market), or -1 when the price is expected to decrease (the expected return is negative and the investor should go short). In practice, unchanged prices are hardly predicted, so we can discard this possibility from the analysis. However, as explained below, 0 can still appear as a 'signal' if prices are expected to decrease but short selling is not allowed.

These characteristics allow us to decompose the expected performance of speculative trading rules into its individual components and to identify the sources of data snooping. Accounting for trading costs, the negative loss function can be defined as:

$$d_{k,t} = \left[\delta_{k,t-1}\zeta_t - \mathbb{1}_{\{\delta_{k,t-1} \neq \delta_{k,t-2}\}}(\varphi + \lambda_t)\right] - \zeta_t \tag{1}$$

where ζ_t represents the market return, φ represents the fixed broker fee, and λ_t is the liquidity and/or price impact cost[5]. The term in the square brackets represents the return obtained by the trading rule, while subtracting the market return yields the excess return. Taking expectations results in the following decomposition of expected trading rule excess performance:

$$\mathbb{E}[d_k] = \left(\mathbb{E}[\delta_{k,t-1}] - 1\right)\mathbb{E}[\zeta_t] + \rho_{\delta,\zeta}\sigma_\delta\sigma_\zeta - \left[\theta(\varphi + \mathbb{E}[\lambda_t]) + \sigma_{\delta,\lambda}\right] \tag{2}$$

where $\rho_{\delta,\zeta} = Corr\left(\delta_{k,t-1}, \zeta_t\right) = Cov(\delta_{k,t-1}, \zeta_t)/\sigma_\delta\sigma_\zeta$ is the correlation between trading signals and market returns, $\sigma_\delta = \sqrt{Var\left(\delta_{k,t-1}\right)}$ is the volatility of trading signals, $\sigma_\zeta = \sqrt{Var(\zeta_t)}$ is the volatility of market returns, $\theta = \mathbb{P}\left(\delta_{k,t-1} \neq \delta_{k,t-2}\right)$ is the probability of making a trade, i.e. the trading frequency of the speculative rule, and $\sigma_{\delta,\lambda} = Cov(\delta_{k,t-1} \neq \delta_{k,t-2}, \lambda_t)$ is the covariance between trade occurrences and

[5] Note that the signal at *t-1* is used in Equation (1) to avoid look-ahead bias.

liquidity costs. When trading in a market that does not allow short selling or in which short trades are not functional, only long trades are possible and the signal function can take either 1 or 0. In this case, δ_k follows a Bernoulli distribution with mean $\mathbb{P}(\delta_{k,t-1}=1)=p$ and variance $p(1-p)$, where p can be interpreted as the tendency (of the trading rule) to participate in the market and $1-p$ as the tendency to stay out of the market. The expected excess performance becomes:

$$\mathbb{E}[d_k] = (p-1)\mathbb{E}[\zeta_t] + \rho_{\delta,\zeta}\sqrt{p(1-p)}\sigma_\zeta - \left[\theta(\varphi + \mathbb{E}[\lambda_t]) + \sigma_{\delta,\lambda}\right] \tag{3}$$

Conversely, when short trades are functional, δ_k can take either 1 or -1. In this case, δ_k can be modelled as the difference between two Bernoulli random variables, one taking 1 or 0 and another taking 0 or -1, with identical probability distributions. In this case, p can be interpreted as the tendency to follow the market and $1-p$ as the tendency to trade against it[6]. The expected excess performance becomes:

$$\mathbb{E}[d_k] = 2(p-1)\mathbb{E}[\zeta_t] + 2\rho_{\delta,\zeta}\sqrt{p(1-p)}\sigma_\zeta - \left[\theta(\varphi + \mathbb{E}[\lambda_t]) + \sigma_{\delta,\lambda}\right] \tag{4}$$

Equations (3) and (4) show that the excess performance of trading rules is a function of two market characteristics–expected return and volatility–and three trading rule properties–bullish (bearish) inclination, predictive accuracy (signal correlation with the market), and trading frequency. In addition, they show that there are three independent components associated with trading rule excess performance. Recognizing that applied researchers only observe and perform tests on finite data samples (see, e.g., the discussion in Romano and Wolf, 2018), they all constitute possible sources of bias in an empirical analysis. Looking from the right to the left, the first component, $\left[\theta(\varphi + \mathbb{E}[\lambda_t]) + \sigma_{\delta,\lambda}\right]$, is related with the level of market frictions in the form of direct and indirect trading costs. Ignoring market frictions or underestimating them would artificially inflate the performance of trading rules, this being a well-known source of bias. Fama (1965) provides an early example in this direction when discussing the results reported by Alexander (1961). As expected, our decomposition shows that this *trading cost effect* is related with trading frequency,

[6] In both Equations (3) and (4), p can be also interpreted as the bullish preference of the trading rules, i.e. the probability that the rule will expect prices to increase and trade with the market, while $1-p$ can be viewed as the bearish preference.

but also with the timing of trades in relation with market liquidity. This implies that ignoring trading costs overestimates the performance of rules that trade more often and especially in thinly traded markets.

The second component contains the correlation between trading signals and market returns, $\rho_{\delta,\zeta}$, which is a proxy for a trading rules' predictive accuracy. If the predictive accuracy is statistically and economically significant, speculative rules can help traders earn 'abnormal' returns. However, investment professionals and researchers routinely mine financial prices in search of better rules, this increasing the chances of finding and using rules that correlate with the market purely by luck and not due to real, economically significant predictive ability. This *spurious correlation effect* can lead to severe underperformance in out-of-sample applications and is the major source of focus for the discussion on data snooping bias in the literature, even if not explicitly identified as such (see, e.g., the seminal work of Brock, Lakonishok, and LeBaron, 1992; or the discussion in Sullivan, Timmermann, and White, 1999). We further note that our decomposition shows that the *spurious correlation effect* can be exacerbated in more volatile markets and for lucky trading rules that have a balanced directional preference, i.e. a 50% probability of trading with the market. Equation (4) further shows that the *spurious correlation effect* is two times larger for trading rules that can also take short positions, implying that tests that allow short trades in markets where they are not functional may additionally inflate trading rule overperformance.

The third component, $(p-1)\mathbb{E}[\zeta_t]$, is a function of a trading rule's directional tendency and market expected return. By construction, $p < 1 \Leftrightarrow p - 1 < 0$ for active trading rules, implying that they benefit investors when $\mathbb{E}[\zeta_t] < 0$, i.e. market prices are expected to decrease, but are disadvantageous otherwise. However, similar to spurious correlation, the increased performance resulting from a bearish preference in downward trending markets could be the result of luck and not skill. For example, the same effect can be replicated by any naïve trading strategy that randomly chooses to enter the market with probability p. As data snooping intensifies, the chances of finding a trading rule that exactly matches the directional preference of the market also increases. Among others, this implies that the *directional preference effect* may be responsible for some trading rules appearing more profitable in periods of declining prices (e.g., Taylor, 2014). In addition, because strong (especially negative) price trends can be associated with higher volatility (Schwert, 2011), the *directional preference effect* may also be responsible for some results that link trading rule overperformance with volatility-related asset pricing factors (e.g., Han, Yang, and Zhou, 2013). Finally, the *directional preference effect* is also two times larger for trading rules that can also take short positions.

In summary, multiple testing procedures that evaluate the performance of speculative trading rules are typically defined using asymptotic considerations, but finite sample limitations and extensive data mining can bias their results. Our decomposition allows us to identify the possible sources of luck and, consequently, false discoveries. First, ignoring trading costs and other market frictions overestimates trading performance. Second, excess performance due to spurious correlation increases when many trading rules are evaluated on the same data sample. Third, trading rules that have a bearish preference appear more profitable in samples where prices experience a negative tendency. Researcher choices that influence these properties, such as data sample location and length, the number of trading rules selected for testing, the amount of trading costs considered, enabling short trades, or even standardizing the test statistic (i.e., redefining the null as $H_0\colon \max \mathbb{E}[d_k]/VAR(d_k) \leq 0$), may materially impact the results by favoring trading rules with lucky characteristics. When researchers do not adequately account for market frictions, limitations associated with finite-sample control of estimation errors, or the data snooping efforts of others, test results and associated inferences would be biased. How much does each factor contribute to data snooping? The next Section provides an answer to this question.

3. Drivers of data snooping bias

3.1. Setup

We perform a simulation exercise designed to investigate the properties of lucky trading rules and their relative contribution to the data snooping bias that might arise in empirical evaluations. Lucky rules are selected from a total of 688,739 technical trading rules, which are defined following Anghel (2020a)[7].

[7] We use the 686,304 trading rules defined by Anghel (2020a, Appendix A), to which we add 2,435 rules constructed from the Trading Range Breakouts (TRB) method, which has been considered by Sullivan, Timmermann, and White (1999).

The data consists of 6x1000=6,000 samples of randomly generated trading prices and volumes spanning one month (approx. 22 observations each)[8], which are obtained from a discretized no-drift geometric Brownian Motion with volatility parameters $\sigma \in \{0.15, 0.20, 0.25, 0.30, 0.35, 0.40\}$. On each sample, we test all rules and isolate the one that earns the highest excess return, i.e. the luckiest trading rule. Then, we evaluate the statistical significance of the results using two multiple testing procedures that are designed to control for data snooping, namely White's (2000) Reality Check test and Hansen's (2005) Superior Predictive Ability test. Both tests control for the Family-wise Error Rate, making them less powerful compared to alternatives that control for (variations of) the less stringent False Discovery Rate (see, e.g., Barras, Scaillet and Wermers, 2010; or Efron, 2012). However, they are also less prone to making Type I errors, false discoveries, which is the metric that we are interested in.

There are three main differences in our approach compared to similar endeavors in the related asset pricing literature, such as the work of Chordia, Goyal and Saretto (2020). First, we focus on analyzing the luck of speculative trading rules when applied on individual time-series (assets) and do not form or test portfolios of assets. This bypasses the need to assume a factor model for the data generating process and implies that our results can be generalized to a wide set of financial asset classes. Second, we assume no dependence structure (autocorrelation) of the time series that we simulate, thus making any test null rejection a false discovery and enforcing the null hypothesis of no trading rule overperformance directly from the construction of the dataset. Third, because we investigate trading rules based on technical analysis, we also simulate other relevant market information alongside Close prices (returns), such as Open, High, and Low prices, plus trading volumes. There are also some important difference compared to the work of Anghel (2020a), on which our analysis builds upon. On the one hand, we measure and use data sample statistics to analyze the interaction between false discoveries and sample characteristics, such as average returns, volatility, skewness, and kurtosis. Among others, this enables us to isolate the contribution that the

[8] The data construction follows Anghel (2020a, Section 4.1) and we do not detail it further here. One month samples are used to ease computational demands. While samples of approximately 22 observations may be considered a limitation of the analysis, Anghel (2020a) finds that the results are not significantly altered when increasing the sample length. This shows that the properties of lucky trading rules and the conclusions inferred from the analysis do not materially change with sample length.

directional preference effect has on the data snooping bias due to lucky trading rules. On the other hand, we do not fix the testing conditions, instead varying them based on methodological choices that applied researchers would need to make. First, we run the tests with and without a fixed broker fee[9]. Second, we run the tests with and without liquidity/price impact costs[10]. Third, we run the tests with and without short trades[11]. Fourth, we run the tests with and without standardizing the statistic. When the statistic is standardized, the test follows the procedure described in Hansen (2005), which additionally handles for the impact of poor performing trading rules. The Bootstrap Bonferroni adjustment of Romano and Wolf (2018) is also considered, but similar to Anghel (2020a) we find no material differences compared to the Hansen (2005) adjustment, thus deciding not to discuss it further (results are available upon request). Conversely, when the statistic is not standardized, the test follows the procedure described in White (2000). Overall, there are 16 different methodological combinations that we test on all 6,000 samples, amounting to a total of 96,000 test runs. Comparing the results enables us to estimate the impact that each choice has on the characteristics of lucky trading rules and on the data snooping bias that each would introduce in empirical tests. Finally, we evaluate the impact of not accounting for the data snooping of others in various testing conditions, by first testing only the luckiest trading rule and then slowly adding the other rules until the entire set is used in the tests. Anghel (2020a) performed such an analysis but only considered the case in which trading costs are enables and short trades are disabled.

3.2. Results–factors that drive the characteristics of lucky trading rules

We first analyze the impact of the different methodological choices and sample properties on the characteristics of lucky trading rules. Table 1 reports the results of a linear regression model that evaluates the drivers of each characteristic. Figure 1 reports how the distribution of each characteristic changes with testing conditions. Figures A1 trough A4 in Appendix A add to the visual analysis by showing how each characteristic interacts with methodological choices and sample statistics such as average returns, volatility,

---

[9] A fee of 1% per round trip is considered, which should be sufficient to account for trading fees in most financial markets.
[10] When considering price impact costs, we simulate trading at the least favorable daily prices, the High price for buy trades and the Low price for sell trades (otherwise, trades are simulated at the average daily price). This should also be sufficient to account for costs associated with the bid-ask spread and price impact.
[11] Short trades might not be possible in some markets but could be ignored even in markets where they are functional.

skewness, and kurtosis. Several results are worth noting. First, we find that an inflated predictive accuracy ($\hat{\rho}$) is the most stable characteristic of lucky trading rules, as shown by the highly significant intercept and low explanatory power of the regression model (R^2 is 0.063). This result implies that the *spurious correlation effect* increases with the number of trading rules being considered but is largely independent from the data sample properties and methodological choices. Nevertheless, we do find that predictive accuracy increases with volatility, showing that the spurious correlation rises in more volatile markets, as predicted by Equations (3) and (4). Also, considering trading fees, ignoring short trades, and standardizing the test statistic help control the effect, but only marginally. Interestingly, we also find a weak positive influence of market returns on signal-to-market correlation, hinting that spurious correlation increases in upward trending markets but is less important when average returns are negative.

Second, trading frequency ($\hat{\theta}$) and, especially, market preference ($\hat{p}$) are more dependent on testing conditions (R^2 are 0.364 and 0.573, respectively). On the one hand, as expected, adjusting for trading costs significantly decreases the trading frequency of lucky rules. Considering short trades and standardizing the test statistic also have a (mostly positive) influence, but in this case the effect is much weaker and depends on market conditions, as the coefficients estimated for the interaction terms have alternating signs. For the most part, trading frequency tends to decrease with average returns and increase with volatility, showing that trading costs, which are mainly responsible for the *trading cost effect*, are also indirectly linked with the *market preference* and *spurious correlation* effects. On the other hand, we find that market preference is strongly determined by market returns: lucky trading rules have a bearish preference when returns decrease (are negative) and a bullish preference when returns increase (are positive). Moreover, adjusting for trading fees and liquidity costs strengthen the effect of returns on market preference by about 20% each[12], a result which supports the aforementioned indirect link between trading costs and the *market preference effect*. Interestingly, the intercept in the market preference equation is 0.50, a value which all other things equal maximizes the variance of the signal function and, in turn, the *spurious correlation effect*.

---

[12] This is calculated as the ratio between γ_{FEE} or γ_{LIQCOST}, respectively, and γ (which is the coefficient showing the base effect of market returns on market preference).

This shows that the "luck" of speculative trading rules takes more forms than previously recognized. We also find that ignoring short trades and standardizing the test statistic weaken the influence of market returns on trading rule directional preference by about -10% each, but that making these methodological choices have an overall positive impact on directional preference.

Table 1. Methodological choices, sample properties, and the characteristics of lucky trading rules

NOTE. This table reports the coefficients estimated for the following linear regression model:

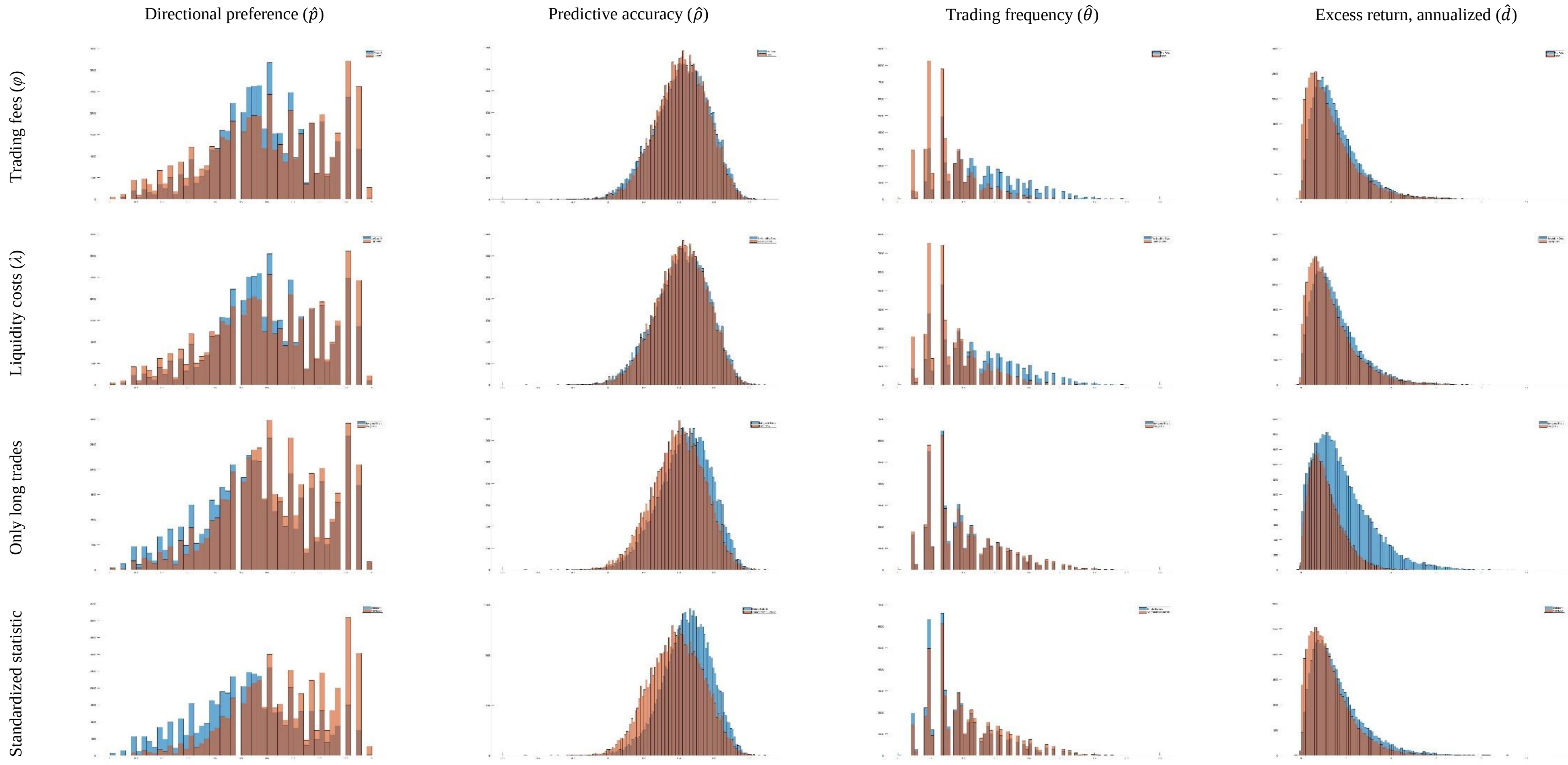
$$\hat{X} = \alpha + \beta\hat{Y} + \gamma\hat{Y}\hat{\zeta} + \delta\hat{Y}\hat{\sigma} + \varepsilon$$

*where $\hat{X}$ is one characteristic of lucky trading rules (either bullish preference-$\hat{p}$, prediction accuracy-$\hat{\rho}$, trading frequency-$\hat{\theta}$, or (annualized) average excess return-$\hat{d}$); $\hat{Y}$ is a vector of dummy explanatory variables representing possible methodological choices (considering trading fees-FEE; considering liquidity costs-LIQCOST; considering only long trades-LONG_ONLY; or standardizing the test statistic-STDSTAT), taking the value of 1 if used and 0 otherwise; $\hat{\zeta}$ is the sample average return; $\hat{\sigma}$ is the sample volatility; β, γ, δ are vectors of coefficients, while ε is the error term. Each regression is estimated on 96,000 data points, which are obtained by running the simulation exercise described in Section 3.1 on 6000 data samples with 16 possible methodological choices. T-statistics are reported in square parenthesis, while ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.*

Dependent variable ($\hat{X}$):	$\hat{p}$	$\hat{\rho}$	$\hat{\theta}$	$\hat{d}$
α	0.5036***	0.4485***	0.2575***	0.0389***
	[171.15]	[142.19]	[137.80]	[4.10]
β_{FEE}	-0.0018	-0.0079***	-0.1404***	-0.1635***
	[-0.71]	[-2.81]	[-84.00]	[-19.27]
β_{LIQCOST}	-0.0026	0.0009	-0.0621***	0.0022
	[-1.02]	[0.31]	[-37.19]	[0.25]
$\beta_{\text{LONG\_ONLY}}$	0.0435***	-0.0412***	-0.0006	-0.0388***
	[16.54]	[-14.61]	[-0.36]	[-4.58]
β_{STDSTAT}	0.1389***	-0.0522***	0.0164***	0.0017
	[52.78]	[-18.51]	[9.82]	[0.21]
γ	0.1245***	0.0024**	-0.0184***	-0.9733***
	[131.64]	[2.37]	[-30.69]	[-319.21]
γ_{FEE}	0.0260***	-0.0043***	0.0137***	-0.0013
	[30.77]	[-4.74]	[25.50]	[-0.49]
γ_{LIQCOST}	0.0221***	-0.0047***	0.0120***	-0.0036
	[26.17]	[-5.24]	[22.40]	[-1.33]
$\gamma_{\text{LONG\_ONLY}}$	-0.0100***	-0.0003	-0.0077***	0.4019***
	[-11.92]	[-0.35]	[-14.37]	[147.37]
γ_{STDSTAT}	-0.0160***	0.0017*	-0.0315***	0.1155***
	[-18.99]	[1.95]	[-58.78]	[42.36]
δ	0.0314***	0.0644***	0.0197***	7.8321***
	[3.06]	[5.86]	[3.03]	[236.87]
δ_{FEE}	0.0091	0.0248**	0.2093***	-0.2251***
	[0.99]	[2.52]	[35.92]	[-7.61]
δ_{LIQCOST}	0.0097	-0.0049	-0.0116**	-0.6962***
	[1.06]	[-0.50]	[-1.99]	[-23.54]
$\delta_{\text{LONG\_ONLY}}$	-0.0057	0.0103	0.0154***	-2.4589***
	[-0.62]	[1.05]	[2.65]	[-83.14]
δ_{STDSTAT}	-0.0386***	-0.0385***	0.0101*	-1.1275***
	[-4.20]	[-3.91]	[1.74]	[-38.12]
Adjusted R^2	0.5734	0.0630	0.3648	0.8468
F-statistic	9218.56***	460.63***	3939.93***	37921.15***

Figure 1. Methodological choices and characteristics of lucky trading rules

NOTE. This figure reports how the characteristics of lucky trading rules changes with methodological choices such as considering trading fees, considering liquidity costs, only considering long trades, or standardizing the test statistic. The distributions are overlaid, with orange bars denoting the frequency when the choice is true and blue bars denoting the frequency when the choice is false. Each distribution is plotted from 48,000 points of data.



Third, when analyzing the impact of the different testing conditions on the apparent profitability of lucky trading rules, we find that the average (annualized) excess returns are to a great degree explained by market conditions and methodological choices (R^2 is 0.846). On the one hand, we find a strong positive impact of sample volatility, showing that the *spurious correlation effect* plays an important role. Nevertheless, this effect can be controlled by eliminating short trades (-31%), standardizing the test statistic (-14%), adjusting for liquidity costs (-9%), and adjusting for trading fees (-3%). On the other hand, and more importantly, we find a strong negative correlation with realized market returns, which shows that the *market preference effect* is the main driver of the apparent overperformance of speculative trading rules, especially when both long and short trades are considered (ignoring short trades decreases the effect by -41%) and the test statistic is not standardized (standardizing the statistic decreases the effect by -12%). Interestingly, while trading fees have a direct negative influence on excess returns, they are not sufficient in eliminating the *directional preference* or *spurious correlation effects*. Also, liquidity costs do not have a direct impact on excess returns, instead being useful only for reducing the impact of volatility, i.e. the *spurious correlation effect*.

3.3. Results– factors that drive false discoveries

The results in Section 3.2 show that trading rules' excess returns are highly dependent on testing conditions. However, applied researchers do not make inferences based solely on raw trading performance, but instead test for statistical significance when accounting for data snooping. Thus, when modern testing frameworks are used, the final results depend not only on the characteristics of the luckiest trading rule, but also on the characteristics of the other rules from the excess returns of which the null distribution is constructed. How do test results behave in the presence of lucky trading rules and when are false discoveries most likely to occur? As discussed in Section 3.1, we test the statistical significance of trading rules excess returns using the RC (White, 2000) and SPA (Hansen, 2005) tests. We start by testing only the luckiest trading rule in each sample and then add the other rules to account for data snooping. For each of the 16 methodological combinations and each of the 6,000 data samples, we estimate the distribution of excess returns for both tests using the Stationary Bootstrap of Politis and Romano (1994) with 1,000 iterations.

Then, we compute critical values and test p-values. Finally, we aggregate the results to estimate the proportion of false discoveries (which is a proxy for the Family-wise error rate) at a significance level of 5%. The results are reported in Table 2. Panel A reports the results obtained on all samples, Panels B and C split the results based on the tendency of the sample returns, while Panel D shows the proportion of false discoveries that arise when average returns are negative, which can be interpreted as the relative contribution of the *directional preference effect* to data snooping bias.

The results show that false discoveries significantly decrease when the number of tested trading rules increases. This supports the findings of Anghel (2020a) and shows that accounting for the data snooping efforts of others is essential in controlling data snooping bias. Also similar to Anghel (2020a), we find that FWER is adequately controlled when at least 2^{13}–2^{17} trading rules are tested, while false discoveries are not eliminated even after accounting for all trading rules from which lucky rules were extracted. However, our analysis extends on the results of Anghel (2020a) and additionally shows that data snooping bias is highly dependent on other methodological choices. First, we find that ignoring trading costs provides and inadequate control of FWER, i.e. false discovery proportions surpass the 5% asymptotically-valid limit, even when all 688,739 trading rules are accounted for. This shows that trading costs, especially trading fees, are a minimum requirement for controlling data snooping bias in empirical tests.

Second, we find that standardizing the test statistic is beneficial to the analysis (reduces false discoveries) when both long and short trades are considered but is detrimental (increases false discoveries) when short trades are ignored. This implies that standardizing the test statistic is a good testing strategy when the volatility of trading signals is expected to be higher–such as in markets where short trades are functional–but could otherwise escalate the data snooping bias. Among others, this finding can be used to explain the result in Anghel (2020a), which considered only long trades and found the RC test better at controlling data snooping compared with the SPA test.

Table 2. Methodological choices, sample properties, and false discoveries

Panel A. Proportion of False Discoveries–all samples

*NOTE. This table reports the proportion of false discoveries (a proxy for the Familywise error rate–FWER) generated by lucky trading rules on randomly generated data, estimated for a significance level of 5% when varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). Each value is aggregated from 6,000 data points, corresponding to the same number of samples on which tests were performed; except for the values in the 'Both' column, which are based on all 96,000 results. Values that fall below the 5% threshold signal a test that offers adequate in-sample control of FWER, and are highlighted using **bolded** text. The bottom three rows report critical values for the excess return distribution (the x-th percentile of the distribution, where x is either 90%, 95%, or 99%).*

FEE		Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIQCOST		Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
LONGONLY		Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
STDSTAT		Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules	1	0.481	0.522	0.755	0.602	0.787	0.385	0.581	0.386	0.615	0.353	0.524	0.336	0.568	0.268	0.387	0.211	0.418
	2	0.429	0.437	0.714	0.533	0.751	0.317	0.535	0.323	0.575	0.294	0.481	0.276	0.522	0.221	0.344	0.170	0.374
	4	0.378	0.361	0.671	0.454	0.706	0.260	0.484	0.263	0.528	0.238	0.435	0.225	0.477	0.177	0.303	0.130	0.335
	8	0.344	0.326	0.636	0.390	0.677	0.233	0.445	0.219	0.492	0.211	0.398	0.186	0.442	0.159	0.275	0.111	0.305
	16	0.329	0.311	0.622	0.368	0.662	0.224	0.429	0.199	0.475	0.203	0.378	0.169	0.424	0.153	0.258	0.100	0.290
	32	0.322	0.300	0.612	0.365	0.656	0.217	0.421	0.195	0.466	0.196	0.369	0.165	0.415	0.149	0.248	0.098	0.281
	64	0.294	0.212	0.594	0.340	0.647	0.158	0.401	0.171	0.454	0.145	0.349	0.145	0.401	0.108	0.231	0.081	0.268
	128	0.294	0.212	0.594	0.340	0.647	0.158	0.401	0.171	0.454	0.145	0.349	0.145	0.401	0.108	0.231	0.081	0.268
	256	0.274	0.212	0.553	0.335	0.603	0.158	0.359	0.168	0.416	0.144	0.306	0.142	0.363	0.107	0.203	0.078	0.241
	512	0.183	0.158	0.363	0.263	0.419	0.114	0.214	0.120	0.268	0.103	0.184	0.099	0.231	0.077	0.116	0.052	0.142
	1024	0.148	0.131	0.308	0.207	0.365	0.087	0.174	0.087	0.227	0.081	0.147	0.070	0.193	0.057	0.089	***0.033***	0.116
	2048	0.147	0.128	0.307	0.206	0.365	0.086	0.174	0.086	0.227	0.080	0.146	0.070	0.192	0.056	0.088	***0.033***	0.116
	4096	0.141	0.122	0.294	0.201	0.351	0.083	0.164	0.083	0.216	0.077	0.138	0.068	0.182	0.054	0.083	***0.032***	0.112
	8192	0.110	0.113	0.211	0.172	0.268	0.077	0.115	0.071	0.157	0.069	0.097	0.059	0.131	***0.049***	0.058	***0.027***	0.081
	16384	0.104	0.105	0.203	0.164	0.260	0.070	0.109	0.066	0.151	0.063	0.092	0.054	0.127	***0.047***	0.053	***0.024***	0.076
	32768	0.072	0.091	0.130	0.135	0.178	0.063	0.060	0.052	0.090	0.056	0.052	***0.043***	0.074	***0.039***	***0.028***	***0.018***	***0.039***
	65536	0.070	0.090	0.127	0.132	0.174	0.061	0.058	0.051	0.087	0.055	0.050	***0.042***	0.070	***0.039***	***0.027***	***0.017***	***0.038***
	131072	0.067	0.088	0.123	0.127	0.168	0.059	0.056	***0.049***	0.083	0.053	***0.047***	***0.041***	0.068	***0.038***	***0.025***	***0.016***	***0.035***
	262144	0.062	0.082	0.114	0.116	0.156	0.055	0.053	***0.043***	0.076	***0.049***	***0.043***	***0.036***	0.064	***0.034***	***0.022***	***0.013***	***0.033***
	524288	0.057	0.079	0.105	0.112	0.145	0.053	***0.046***	***0.040***	0.070	***0.047***	***0.039***	***0.035***	0.059	***0.032***	***0.019***	***0.013***	***0.028***
	688739	***0.049***	0.072	0.090	0.095	0.126	0.050	***0.037***	***0.033***	0.058	***0.044***	***0.030***	***0.029***	0.050	***0.029***	***0.015***	***0.010***	***0.024***
ExRet Pctl. 90%		2.630	3.683	2.936	2.197	2.101	3.458	2.839	1.891	1.790	3.468	2.835	1.896	1.794	3.300	2.774	1.701	1.578
ExRet Pctl. 95%		3.247	4.375	3.463	2.576	2.454	4.163	3.383	2.234	2.122	4.179	3.402	2.234	2.132	3.997	3.410	2.042	1.909
ExRet Pctl. 99%		4.772	5.898	4.752	3.298	3.147	5.653	4.913	2.987	2.844	5.722	4.797	3.009	2.846	5.514	4.884	2.814	2.672

Panel B. Proportion of False Discoveries–samples with positive average return

NOTE. This table reports the proportion of false discoveries (a proxy for the Familywise error rate–FWER) generated by lucky trading rules on randomly generated data, estimated for a significance level of 5% when varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). Each value is aggregated from 2,961 data points, corresponding to the same number of samples on which average returns are positive; except for the values in the 'Both' column, which are based on 47,376 results. Values that fall below the 5% threshold signal a test that offers adequate in-sample control of FWER, and are highlighted using ***bolded*** *text. The bottom three rows report critical values for the excess return distribution (the x-th percentile of the distribution, where x is either 90%, 95%, or 99%).*

FEE		Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIQCOST		Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
LONGONLY		Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
STDSTAT		Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules	1	0.188	0.163	0.531	0.255	0.584	0.056	0.272	0.052	0.306	***0.043***	0.220	***0.033***	0.263	***0.015***	0.091	***0.006***	0.118
	2	0.149	0.093	0.464	0.163	0.519	***0.029***	0.225	***0.028***	0.258	***0.025***	0.181	***0.018***	0.215	***0.008***	0.068	***0.004***	0.086
	4	0.115	0.051	0.396	0.086	0.444	***0.014***	0.175	***0.010***	0.210	***0.011***	0.142	***0.008***	0.170	***0.003***	0.054	***0.001***	0.065
	8	0.093	***0.031***	0.350	***0.024***	0.398	***0.007***	0.148	***0.003***	0.173	***0.006***	0.116	***0.002***	0.140	***0.002***	***0.041***	***0.000***	0.054
	16	0.085	***0.023***	0.330	***0.012***	0.377	***0.005***	0.137	***0.000***	0.160	***0.005***	0.103	***0.001***	0.128	***0.001***	***0.035***	***0.000***	***0.048***
	32	0.083	***0.020***	0.321	***0.011***	0.369	***0.004***	0.133	***0.000***	0.156	***0.004***	0.099	***0.001***	0.123	***0.001***	***0.033***	***0.000***	***0.045***
	64	0.075	***0.000***	0.300	***0.006***	0.357	***0.000***	0.118	***0.000***	0.148	***0.000***	0.086	***0.000***	0.115	***0.000***	***0.028***	***0.000***	***0.042***
	128	0.075	***0.000***	0.300	***0.006***	0.357	***0.000***	0.118	***0.000***	0.148	***0.000***	0.086	***0.000***	0.115	***0.000***	***0.028***	***0.000***	***0.042***
	256	0.064	***0.000***	0.260	***0.005***	0.309	***0.000***	0.099	***0.000***	0.126	***0.000***	0.068	***0.000***	0.095	***0.000***	***0.022***	***0.000***	***0.035***
	512	***0.027***	***0.000***	0.115	***0.001***	0.146	***0.000***	***0.035***	***0.000***	0.051	***0.000***	***0.027***	***0.000***	***0.036***	***0.000***	***0.006***	***0.000***	***0.009***
	1024	***0.017***	***0.000***	0.071	***0.000***	0.103	***0.000***	***0.021***	***0.000***	***0.035***	***0.000***	***0.015***	***0.000***	***0.025***	***0.000***	***0.003***	***0.000***	***0.004***
	2048	***0.017***	***0.000***	0.070	***0.000***	0.103	***0.000***	***0.021***	***0.000***	***0.035***	***0.000***	***0.015***	***0.000***	***0.024***	***0.000***	***0.003***	***0.000***	***0.004***
	4096	***0.015***	***0.000***	0.064	***0.000***	0.094	***0.000***	***0.018***	***0.000***	***0.030***	***0.000***	***0.012***	***0.000***	***0.022***	***0.000***	***0.002***	***0.000***	***0.004***
	8192	***0.008***	***0.000***	***0.032***	***0.000***	0.053	***0.000***	***0.008***	***0.000***	***0.014***	***0.000***	***0.005***	***0.000***	***0.009***	***0.000***	***0.001***	***0.000***	***0.002***
	16384	***0.007***	***0.000***	***0.028***	***0.000***	***0.049***	***0.000***	***0.006***	***0.000***	***0.012***	***0.000***	***0.004***	***0.000***	***0.008***	***0.000***	***0.000***	***0.000***	***0.002***
	32768	***0.002***	***0.000***	***0.009***	***0.000***	***0.014***	***0.000***	***0.001***	***0.000***	***0.003***	***0.000***	***0.001***	***0.000***	***0.003***	***0.000***	***0.000***	***0.000***	***0.000***
	65536	***0.002***	***0.000***	***0.008***	***0.000***	***0.014***	***0.000***	***0.001***	***0.000***	***0.003***	***0.000***	***0.001***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.000***
	131072	***0.002***	***0.000***	***0.007***	***0.000***	***0.012***	***0.000***	***0.001***	***0.000***	***0.003***	***0.000***	***0.001***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.000***
	262144	***0.001***	***0.000***	***0.006***	***0.000***	***0.010***	***0.000***	***0.001***	***0.000***	***0.002***	***0.000***	***0.001***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.000***
	524288	***0.001***	***0.000***	***0.004***	***0.000***	***0.009***	***0.000***	***0.001***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.000***
	688739	***0.001***	***0.000***	***0.003***	***0.000***	***0.007***	***0.000***	***0.000***	***0.000***	***0.002***	***0.000***	***0.000***	***0.000***	***0.001***	***0.000***	***0.000***	***0.000***	***0.000***
ExRet Pctl. 90%		1.596	2.346	1.889	1.457	1.393	2.083	1.660	1.159	1.095	2.094	1.645	1.168	1.106	1.883	1.490	0.975	0.895
ExRet Pctl. 95%		1.967	2.695	2.256	1.641	1.598	2.424	2.035	1.346	1.276	2.429	2.044	1.366	1.318	2.214	1.859	1.158	1.075
ExRet Pctl. 99%		2.740	3.284	3.072	2.064	1.994	3.029	2.796	1.742	1.658	3.061	2.859	1.776	1.718	2.880	2.541	1.508	1.429

Panel C. Proportion of False Discoveries–samples with negative average return

NOTE. This table reports the proportion of false discoveries (a proxy for the Familywise error rate–FWER) generated by lucky trading rules on randomly generated data, estimated for a significance level of 5% when varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). Each value is aggregated from 3,039 data points, corresponding to the same number of samples on which average market returns are negative; except for the values in the 'Both' column, which are based on 48,624 results. Values that fall below the 5% threshold signal a test that offers adequate in-sample control of FWER, and are highlighted using ***bolded*** *text. The bottom three rows report critical values for the excess return distribution (the x-th percentile of the distribution, where x is either 90%, 95%, or 99%).*

FEE	Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIQCOST	Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
LONGONLY	Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
STDSTAT	Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules: 1	0.767	0.872	0.973	0.940	0.986	0.705	0.882	0.711	0.917	0.655	0.821	0.631	0.865	0.514	0.676	0.411	0.711
2	0.702	0.773	0.957	0.893	0.978	0.599	0.838	0.611	0.884	0.555	0.774	0.528	0.821	0.427	0.613	0.332	0.654
4	0.634	0.663	0.939	0.813	0.961	0.500	0.784	0.509	0.838	0.458	0.721	0.437	0.775	0.346	0.545	0.256	0.597
8	0.588	0.613	0.914	0.746	0.948	0.454	0.735	0.430	0.803	0.412	0.673	0.365	0.735	0.312	0.504	0.219	0.550
16	0.566	0.591	0.905	0.714	0.940	0.437	0.714	0.394	0.782	0.395	0.647	0.333	0.711	0.300	0.475	0.198	0.525
32	0.555	0.573	0.895	0.708	0.936	0.424	0.702	0.385	0.769	0.382	0.632	0.324	0.698	0.293	0.458	0.193	0.510
64	0.508	0.419	0.881	0.665	0.930	0.312	0.678	0.338	0.752	0.285	0.606	0.287	0.679	0.212	0.429	0.160	0.488
128	0.508	0.419	0.881	0.665	0.930	0.312	0.678	0.338	0.752	0.285	0.606	0.287	0.679	0.212	0.429	0.160	0.488
256	0.479	0.418	0.838	0.657	0.889	0.312	0.612	0.331	0.699	0.284	0.537	0.280	0.625	0.210	0.380	0.155	0.442
512	0.335	0.312	0.605	0.519	0.686	0.224	0.388	0.237	0.480	0.204	0.337	0.195	0.420	0.151	0.224	0.103	0.270
1024	0.276	0.259	0.538	0.408	0.619	0.172	0.323	0.172	0.415	0.159	0.275	0.139	0.356	0.113	0.171	0.065	0.224
2048	0.274	0.253	0.538	0.406	0.619	0.170	0.323	0.170	0.414	0.158	0.274	0.137	0.355	0.111	0.171	0.064	0.224
4096	0.264	0.241	0.519	0.396	0.600	0.164	0.306	0.165	0.397	0.152	0.261	0.134	0.338	0.107	0.162	0.063	0.216
8192	0.209	0.222	0.386	0.340	0.477	0.151	0.219	0.141	0.296	0.135	0.186	0.116	0.249	0.097	0.113	0.054	0.157
16384	0.198	0.208	0.372	0.323	0.466	0.138	0.208	0.130	0.286	0.124	0.177	0.107	0.242	0.092	0.104	***0.047***	0.148
32768	0.139	0.179	0.247	0.266	0.337	0.123	0.117	0.103	0.174	0.110	0.101	0.084	0.143	0.078	0.054	***0.035***	0.077
65536	0.136	0.177	0.243	0.260	0.331	0.121	0.114	0.101	0.169	0.109	0.097	0.083	0.136	0.077	0.052	***0.034***	0.075
131072	0.131	0.173	0.235	0.251	0.320	0.116	0.109	0.097	0.161	0.105	0.092	0.081	0.133	0.075	***0.050***	***0.031***	0.069
262144	0.121	0.162	0.219	0.229	0.298	0.108	0.103	0.084	0.148	0.096	0.084	0.072	0.125	0.067	***0.044***	***0.026***	0.064
524288	0.112	0.157	0.203	0.221	0.277	0.105	0.090	0.078	0.135	0.092	0.076	0.068	0.115	0.063	***0.038***	***0.025***	0.055
688739	0.097	0.143	0.173	0.188	0.242	0.099	0.073	0.065	0.113	0.086	0.060	0.056	0.098	0.056	***0.030***	***0.020***	***0.047***
ExRet Pctl. 90%	3.217	4.371	3.428	2.567	2.440	4.150	3.367	2.227	2.112	4.166	3.380	2.227	2.124	3.985	3.406	2.039	1.904
ExRet Pctl. 95%	3.851	5.108	3.944	2.946	2.790	4.890	3.962	2.603	2.447	4.912	3.947	2.608	2.470	4.714	4.040	2.415	2.253
ExRet Pctl. 99%	5.335	6.328	5.270	3.614	3.372	6.158	5.492	3.233	3.066	6.155	5.426	3.316	3.095	5.999	5.516	3.076	2.910

Panel D. Relative Proportion of False Discoveries–samples with negative average return

NOTE. This table reports the number of false discoveries estimated on bearish samples (samples exhibiting a negative average return) divided by the total number of false discoveries estimated on all samples, using a significance level of 5% and while varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). The results can be used to infer the relative contribution of the directional preference effect to the aggregate proportion of false discoveries (data snooping bias) caused by lucky trading rules.

FEE		Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIQCOST		Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
LONGONLY		Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
STDSTAT		Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules	1	60.9%	68.8%	29.7%	57.7%	25.9%	85.4%	53.2%	86.4%	50.3%	87.8%	58.1%	90.1%	53.7%	94.4%	76.6%	97.3%	71.8%
	2	65.3%	78.8%	35.0%	69.3%	31.0%	91.0%	58.0%	91.2%	55.1%	91.6%	62.4%	93.6%	58.7%	96.2%	80.2%	97.6%	77.0%
	4	69.5%	86.0%	41.0%	81.1%	37.1%	94.5%	63.8%	96.3%	60.2%	95.2%	67.5%	96.3%	64.3%	98.1%	82.1%	99.2%	80.6%
	8	72.9%	90.4%	45.0%	93.8%	41.2%	97.0%	66.9%	98.8%	64.8%	97.3%	71.0%	98.9%	68.3%	98.9%	85.2%	99.7%	82.3%
	16	74.0%	92.5%	46.9%	96.7%	43.0%	97.6%	68.1%	100.0%	66.3%	97.7%	72.8%	99.6%	69.8%	99.1%	86.3%	100.0%	83.3%
	32	74.3%	93.3%	47.5%	96.8%	43.7%	98.0%	68.3%	100.0%	66.6%	98.1%	73.2%	99.6%	70.3%	99.3%	86.7%	100.0%	83.9%
	64	74.5%	100.0%	49.5%	98.3%	44.8%	100.0%	70.7%	100.0%	67.5%	100.0%	75.4%	100.0%	71.4%	100.0%	87.9%	100.0%	84.4%
	128	74.5%	100.0%	49.5%	98.3%	44.8%	100.0%	70.7%	100.0%	67.5%	100.0%	75.4%	100.0%	71.4%	100.0%	87.9%	100.0%	84.4%
	256	76.8%	100.0%	53.0%	98.4%	48.7%	100.0%	72.3%	100.0%	69.7%	100.0%	77.7%	100.0%	73.9%	100.0%	89.4%	100.0%	85.3%
	512	85.4%	100.0%	68.4%	99.7%	65.2%	100.0%	83.7%	100.0%	81.0%	100.0%	85.5%	100.0%	84.3%	100.0%	95.1%	100.0%	93.3%
	1024	88.3%	100.0%	76.9%	100.0%	71.7%	100.0%	88.0%	100.0%	84.5%	100.0%	89.6%	100.0%	87.2%	100.0%	96.2%	100.0%	96.2%
	2048	88.3%	100.0%	77.1%	100.0%	71.6%	100.0%	88.0%	100.0%	84.7%	100.0%	89.6%	100.0%	87.3%	100.0%	96.2%	100.0%	96.2%
	4096	89.1%	100.0%	78.2%	100.0%	73.1%	100.0%	89.1%	100.0%	85.9%	100.0%	91.0%	100.0%	88.1%	100.0%	97.6%	100.0%	96.4%
	8192	92.9%	100.0%	84.8%	100.0%	80.2%	100.0%	92.9%	100.0%	91.0%	100.0%	94.4%	100.0%	92.8%	100.0%	98.8%	100.0%	97.5%
	16384	93.3%	100.0%	86.0%	100.0%	81.3%	100.0%	94.1%	100.0%	91.7%	100.0%	95.2%	100.0%	93.3%	100.0%	99.4%	100.0%	97.3%
	32768	97.3%	100.0%	93.0%	100.0%	92.0%	100.0%	98.3%	100.0%	96.6%	100.0%	98.0%	100.0%	96.3%	100.0%	100.0%	100.0%	99.1%
	65536	97.4%	100.0%	93.9%	100.0%	92.1%	100.0%	98.3%	100.0%	96.9%	100.0%	98.0%	100.0%	96.6%	100.0%	100.0%	100.0%	99.1%
	131072	97.6%	100.0%	94.2%	100.0%	92.6%	100.0%	98.2%	100.0%	96.7%	100.0%	98.6%	100.0%	97.0%	100.0%	100.0%	100.0%	99.0%
	262144	97.8%	100.0%	94.9%	100.0%	93.5%	100.0%	98.7%	100.0%	96.9%	100.0%	98.4%	100.0%	97.4%	100.0%	100.0%	100.0%	99.0%
	524288	97.9%	100.0%	95.8%	100.0%	93.5%	100.0%	98.5%	100.0%	96.6%	100.0%	99.1%	100.0%	97.1%	100.0%	100.0%	100.0%	100.0%
	688739	98.3%	100.0%	96.2%	100.0%	94.4%	100.0%	99.1%	100.0%	97.1%	100.0%	100.0%	100.0%	97.3%	100.0%	100.0%	100.0%	100.0%

Table 3. Drivers of false discoveries

NOTE. This table reports the coefficients estimated for the following linear and logit regression models:

$$\hat{X} = \alpha + \beta\hat{Y} + \gamma\hat{Y}\hat{\zeta} + \delta\hat{Y}\hat{\sigma} + \varepsilon$$

$$\mathbb{P}(\hat{X} \leq 0.05) = \frac{1}{1 + e^{-(\alpha+\beta\hat{Y}+\gamma\hat{Y}\hat{\zeta}+\delta\hat{Y}\hat{\sigma}+\varepsilon)}}$$

*where $\hat{X}$ is the test p-value, obtained either from the RC test when the statistic is not standardized or from the SPA test when the statistic is standardized (for each data sample and each methodological choice, we test the luckiest rule and vary the number of additional rules from which bootstrapped excess returns are obtained, this resulting in multiple p-values on the same sample; here, we only report the results based on the p-value obtained when no additional rule is considered–denoted as PVAL1–and the p-value obtained when all other 688,739 trading rules are considered–denoted as PVAL21); $\hat{Y}$ is a vector of dummy explanatory variables representing possible methodological choices (considering trading fees-FEE; considering liquidity costs-LIQCOST; considering only long trades-LONG_ONLY; or standardizing the test statistic-STDSTAT), taking the value of 1 if used and 0 otherwise; $\hat{\zeta}$ is the sample average return; $\hat{\sigma}$ is the sample volatility; β, γ, δ are vectors of coefficients, while ε is the error term. Each regression is estimated on 96,000 data points, which are obtained by running the simulation exercise described in Section 3.1 on 6000 data samples with 16 possible methodological choices. T-statistics are reported in square parenthesis, while ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.*

Linear Model			Logit Model		
Dependent variable ($\hat{X}$):	PVAL1	PVAL21	Dependent variable ($\hat{X}$):	PVAL1	PVAL21
α	0.0820***	0.3578***	α	0.0309	-0.7030***
	[49.95]	[97.23]		[0.48]	[-4.52]
β_{FEE}	0.1096***	0.3057***	β_{FEE}	-1.9905***	-2.2644***
	[74.62]	[92.88]		[-34.34]	[-14.73]
β_{LIQCOST}	0.0486***	0.1251***	β_{LIQCOST}	-0.3865***	-0.7287***
	[33.08]	[38.01]		[-6.71]	[-5.08]
$\beta_{\text{LONG\_ONLY}}$	0.0069***	0.0510***	$\beta_{\text{LONG\_ONLY}}$	0.1726***	-0.2684**
	[4.74]	[15.50]		[3.01]	[-1.99]
β_{STDSTAT}	-0.0704***	-0.0021	β_{STDSTAT}	0.4458***	-0.3403**
	[-47.95]	[-0.65]		[7.71]	[-2.39]
γ	0.0635***	0.2364***	γ	-3.6208***	-7.0374***
	[120.33]	[199.86]		[-77.59]	[-46.85]
γ_{FEE}	0.0258***	0.0077***	γ_{FEE}	-0.1274***	-0.1408
	[54.68]	[7.31]		[-3.49]	[-1.45]
γ_{LIQCOST}	0.0210***	0.0082***	γ_{LIQCOST}	-0.1754***	-0.2155**
	[44.47]	[7.76]		[-4.86]	[-2.27]
$\gamma_{\text{LONG\_ONLY}}$	0.0000	0.0006	$\gamma_{\text{LONG\_ONLY}}$	0.2278***	1.0572***
	[0.16]	[0.63]		[6.58]	[11.80]
γ_{STDSTAT}	-0.0410***	-0.0343***	γ_{STDSTAT}	0.9490***	3.4017***
	[-86.92]	[-32.50]		[24.18]	[26.32]
δ	-0.0160***	0.0109	δ	0.8615***	-37.1855***
	[-2.79]	[0.85]		[3.58]	[-34.98]
δ_{FEE}	-0.2185***	-0.5302***	δ_{FEE}	0.8652***	-0.0394
	[-42.65]	[-46.21]		[4.02]	[-0.05]
δ_{LIQCOST}	-0.0436***	-0.0079	δ_{LIQCOST}	-3.9339***	-5.1633***
	[-8.51]	[-0.69]		[-18.16]	[-6.54]
$\delta_{\text{LONG\_ONLY}}$	-0.0191***	-0.0985***	$\delta_{\text{LONG\_ONLY}}$	0.0227	7.9359***
	[-3.74]	[-8.59]		[0.10]	[10.98]
δ_{STDSTAT}	0.0458***	-0.0421***	δ_{STDSTAT}	5.5390***	21.4303***
	[8.95]	[-3.67]		[25.31]	[22.75]
Adjusted R^2	0.5685	0.7023	Pseudo-R^2	0.5120	0.6273
F-statistic	9035.40***	16181.15***	F-statistic	68075.23***	23730.20***

Third, and most importantly, we find that false discoveries are overwhelmingly concentrated on data samples that exhibit negative average returns, irrespective of methodological choices. When only the luckiest trading rule is tested, about 60% of false discoveries arise in downward trading markets. However, as the number of trading rules increases, we end up with over 98% of false discoveries arising when conditions are bearish, compared with no more than 2% arising when conditions are bullish. This asymmetry shows that the directional preference of lucky trading rules is the main driver of false discoveries in empirical tests, while spurious correlation is less important and becomes irrelevant when accounting for the data snooping efforts of others.

We complement the analysis by estimating several regression models that quantify how the different methodological choices and sample properties drive false discoveries[13]. The results are reported in Table 3. We focus on the results of the logit model, which enables us to make inferences about the probability of making false discoveries. On the one hand, when only the luckiest trading rule is tested (the data snooping efforts of others is not accounted for), we find that there is a 50% default probability of rejecting the null, this increasing when the statistic is standardized (+11%) and decreasing when trading fees (-38%), liquidity costs (-10%), or short trades are considered (-4%). Also, we find that the probability of rejecting the null is strongly (negatively) determined by market returns, showing that the *directional preference* of lucky trading rules is one of the main drivers of false discoveries. All other things equal, a 1% per year decline in sample average returns increases the probability of rejecting the null by about 0.5%. This effect is slightly stronger when trading costs and short trades are considered, but weaker when the statistic is standardized. False discoveries are also positively related to volatility but to a lesser extent, showing that *spurious correlation* is less important in single hypothesis tests. Further, this effect is stronger when subtracting trading fees or standardizing the test statistic but weaker when subtracting liquidity costs. The latter result shows that adjusting for liquidity costs can help reduce the spurious correlation effect.

On the other hand, when all of the trading rules in the universe are accounted for, the default probability of rejecting the null drops to about 33% and can be reduced further with various methodological choices, such as subtracting trading fees (-28%) and liquidity costs (-14%). Now, standardizing the test statistic also helps reduce false discoveries (-7%), while considering short trades increases them (+6%).

[13] We only report the results from the 4 most relevant models. In alternative specifications, we incorporate quadratic terms and/or samples skewness and kurtosis as additional explanatory variables, finding mostly weak and/or insignificant interactions.

Making the most conservative choices would result in a net decrease of -32% (-96% in relative terms) in the probability of rejecting the null, which would reach only 1.33%. However, these baseline results do not account for the influence of sample properties, which seems to be very important. Complementing the discussion based on the results in Table 2, we find that market returns have a strong negative influence on null rejections, showing that false discoveries significantly increase in downward trending markets. The influence of market returns on false discoveries can be reduced by standardizing the test statistic or disregarding short trades, although the latter adjustment should only be performed in markets where short trades are not functional, otherwise a loss of information would occur. Surprisingly, we also find a highly significant negative impact of volatility on test null rejections (δ has a t-stat of -34.98), a result which goes against previous observations. Because the sign is reversed compared with the baseline model (estimated for PVAL1), the results shows that higher volatility does contribute to the spurious correlation effect, but that it has a higher, asymmetric impact on the trading rules in the tested set and, thus, on the null distribution of excess returns, compared to the excess returns of the luckiest trading rules. More specifically, when volatility increases, the distribution and corresponding critical values increase at a higher rate compared to the test statistic, thus decreasing the probability of making false discoveries. This result is very important, as it implies that (not) controlling for the data snooping efforts of others has a higher impact on test results performed in more volatile markets. Nevertheless, we observe that standardizing the test statistic eliminates more than half of the influence of volatility on null rejections, which reinforces the conclusion that this testing strategy should be used when volatility is expected to be high.

Overall, our results show the complex inner workings of how data snooping bias could arise in empirical tests. Our key, novel findings are as follows: (1) Data snooping bias in empirical tests concerned with the overperformance of trading rules is highly dependent on methodological choices and sample properties, especially market returns and volatility. Different testing conditions change the balance between the three different effects that alter testing results (market preference, spurious correlation, and trading costs) and can lead to significantly different outcomes. (2) Accounting for the data snooping of others–i.e. considering all trading rules from which lucky rules are obtained–is the main condition for controlling false discoveries, but its efficacy can substantially vary from one test to another. In some conditions, it is not sufficient in providing adequate finite-sample control of data snooping bias. (3) Subtracting trading fees and liquidity costs constitutes a minimum requirement for the goal of obtaining unbiased results in tests that meet the aforementioned condition. On average, false discoveries decrease by -32% when subtracting trading fees and by -24% when considering liquidity costs[14]. However, trading costs also have an indirect impact on the other effects. In particular, when liquidity costs are considered, the *market preference* and *spurious correlation effects* are amplified. (4) Although allowing short trades inflates the excess performance of lucky trading rules, there is little difference in terms of false discoveries compared with the case when only long trades are possible–false discoveries actually decrease on average by -2% when short trades are allowed. (5) The effects of standardizing the test statistic depend on the volatility of trading signals, which is a function of market volatility and whether or not short trades are allowed. Standardizing the test statistic is effective for reducing false discoveries when volatility is high or in markets where short trades are functional, but is counterproductive otherwise. Also, standardizing the test statistic works only when testing a sufficient number of trading rules as to account for their luck. Otherwise, in single hypothesis tests or in tests that consider few trading rules, using a simple statistic would be more effective in reducing data snooping bias. (6) Most importantly, false discoveries are not eliminated even when employing conservative testing strategies. Our analysis shows that this result is due to the *directional preference effect* of lucky trading rules, which is the main driver of data snooping bias by a significant margin. Specifically, false discoveries are overwhelmingly concentrated on samples that exhibit negative average returns. Also, while accounting for the data snooping efforts of others eliminates false discoveries in upward trending markets (implying that it is effective at curbing the *spurious correlation effect*), it is far less effective when market returns have a negative tendency. (7) False discoveries are on average negatively associated with market volatility, even though this varies with specific testing conditions. This result contradicts the

[14] The values are calculated as one subtracted from the slope of the regression between average false discovery proportions obtained when trading costs are considered and average false discovery proportions obtained when trading costs are ignored. The regression is performed for the subset of tests reported in Panel A of Table 2.

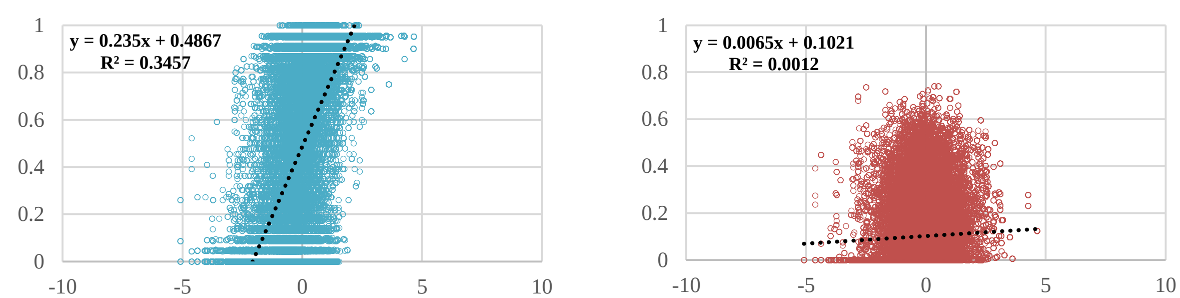
conclusion of Anghel (2020a), which only performed tests with trading costs enables and short trades disabled. This result highlights that methodological choices should be tailored to the expected market conditions. For example, the influence of volatility on false discoveries can be reduced by ignoring short trades or standardizing the tests statistic.

4. The directional preference effect

The results reported in Section 3 show that the *directional preference effect* is the main driver of false discoveries in empirical tests concerned with the overperformance of trading rules. They also show that the effect persists even after accounting for the set of trading rules from which lucky rules are selected. In this Section, we perform an in-depth investigation of its characteristics. First, we try to replicate previous results by testing suitably constructed alternatives rules that are not derived from technical analysis. Specifically, we consider a set of $C = 10{,}000$ naïve trading rules whose signal function follows $\delta_c \sim Bernoulli(p)$, where $p = 0.5$ is selected to enforce a balanced directional preference that maximizes the *spurious correlation effect* in sideways trending markets. The results are reported in Table 4 and Figure 2. On the one hand, we find that the directional preference of naïve trading rules strongly depends on market returns, while spurious correlation does not. On the other hand, we find non-negligible amounts of false discoveries that are concentrated on samples exhibiting negative average returns. Taken together, the results imply that previous results and conclusions are not unique to technical trading rules, but instead can arise for all lucky trading strategies that rely on a signal function to deliver buy and sell signals that help investors time individual assets.

Figure 2. Data sample properties and characteristics of naïve trading rules

NOTE. This figure plots some characteristics of lucky naïve trading rules (y-axis) against average market returns (x-axis).



Panel A. Directional preference *Panel B. Spurious correlation*

Table 4. Proportion of False Discoveries–tests on samples with negative average returns–only naïve trading rule

NOTE. This table reports the proportion of false discoveries (a proxy for the Familywise error rate–FWER) generated by naïve trading rules on randomly generated data, estimated for a significance level of 5% when varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). Each value is aggregated from 3,039 data points, corresponding to the same number of samples on which average market returns are negative; except for the values in the 'Both' column, which are based on 48,624 results. Values that fall below the 5% threshold signal a test that offers adequate in-sample control of FWER, and are highlighted using ***bolded*** *text. The values in parenthesis represent the relative contribution of negative samples to the total number of false discoveries.*

FEE		Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIQCOST		Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
LONGONLY		Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
STDSTAT		Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules	1	0.086	0.127	0.166	0.097	0.170	0.084	0.104	***0.009***	0.098	0.079	0.100	***0.008***	0.095	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.6%)	(93.6%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	2	0.086	0.127	0.166	0.097	0.170	0.084	0.104	***0.009***	0.098	0.079	0.100	***0.008***	0.095	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.6%)	(93.6%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	4	0.086	0.127	0.166	0.097	0.170	0.084	0.104	***0.009***	0.098	0.079	0.100	***0.008***	0.095	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.6%)	(93.6%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	8	0.085	0.126	0.165	0.097	0.169	0.084	0.104	***0.009***	0.098	0.079	0.100	***0.008***	0.094	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.5%)	(93.5%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	16	0.085	0.125	0.165	0.097	0.169	0.084	0.104	***0.009***	0.098	0.079	0.100	***0.008***	0.094	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.5%)	(93.5%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	32	0.085	0.124	0.165	0.096	0.169	0.084	0.104	***0.009***	0.098	0.078	0.100	***0.008***	0.094	0.066	0.085	***0.001***	0.082
		(97.4%)	(98.5%)	(93.5%)	(99.8%)	(92.5%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.2%)	***(100.0%)***	(98.1%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	64	0.085	0.123	0.165	0.096	0.169	0.083	0.103	***0.009***	0.097	0.078	0.099	***0.008***	0.094	0.066	0.084	***0.001***	0.081
		(97.4%)	(98.8%)	(93.5%)	(99.8%)	(92.6%)	(100.0%)	(97.4%)	***(100.0%)***	(97.3%)	(100.0%)	(98.1%)	***(100.0%)***	(98.0%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	128	0.084	0.117	0.164	0.096	0.169	0.082	0.102	***0.009***	0.097	0.077	0.098	***0.008***	0.093	0.065	0.083	***0.001***	0.081
		(97.4%)	(98.7%)	(93.5%)	(99.8%)	(92.6%)	(100.0%)	(97.4%)	***(100.0%)***	(97.2%)	(100.0%)	(98.1%)	***(100.0%)***	(98.0%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	256	0.082	0.109	0.162	0.095	0.169	0.078	0.100	***0.009***	0.096	0.074	0.095	***0.007***	0.093	0.063	0.080	***0.001***	0.078
		(97.4%)	(98.9%)	(93.6%)	(99.8%)	(92.7%)	(100.0%)	(97.3%)	***(100.0%)***	(97.2%)	(100.0%)	(98.2%)	***(100.0%)***	(98.0%)	(100.0%)	(99.4%)	***(100.0%)***	(99.4%)
	512	0.079	0.098	0.160	0.094	0.168	0.075	0.097	***0.009***	0.094	0.070	0.092	***0.007***	0.090	0.061	0.077	***0.001***	0.075
		(97.4%)	(99.5%)	(93.6%)	(99.8%)	(92.7%)	(100.0%)	(97.4%)	***(100.0%)***	(97.2%)	(100.0%)	(98.4%)	***(100.0%)***	(98.2%)	(100.0%)	(99.3%)	***(100.0%)***	(99.3%)
	1024	0.075	0.090	0.155	0.093	0.165	0.070	0.092	***0.008***	0.092	0.067	0.086	***0.006***	0.087	0.056	0.070	***0.001***	0.070
		(97.5%)	(99.8%)	(93.9%)	(99.8%)	(92.9%)	(100.0%)	(97.3%)	***(100.0%)***	(97.3%)	(100.0%)	(98.5%)	***(100.0%)***	(98.1%)	(100.0%)	(99.5%)	***(100.0%)***	(99.5%)
	2048	0.069	0.084	0.145	0.088	0.159	0.065	0.082	***0.007***	0.085	0.062	0.076	***0.006***	0.079	0.051	0.060	***0.001***	0.061
		(97.7%)	(100.0%)	(94.0%)	(100.0%)	(93.3%)	(100.0%)	(97.8%)	***(100.0%)***	(97.5%)	(100.0%)	(98.7%)	***(100.0%)***	(98.3%)	(100.0%)	(99.7%)	***(100.0%)***	(99.7%)
	4096	0.062	0.080	0.129	0.080	0.146	0.061	0.067	***0.006***	0.074	0.058	0.063	***0.005***	0.069	***0.047***	***0.049***	***0.001***	0.051
		(97.8%)	(100.0%)	(93.9%)	(100.0%)	(93.9%)	(100.0%)	(98.3%)	***(100.0%)***	(98.0%)	(100.0%)	(98.4%)	***(100.0%)***	(98.5%)	***(100.0%)***	***(99.7%)***	***(100.0%)***	(100.0%)
	8192	***0.047***	0.078	0.086	0.069	0.094	0.060	***0.047***	***0.006***	0.051	0.057	***0.044***	***0.005***	***0.047***	***0.046***	***0.034***	***0.001***	***0.037***
		(98.4%)	(100.0%)	(94.9%)	(100.0%)	(93.9%)	(100.0%)	***(98.9%)***	***(100.0%)***	(98.4%)	(100.0%)	***(99.2%)***	***(100.0%)***	***(98.9%)***	***(100.0%)***	***(100.0%)***	***(100.0%)***	***(100.0%)***
	10000	***0.041***	0.078	0.063	0.068	0.068	0.060	***0.037***	***0.006***	***0.037***	0.057	***0.036***	***0.005***	***0.037***	***0.046***	***0.031***	***0.001***	***0.032***
		(99.2%)	(100.0%)	(97.6%)	(100.0%)	(96.8%)	(100.0%)	***(99.1%)***	***(100.0%)***	***(98.6%)***	(100.0%)	***(99.1%)***	***(100.0%)***	***(99.1%)***	***(100.0%)***	***(100.0%)***	***(100.0%)***	***(100.0%)***

Table 5. Proportion of False Discoveries–tests on samples with negative average returns–technical & naïve trading rule

NOTE. This table reports the proportion of false discoveries (a proxy for the Familywise error rate–FWER) generated when testing technical and naïve trading rules together on randomly generated data, estimated for a significance level of 5% when varying the testing conditions (considering trading fees–FEE; considering liquidity costs–LIQCOST; only considering long trades–LONG_ONLY; or standardizing the test statistic–STDSTAT). Each value is aggregated from 3,039 data points, corresponding to the same number of samples on which average market returns are negative; except for the values in the 'Both' column, which are based on 48,624 results. Values that fall below the 5% threshold signal a test that offers adequate in-sample control of FWER, and are highlighted using ***bolded*** *text. The values in parenthesis represent the relative contribution of negative samples to the total amount of false discoveries.*

	FEE	Both	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	LIQCOST	Both	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
	LONGONLY	Both	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
	STDSTAT	Both	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of trading rules	1	0.361	0.813	0.555	0.167	0.508	0.626	0.383	0.088	0.421	0.568	0.339	0.074	0.372	0.397	0.202	***0.031***	0.235
		(95.4%)	(95.7%)	(89.9%)	(100.0%)	(87.7%)	(98.9%)	(94.7%)	(100.0%)	(93.7%)	(99.2%)	(95.6%)	(100.0%)	(94.4%)	(99.9%)	(99.2%)	***(100.0%)***	(98.1%)
	2	0.344	0.734	0.554	0.167	0.508	0.546	0.381	0.088	0.420	0.503	0.337	0.073	0.371	0.354	0.201	***0.031***	0.233
		(95.5%)	(96.5%)	(90.0%)	(100.0%)	(87.8%)	(99.2%)	(94.8%)	(100.0%)	(93.7%)	(99.2%)	(95.5%)	(100.0%)	(94.4%)	(99.9%)	(99.2%)	***(100.0%)***	(98.1%)
	4	0.325	0.645	0.552	0.167	0.506	0.470	0.378	0.088	0.418	0.426	0.334	0.073	0.370	0.308	0.199	***0.031***	0.233
		(95.6%)	(97.7%)	(90.1%)	(100.0%)	(87.8%)	(99.5%)	(94.8%)	(100.0%)	(93.8%)	(99.6%)	(95.8%)	(100.0%)	(94.4%)	(100.0%)	(99.2%)	***(100.0%)***	(98.1%)
	8	0.314	0.600	0.549	0.167	0.503	0.433	0.374	0.088	0.411	0.390	0.329	0.073	0.364	0.284	0.196	***0.031***	0.229
		(95.6%)	(98.2%)	(90.2%)	(100.0%)	(87.7%)	(99.6%)	(95.0%)	(100.0%)	(93.8%)	(99.7%)	(96.0%)	(100.0%)	(94.5%)	(100.0%)	(99.2%)	***(100.0%)***	(98.2%)
	16	0.308	0.580	0.545	0.167	0.502	0.420	0.367	0.088	0.407	0.377	0.325	0.073	0.359	0.276	0.190	***0.031***	0.225
		(95.6%)	(98.5%)	(90.2%)	(100.0%)	(87.8%)	(99.8%)	(95.1%)	(100.0%)	(93.9%)	(99.7%)	(95.9%)	(100.0%)	(94.5%)	(100.0%)	(99.1%)	***(100.0%)***	(98.1%)
	32	0.304	0.563	0.542	0.167	0.500	0.408	0.363	0.088	0.406	0.366	0.320	0.073	0.355	0.271	0.187	***0.031***	0.222
		(95.6%)	(98.7%)	(90.2%)	(100.0%)	(87.9%)	(99.8%)	(95.0%)	(100.0%)	(93.8%)	(99.7%)	(96.0%)	(100.0%)	(94.4%)	(100.0%)	(99.1%)	***(100.0%)***	(98.1%)
	64	0.279	0.418	0.542	0.167	0.500	0.312	0.363	0.088	0.404	0.285	0.318	0.073	0.354	0.209	0.185	***0.031***	0.220
		(95.4%)	(100.0%)	(90.2%)	(100.0%)	(87.8%)	(100.0%)	(95.0%)	(100.0%)	(93.8%)	(100.0%)	(95.9%)	(100.0%)	(94.4%)	(100.0%)	(99.1%)	***(100.0%)***	(98.1%)
	128	0.279	0.418	0.542	0.167	0.500	0.312	0.363	0.088	0.404	0.285	0.318	0.073	0.354	0.209	0.185	***0.031***	0.220
		(95.4%)	(100.0%)	(90.2%)	(100.0%)	(87.8%)	(100.0%)	(95.0%)	(100.0%)	(93.8%)	(100.0%)	(95.9%)	(100.0%)	(94.4%)	(100.0%)	(99.1%)	***(100.0%)***	(98.1%)
	256	0.272	0.418	0.522	0.167	0.488	0.311	0.347	0.088	0.393	0.284	0.299	0.073	0.341	0.208	0.176	***0.031***	0.211
		(95.7%)	(100.0%)	(90.6%)	(100.0%)	(88.3%)	(100.0%)	(95.2%)	(100.0%)	(94.2%)	(100.0%)	(96.0%)	(100.0%)	(94.3%)	(100.0%)	(99.1%)	***(100.0%)***	(98.6%)
	512	0.222	0.312	0.438	0.167	0.444	0.224	0.275	0.087	0.323	0.204	0.237	0.073	0.278	0.151	0.137	***0.030***	0.175
		(96.2%)	(100.0%)	(91.9%)	(100.0%)	(89.9%)	(100.0%)	(96.3%)	(100.0%)	(95.2%)	(100.0%)	(96.6%)	(100.0%)	(95.4%)	(100.0%)	(99.5%)	***(100.0%)***	(98.9%)
	1024	0.204	0.259	0.419	0.167	0.439	0.172	0.257	0.083	0.308	0.159	0.222	0.071	0.269	0.113	0.127	***0.029***	0.167
		(96.4%)	(100.0%)	(92.3%)	(100.0%)	(90.5%)	(100.0%)	(96.5%)	(100.0%)	(95.9%)	(100.0%)	(97.3%)	(100.0%)	(95.9%)	(100.0%)	(99.5%)	***(100.0%)***	(99.2%)
	2048	0.203	0.253	0.419	0.167	0.439	0.170	0.257	0.083	0.308	0.158	0.222	0.071	0.269	0.111	0.127	***0.029***	0.167
		(96.4%)	(100.0%)	(92.3%)	(100.0%)	(90.5%)	(100.0%)	(96.5%)	(100.0%)	(95.9%)	(100.0%)	(97.3%)	(100.0%)	(95.9%)	(100.0%)	(99.5%)	***(100.0%)***	(99.2%)
	4096	0.198	0.241	0.406	0.167	0.434	0.164	0.249	0.083	0.302	0.152	0.215	0.070	0.266	0.107	0.124	***0.029***	0.162
		(96.5%)	(100.0%)	(92.4%)	(100.0%)	(90.8%)	(100.0%)	(96.9%)	(100.0%)	(96.1%)	(100.0%)	(97.5%)	(100.0%)	(96.0%)	(100.0%)	(99.5%)	***(100.0%)***	(99.2%)
	8192	0.170	0.222	0.335	0.166	0.380	0.151	0.196	0.082	0.253	0.135	0.168	0.068	0.217	0.097	0.093	***0.027***	0.129
		(97.2%)	(100.0%)	(94.0%)	(100.0%)	(92.0%)	(100.0%)	(98.3%)	(100.0%)	(96.7%)	(100.0%)	(98.3%)	(100.0%)	(96.9%)	(100.0%)	(99.6%)	***(100.0%)***	(99.2%)
	16384	0.165	0.208	0.330	0.166	0.376	0.138	0.191	0.081	0.248	0.124	0.160	0.068	0.211	0.092	0.090	***0.027***	0.124
		(97.4%)	(100.0%)	(94.4%)	(100.0%)	(92.3%)	(100.0%)	(98.6%)	(100.0%)	(97.4%)	(100.0%)	(98.6%)	(100.0%)	(97.1%)	(100.0%)	(99.6%)	***(100.0%)***	(99.2%)
	32768	0.124	0.177	0.237	0.164	0.309	0.121	0.112	0.074	0.162	0.109	0.094	0.064	0.133	0.077	0.052	***0.024***	0.072
		(98.6%)	(100.0%)	(97.4%)	(100.0%)	(95.6%)	(100.0%)	(99.1%)	(100.0%)	(98.6%)	(100.0%)	(99.0%)	(100.0%)	(98.3%)	(100.0%)	(100.0%)	***(100.0%)***	(99.5%)
	65536	0.126	0.179	0.242	0.165	0.313	0.123	0.116	0.075	0.168	0.110	0.097	0.064	0.137	0.078	0.054	***0.025***	0.075
		(98.6%)	(100.0%)	(96.8%)	(100.0%)	(95.5%)	(100.0%)	(99.2%)	(100.0%)	(98.6%)	(100.0%)	(99.0%)	(100.0%)	(98.3%)	(100.0%)	(100.0%)	***(100.0%)***	(99.6%)
	131072	0.120	0.173	0.229	0.164	0.302	0.116	0.109	0.073	0.155	0.105	0.090	0.064	0.130	0.075	***0.049***	***0.024***	0.068
		(98.7%)	(100.0%)	(97.6%)	(100.0%)	(95.9%)	(100.0%)	(99.1%)	(100.0%)	(98.5%)	(100.0%)	(99.3%)	(100.0%)	(98.5%)	(100.0%)	***(100.0%)***	***(100.0%)***	(99.5%)
	262144	0.113	0.162	0.213	0.163	0.286	0.108	0.103	0.070	0.142	0.096	0.083	0.060	0.122	0.067	***0.043***	***0.023***	0.064
		(98.8%)	(100.0%)	(97.7%)	(100.0%)	(96.0%)	(100.0%)	(99.4%)	(100.0%)	(98.4%)	(100.0%)	(99.2%)	(100.0%)	(98.7%)	(100.0%)	***(100.0%)***	***(100.0%)***	(99.5%)
	524288	0.106	0.157	0.200	0.159	0.269	0.105	0.089	0.068	0.132	0.092	0.075	0.058	0.114	0.063	***0.038***	***0.022***	0.054
		(98.8%)	(100.0%)	(97.9%)	(100.0%)	(96.2%)	(100.0%)	(99.3%)	(100.0%)	(98.3%)	(100.0%)	(99.6%)	(100.0%)	(98.6%)	(100.0%)	***(100.0%)***	***(100.0%)***	(100.0%)
	688739	0.094	0.143	0.173	0.154	0.242	0.099	0.073	0.062	0.113	0.086	0.060	0.053	0.097	0.056	***0.030***	***0.020***	***0.047***
		(99.0%)	(100.0%)	(98.1%)	(100.0%)	(96.5%)	(100.0%)	(99.5%)	(100.0%)	(98.6%)	(100.0%)	(100.0%)	(100.0%)	(98.7%)	(100.0%)	***(100.0%)***	***(100.0%)***	***(100.0%)***

Second, investigate methodological improvements that would possibly control the data snooping bias associated to the *directional preference effect* in downward trending markets. As the bearish tendency of lucky rules in periods of declining prices is the main factor leading to false discoveries, we consider supplementing the set of tested trading rules with suitably constructed alternative strategies that have the same directional preference as the luckiest rule in the sample. We use a set of $C = 20{,}000$ naïve trading rules whose signal function follows a Bernoulli distribution, this time with parameter $\hat{p}_m$, where m denotes the luckiest trading rules in a sample and $\hat{p}_m$ its directional preference–i.e., $\delta_c \sim Bernoulli(\hat{p}_m), c = 1..C, m = \underset{k=1..K}{\operatorname{argmax}} \hat{d}_k$. The testing procedure is modified to additionally bootstrap from the excess return series of all naïve alternatives, with the aim of improving the finite-sample properties of the null distribution and making it more robust to the influence of lucky directional preference. The results are reported in Table 4[15] and show that adding the set of naïve rules significantly decreases the amount of false discoveries in single hypothesis tests (i.e., when only the luckiest rule from the original set is considered), but do not have a material impact on false discoveries after accounting for the other technical trading rules from which lucky rules are selected. Specifically, compared to the benchmark results (reported in Panels C and D of Table 2), false discoveries decrease on average by 40.6 percentage points when only the luckiest technical trading rule is tested but only by 0.3 percentage points when all 688,739 rules are tested together. Also, the relative contribution of directional preference to data snooping is unchanged, while the adjustment has no material impact on results obtained when short trades are enabled. The results show that the *directional preference effect* is resilient to methodological strategies designed to eliminate it, implying that false discoveries may increase in empirical tests that use samples in which prices have a tendency to decline.

5. Lucky trading rules in the cryptocurrency market

We perform an empirical exercise to evaluate if the simulation results hold when considering real market conditions and also to show how our previous findings can be used by researchers and investment

[15] All other results follow similar patterns and do not report them (results are available upon request). We center on negative returns because this is where false discoveries predominantly appear.

professionals to account for lucky trading rules. We choose to investigate the cryptocurrency market, which is relevant for our analysis because it represents is relatively young financial markets that has been found to be "less efficient" than it's more traditional counterparts. Also, the analysis of trading rule overperformance in the cryptocurrency has revealed evidence in favor of their economically significant predictive ability. However, data snooping bias is not a concern of most studies, which begs the question: are trading rules really profitable in the cryptocurrency market after adjusting for luck? In this Section, we perform such a data-snooping-free evaluation of technical trading rule performance in this market. Anghel (2020b) performed a similar analysis but only tested 67,480 rules based on technical analysis and 5 optimized rules based on machine learning techniques. Here, we expand the investigation by testing the entire set of 688,739 trading rules defined previously. More importantly, we not only test statistical significance, but also evaluate economic significance by comparing test results with the results obtained on randomly generated data, which constitute our benchmark. If the rates of positive discoveries (test null rejections) exceeded what we might expect given the bounds of randomness, than trading rules can be deemed as economically significant. Moreover, we decompose the results by sample average return to evaluate the role played by directional preference and signal-to-market correlation in shaping the results. The data sample is obtained from Anghel (2020b) and consists of trading histories for 861 cryptocurrencies collected on February 10, 2020 from www.coinmarketcap.com (sample statistics are reported in the Data Appendix provided by Anghel, 2020b). We split the data by calendar month and perform tests on monthly samples of at least 22 observations. On average, we investigate roughly 37 months of trading activity for each cryptoasset, leading to a total of 32,014 tests.

Our methodological choices are tailored to the general conditions encountered by investors when trading in this market. First, trading fees vary by broker but can be as low as 0%, so we decide not to consider them. Second, price impact costs should be important in a relatively young and thinly traded market, so we decide to incorporate them. Third, short trades are possible, but only when opening accounts at some brokers and mostly for top-tier assets. In general, we assess that short trades cannot currently be considered as fully functional in the cryptocurrency market, so we decide to ignore them. Finally, we choose to employ both a simple and a standardized statistic. The cryptocurrency market is on average more volatile compared to its competitors, which suggests that a standardized statistic should be used in principle. However, because we only consider long positions, we expect that a simple statistic would better control for false discoveries.

The results for the empirical exercise are reported in Table 10. When considering all samples, we find that the tests performed in the cryptocurrency market reject their null hypotheses 50%-90% less often compared to the benchmark, implying that technical trading rules do not generally have a superior predictive ability in this market after adjusting for data snooping. A similar result is obtained on the subset of samples exhibiting negative average returns, implying that technical trading rules have especially no use in periods of declining cryptocurrency prices. Interestingly, testing few rules–including using single hypothesis tests–results in more null rejections on bearish samples compared with bullish samples, which is consistent with our earlier results obtained in the simulation exercise. However, increasing the number of rules–including testing all 688,739 rules– results in more null rejections on bullish samples. This implies that false discoveries arising from the *directional preference effect* are well controlled after adjusting for data snooping in the cryptocurrency market, while some positive discoveries linked to signal-to-market correlation could have some merit.

Table 10. Positive discoveries in the cryptocurrency market

NOTE. This table reports the proportion of positive discoveries–i.e. test null rejections (at the 5% significance level) divided by the total number of tests performed–for tests evaluating technical trading rules in the cryptocurrency market. In total, 32,014 tests are performed on data samples containing between 22 and 31 observations, out of which 18,809 are bearish ($\hat{\zeta} < 0$) and 13,205 are bullish ($\hat{\zeta} \geq 0$). Prop($\hat{\zeta} < 0$) *denotes the ratio between null rejections obtained on bearish markets and the total number of null rejections. The benchmark results are extracted from Table 2. Null rejection rates above the randomness benchmark are highlighted in **bolded** text.*

Panel A. Basic statistic (RC test)

Cryptocurrency data					Randomly generated data (benchmark)				
# TTRs	$\hat{\zeta} \in \mathbb{R}$	$\hat{\zeta} < 0$	$\hat{\zeta} \geq 0$	Prop($\hat{\zeta} < 0$)	# TTRs	$\hat{\zeta} \in \mathbb{R}$	$\hat{\zeta} < 0$	$\hat{\zeta} \geq 0$	Prop($\hat{\zeta} < 0$)
1	0.102	0.164	0.012	95.0%	1	0.386	0.711	0.052	86.4%
2	0.087	0.141	0.011	94.8%	2	0.323	0.611	0.028	91.2%
4	0.071	0.113	0.010	94.2%	4	0.263	0.509	0.010	96.3%
8	0.055	0.087	***0.009***	93.0%	8	0.219	0.430	0.003	98.8%
16	0.045	0.070	***0.009***	92.1%	16	0.199	0.394	0.000	100.0%
32	0.039	0.060	***0.008***	91.4%	32	0.195	0.385	0.000	100.0%
64	0.029	0.044	***0.008***	89.0%	64	0.171	0.338	0.000	100.0%
128	0.029	0.044	***0.008***	89.0%	128	0.171	0.338	0.000	100.0%
256	0.028	0.042	***0.008***	88.6%	256	0.168	0.331	0.000	100.0%
512	0.016	0.022	***0.008***	80.7%	512	0.120	0.237	0.000	100.0%
1024	0.008	0.009	***0.008***	62.5%	1024	0.087	0.172	0.000	100.0%
2048	0.008	0.008	***0.008***	61.4%	2048	0.086	0.170	0.000	100.0%

4096	0.008	0.008	***0.008***	59.3%	4096	0.083	0.165	0.000	100.0%
8192	0.006	0.004	***0.007***	44.6%	8192	0.071	0.141	0.000	100.0%
16384	0.005	0.003	***0.007***	36.2%	16384	0.066	0.130	0.000	100.0%
32768	0.004	0.002	***0.007***	23.0%	32768	0.052	0.103	0.000	100.0%
65536	0.004	0.001	***0.007***	20.5%	65536	0.051	0.101	0.000	100.0%
131072	0.004	0.001	***0.007***	19.8%	131072	0.049	0.097	0.000	100.0%
262144	0.004	0.001	***0.007***	15.7%	262144	0.043	0.084	0.000	100.0%
524288	0.004	0.001	***0.007***	14.9%	524288	0.040	0.078	0.000	100.0%
688739	0.003	0.001	***0.007***	11.8%	688739	0.033	0.065	0.000	100.0%

Panel B. Standardized statistic (SPA test)

Cryptocurrency data					Randomly generated data (benchmark)				
# TTRs	$\hat{\zeta} \in \mathbb{R}$	$\hat{\zeta} < 0$	$\hat{\zeta} \geq 0$	Prop($\hat{\zeta} < 0$)	# TTRs	$\hat{\zeta} \in \mathbb{R}$	$\hat{\zeta} < 0$	$\hat{\zeta} \geq 0$	Prop($\hat{\zeta} < 0$)
1	0.303	0.452	0.090	87.7%	1	0.615	0.917	0.306	50.3%
2	0.283	0.424	0.082	88.1%	2	0.575	0.884	0.258	55.1%
4	0.256	0.383	0.073	88.1%	4	0.528	0.838	0.210	60.2%
8	0.226	0.340	0.064	88.4%	8	0.492	0.803	0.173	64.8%
16	0.196	0.297	0.052	89.1%	16	0.475	0.782	0.160	66.3%
32	0.170	0.259	0.043	89.5%	32	0.466	0.769	0.156	66.6%
64	0.148	0.226	0.036	90.0%	64	0.454	0.752	0.148	67.5%
128	0.148	0.226	0.036	90.0%	128	0.454	0.752	0.148	67.5%
256	0.130	0.199	0.031	90.3%	256	0.416	0.699	0.126	69.7%
512	0.068	0.103	0.017	89.5%	512	0.268	0.480	0.051	81.0%
1024	0.051	0.078	0.014	88.6%	1024	0.227	0.415	0.035	84.5%
2048	0.051	0.077	0.014	88.7%	2048	0.227	0.414	0.035	84.7%
4096	0.048	0.072	0.013	88.3%	4096	0.216	0.397	0.030	85.9%
8192	0.036	0.053	0.012	86.1%	8192	0.157	0.296	0.014	91.0%
16384	0.031	0.045	0.012	84.2%	16384	0.151	0.286	0.012	91.7%
32768	0.016	0.020	***0.011***	71.9%	32768	0.090	0.174	0.003	96.6%
65536	0.016	0.019	***0.011***	71.3%	65536	0.087	0.169	0.003	96.9%
131072	0.015	0.018	***0.011***	69.3%	131072	0.083	0.161	0.003	96.7%
262144	0.014	0.016	***0.011***	67.2%	262144	0.076	0.148	0.002	96.9%
524288	0.013	0.014	***0.011***	65.1%	524288	0.070	0.135	0.002	96.6%
688739	0.012	0.012	***0.011***	61.3%	688739	0.058	0.113	0.002	97.1%

A detailed analysis reveals 97 instances when the RC null hypothesis is rejected on samples with positive average returns, which constitutes 0.7% of all bullish samples (0.3% of all samples). This is higher than the proportion estimated on randomly generated data, which was 0%. Similarly, there are 146 instances when the SPA null is rejected on samples with positive average returns, which constitutes 1.1% of all bullish samples (0.4% of all samples), again being higher than the proportion estimated on randomly generated data, which was 0.2%. Null rejections decline when testing more rules, implying that a substantial amount of false discoveries due to the *spurious correlation effect* can be eliminated when controlling for data snooping. However, the differences of 0.3% for the RC test and 0.9% for the SPA test constitute evidence in favor of technical trading rules being capable of successfully timing cryptocurrency prices and earning statistically significant excess returns in some periods of increasing prices. Nevertheless, we consider a 0.3%-0.9% success rate to be insufficient, as it would not encourage most investors to use trading rules to make real investment decisions. As a result, and taking into account the entirety of results, we conclude that speculative trading rules cannot be considered as having an economically significant superior predictive ability on the cryptocurrency market. This conclusion contradicts most of the previous findings reported in the literature, implying that data snooping in related tests would be the main factor driving the result.

6. Conclusions

While some discussions exist on how to design and conduct a proper empirical exercise when examining the profitability of speculative trading strategies, especially in the era of machine learning (e.g., Arnott, Harvey, and Markowitz, 2019), applied researchers often go their own, subjective way. The lack of evidence regarding how each methodological choice leads to false discoveries does not help in encouraging a homogenous approach. This increases the risk that positive findings reported in the literature may be biased due to data snooping. Although the effects of ignoring trading costs or the data snooping efforts of others have been studied before, a quantitative assessments of their exact impact in terms of false discoveries has rarely been provided. Also alternative sources of false discoveries have not been examined.

Besides showing that false discoveries are driven by several methodological choices–including the use (or not) of short trades, the standardization (or not) of the test statistic, or the choice for a particular data sample–our analysis reveals the contribution that each factor has on data snooping bias. Our main finding is that up to 100% of false discoveries can be attributed to a bearish directional tendency of lucky trading rules in periods of declining prices. While the effects of spurious correlation are controlled when using adequate multiple testing procedure, the same is not true for directional tendency. This *directional preference effect* is not unique to trading rules based on technical analysis but instead can arise when testing any set of speculative strategies that output trading signals to investors. Moreover, we find that the *directional preference effect* is very resilient to methodological adjustments designed to eliminate it, implying that it may be very prevalent in empirical tests.

We also perform an empirical exercise to demonstrate how our results can be applied in tests examining speculative trading strategies. We focus on the emerging cryptocurrency market and tune our methodological choices to adequately account for luck. On the one hand, the results show that the *directional preference effect* is important, especially when the number of trading rules being tested is small. On the other hand, we find that the proportion of null rejections obtained in multiple hypothesis tests does not exceed the bounds of luckiness, with the exception of less than 1% of samples on which market prices had a tendency to increase. From a statistical perspective, this shows that the cryptocurrency cannot be considered as informationally efficient. However, from an economic perspective, this shows that "abnormal" profit opportunities are insignificant and likely unattainable to investors that trade in this market. Overall, our results rather support the Efficient Market Hypothesis even for cryptocurrencies.

Overall, this paper shows that applied researchers should carefully tailor their methodological approaches to the testing problem that they face, so that they minimize the chances of making false discoveries. For example, a standardized the test statistic should be used when the volatility of trading signals is expected to be high–such as in thinly traded markets or in markets where short trades are functional–but can otherwise inflate the number of false discoveries. However, when selecting and using data sample(s) that lean bearish, additional precautions should be taken as to avoid data snooping bias. In this regard, as econometric tools that can explicitly handle such situations have not been discussed, future research endeavors should strive to develop and use novel testing frameworks that can handle the directional preference of lucky trading rules.

Acknowledgements

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-Pl-l.1-PD-2019-0016, within PNCDI III.

References

Alexander, S.S., 1961. Price movements in speculative markets: Trends or random walks. Industrial Management Review, 2(2), pp.25-46.

Al-Yahyaee, K.H., Mensi, W. and Yoon, S.M., 2018. Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. Finance Research Letters, 27, pp.228-234.

Anghel, D.G., 2020a. Data Snooping Bias in Tests of the Relative Performance of Multiple Forecasting Models. Unpublished manuscript, https://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2020-Dublin/papers/EFMA%202020_stage-1301_question-Full%20Paper_id-114.pdf.

Anghel, D.G., 2020b. A reality check on trading rule performance in the cryptocurrency market: Machine learning vs. technical analysis. Finance Research Letters, 101655.

Arnott, R., Harvey, C.R. and Markowitz, H., 2019. A backtesting protocol in the era of machine learning. The Journal of Financial Data Science, 1(1), pp.64-74.

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

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

Benjamini, Y. and 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), 57(1), pp.289-300.

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

Chordia, T., Goyal, A. and Saretto, A., 2020. Anomalies and false rejections. The Review of Financial Studies, 33(5), pp.2134-2179.

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

Corbet, S., Eraslan, V., Lucey, B. and Sensoy, A., 2019. The effectiveness of technical trading rules in cryptocurrency markets. Finance Research Letters, 31, pp.32-37.

Cowles 3rd, A., 1933. Can stock market forecasters forecast?. Econometrica: Journal of the Econometric Society, pp.309-324.

Dichtl, H., Drobetz, W., Neuhierl, A. and Wendt, V.S., 2020. Data snooping in equity premium prediction. International Journal of Forecasting, 37, pp.72-94.

Efron, B., 2012. Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press.

Fama, E.F., 1965. The behavior of stock-market prices. The Journal of Business, 38(1), pp.34-105.

Fischer, T.G., Krauss, C. and Deinert, A., 2019. Statistical arbitrage in cryptocurrency markets. Journal of Risk and Financial Management, 12(1), 31.

Grobys, K., Ahmed, S. and Sapkota, N., 2020. Technical trading rules in the cryptocurrency market. Finance Research Letters, 32, 101396.

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.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics, 23(4), pp.365-380.

Harvey, C. R., & Liu, Y. (2020). Lucky factors. Available at SSRN 2528780.

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

Harvey, C.R., Liu, Y. and Saretto, A., 2020. An Evaluation of Alternative Multiple Testing Methods for Finance Applications. The Review of Asset Pricing Studies, 10(2), pp.199-248.

Harvey, C.R., Liu, Y. and Zhu, H., 2016. … and the cross-section of expected returns. The Review of Financial Studies, 29(1), pp.5-68.

Henrique, B.M., Sobreiro, V.A. and Kimura, H., 2019. Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, 124, pp.226-251.

Jiang, S., Li, X. and Wang, S., 2020. Exploring evolution trends in cryptocurrency study: from underlying technology to economic applications. Finance Research Letters, 101532.

Linnainmaa, J.T. and Roberts, M.R., 2018. The history of the cross-section of stock returns. The Review of Financial Studies, 31(7), pp.2606-2649.

Liu, W., Liang, X. and Cui, G., 2020. Common risk factors in the returns on cryptocurrencies. Economic Modelling, 86, pp.299-305.

Liu, Y., Tsyvinski, A. and Wu, X., 2019. Common risk factors in cryptocurrency (No. w25882). National Bureau of Economic Research.

McLean, R.D. and Pontiff, J., 2016. Does academic research destroy stock return predictability?. The Journal of Finance, 71(1), pp.5-32.

Nazário, R.T.F., e Silva, J.L., Sobreiro, V.A. and Kimura, H., 2017. A literature review of technical analysis on stock markets. The Quarterly Review of Economics and Finance, 66, pp.115-126.

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

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

Romano, J.P. and Wolf, M., 2018, January. Multiple testing of one-sided hypotheses: combining Bonferroni and the bootstrap. In International Conference of the Thailand Econometrics Society (pp.78-94). Springer, Cham.

Schwert, G.W., 2011. Stock volatility during the recent financial crisis. European Financial Management, 17(5), pp.789-805.

Sensoy, A., 2019. The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. Finance Research Letters, 28, pp.68-73.

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

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

Welch, I. and Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), pp.1455-1508.

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

Zhang, W., Wang, P., Li, X. and Shen, D., 2018. The inefficiency of cryptocurrency and its cross-correlation with Dow Jones Industrial Average. Physica A: Statistical Mechanics and its Applications, 510, pp.658-670.

Appendix A. Methodological choices, data sample properties, and lucky trading rules

Figure A1. Data sample features and the properties of lucky trading rules–grouped by Trading Fees

NOTE. This figure reports how the characteristics of lucky trading rules relate with methodological choices such as considering trading fees, considering liquidity costs, only considering long trades, or standardizing the test statistic. Red dots denote results when the choice (subtracting trading fees) is true and black dots denote results when the choice is false. Each group of results is plotted from 48,000 points of data.

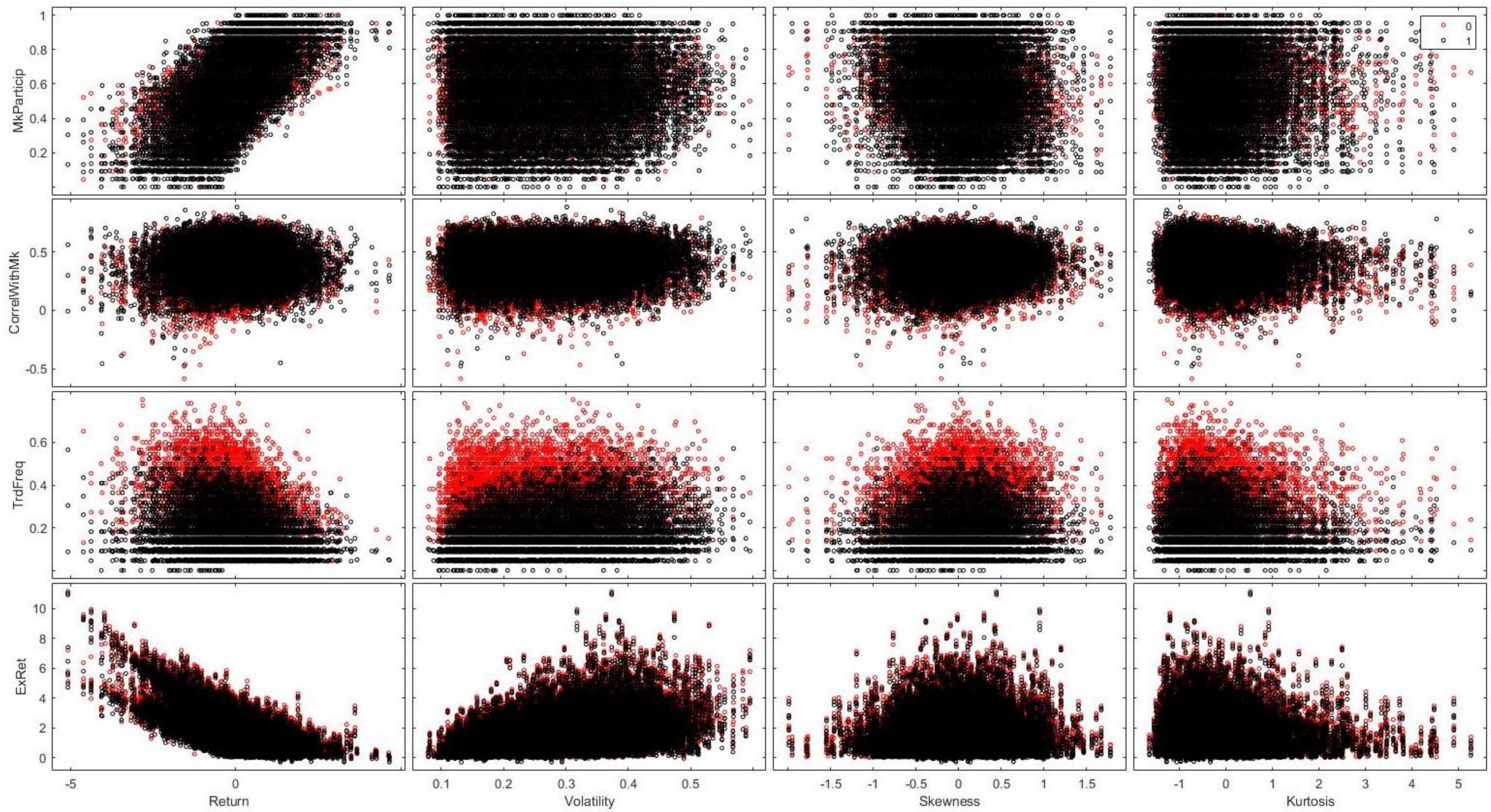


Figure A2. Data sample features and the properties of lucky trading rules–grouped by Liquidity Cost

NOTE. This figure reports how the characteristics of lucky trading rules relate with methodological choices such as considering trading fees, considering liquidity costs, only considering long trades, or standardizing the test statistic. Red dots denote results when the choice (subtracting liquidity costs) is true and black dots denote results when the choice is false. Each group of results is plotted from 48,000 points of data.

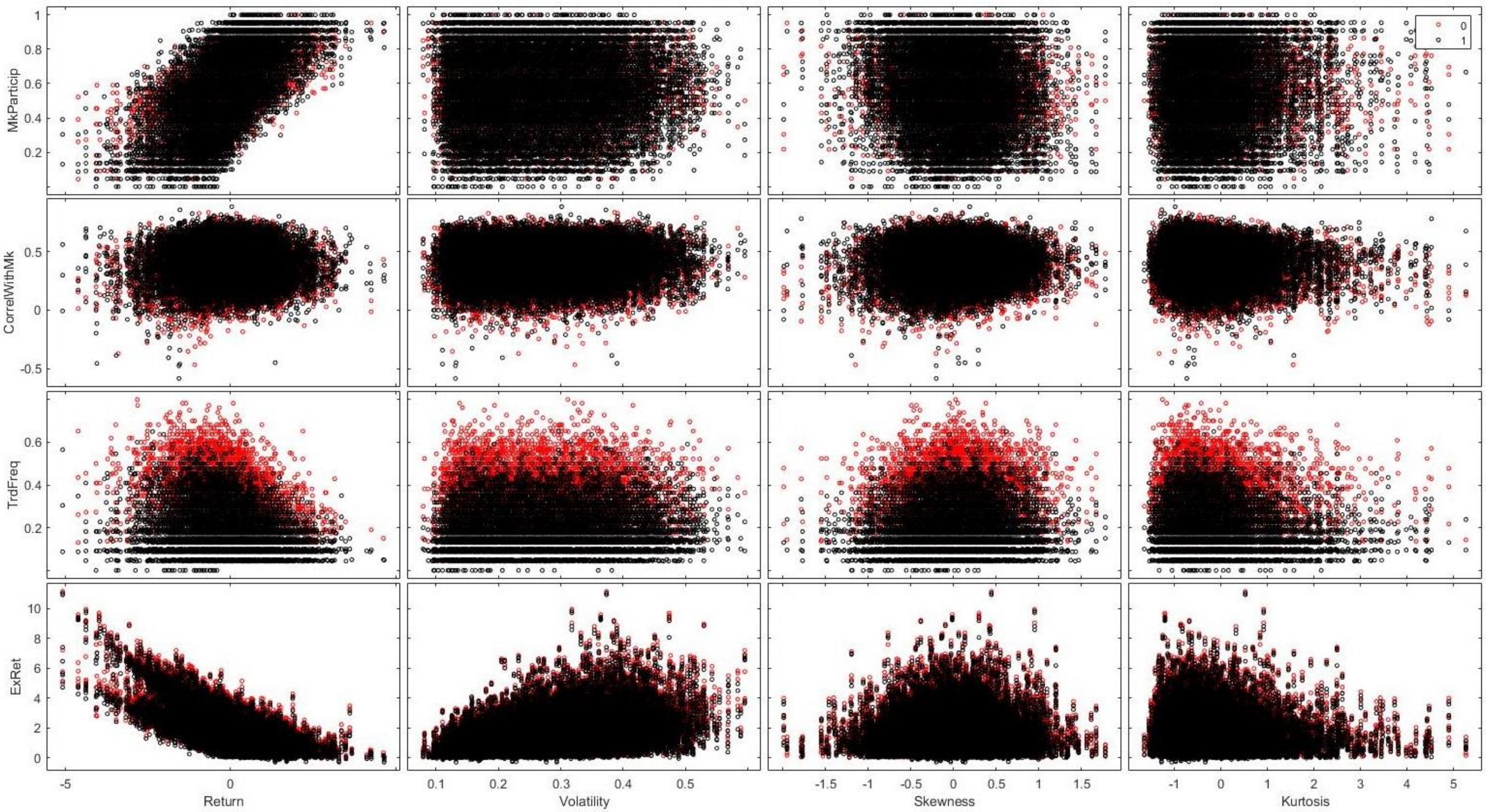


Figure A3. Data sample features and the properties of lucky trading rules–grouped by Short Trades

NOTE. This figure reports how the characteristics of lucky trading rules relate with methodological choices such as considering trading fees, considering liquidity costs, only considering long trades, or standardizing the test statistic. Red dots denote results when the choice (enabling short positions) is true and black dots denote results when the choice is false. Each group of results is plotted from 48,000 points of data.

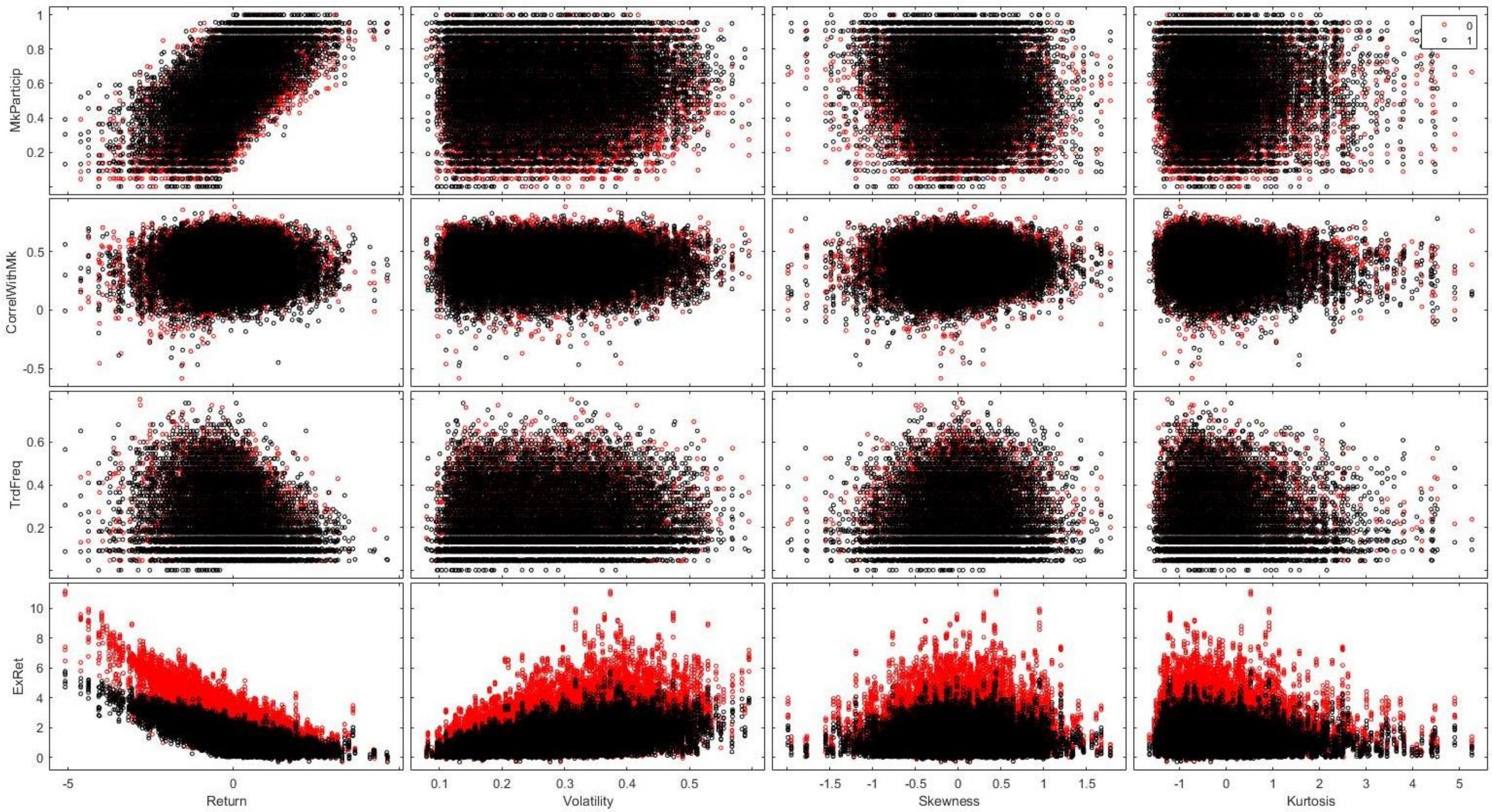


Figure A4. Data sample features and the properties of lucky trading rules–grouped by Standardized Statistic

NOTE. This figure reports how the characteristics of lucky trading rules relate with methodological choices such as considering trading fees, considering liquidity costs, only considering long trades, or standardizing the test statistic. Red dots denote results when the choice (standardizing the test statistic) is true and black dots denote results when the choice is false. Each group of results is plotted from 48,000 points of data.