Technical Trading Revisited: Persistence Tests, Transaction Costs, and False Discoveries *

Pierre Bajgrowicz† Olivier Scaillet‡

First version: May 2007
This version: January 2008

Abstract

We revisit the apparent historical success of technical trading rules on daily prices of the Dow Jones index. First, we use the False Discovery Rate as a new approach to data snooping. The advantage of the FDR over existing methods is that it is more powerful and not restricted only to the best rule in the sample. Second, we perform persistence tests and conclude that an investor would not have been able to select ex ante the future best-performing rules. Finally, we show that the performance fully disappears once transaction costs are taken into account.

*We thank Michel Dubois, Bernard Dumas, Laurent Frésard, as well as seminar participants at the EC2 conference 2007, the Swiss Doctoral Workshop in Finance 2007, and the AFFI 2007 Conference for helpful comments. Both authors received financial support by the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK).

†HEC Genève, Boulevard du Pont d'Arve 40, 1211 Geneva 4, Switzerland. Voice: +41 22 379 85 28. Fax: +41 22 379 81 04. E-mail: pierre.bajgrowicz@hec.unige.ch.

‡HEC Genève and Swiss Finance Institute, Boulevard du Pont d'Arve 40, 1211 Geneva 4, Switzerland. Voice: +41 22 379 88 16. Fax: +41 22 379 81 04. E-mail: olivier.scaillet@hec.unige.ch.

1 Introduction

Technical analysis involves the study of past price and volume data in order to predict future prices. Its ability to generate profits is the subject of a continuous debate. Academics have long been skeptical about the usefulness of this form of analysis. They argue that it is inconsistent with the theory of market efficiency, which states that all available information must be reflected in security prices. Practitioners, in spite of the contempt among academics, have historically devoted significant resources to technical trading rules. A substantial segment of the investment industry employs indicators which include moving averages, support and resistance levels, and other filter rules. In hopes of resolving this conflict, researchers have undertaken numerous studies of technical analysis. The resulting literature has found mixed results and academics have yet to agree on whether technical analysis can be used to generate economic performance.

In this paper we revisit existing results on the performance of technical trading rules by focusing on three issues that have been only partly addressed in the literature. The first issue known as data snooping is the problem whether the apparent performance of a particular trading rule is really significant and not simply due to luck and the abuse of data mining techniques. The second issue is whether and how an investor can possibly select the best technical rules prior to committing his money. The third issue is the impact of transaction costs on the performance of the strategies. To investigate these three issues, we use the same framework as Sullivan, Timmermann and White (1999) (later STW), who examine the performance of $7'846$ technical trading rules on 100 years of daily prices of the Dow Jones Industrial Average (DJIA) index. During the 100-year period and before transaction costs, technical rules seem to be a useful tool to generate superior returns, even after accounting for data snooping.

To illustrate the first issue, consider an investor who back-tests several quantitative trading strategies on historical data. By looking long enough and hard enough on a given set of data, he will always find a mechanical trading rule that works, even if it does not genuinely possess predictive power over asset returns. The observed good performance may simply be due to chance rather than to any merit inherent in the strategy yielding the returns. Similarly, if you put enough monkeys on typewriters, one of the monkeys will write the Iliad in ancient Greek. But would you bet any money that he is going to write the Odyssey next? Because of the sheer size of the sample, you are likely to find a lucky monkey once in a while. The same applies to traders. These two examples of data snooping problems could be easily resolved with a meaningful out-of-sample experiment. Unfortunately, unlike in engineering sciences, it is quasi impossible to design replicable experiments in finance, and typically only a single history for a given phenomenon of interest is available. Researchers have long been aware of the dangers of data snooping, but often have no other choice than to reuse existing data[1]. Recently, techniques have been proposed to rigorously measure the effect of data snooping, e.g., the bootstrap reality check (BRC) of White (2000), and the stepwise multiple testing method of Romano and Wolf (2005) (later RW).

The first contribution of this paper is to use the False Discovery Rate (FDR) as a new measure of data snooping. The FDR was first introduced by Benjamini and Hochberg (1995), and extended by Storey (2002). It is defined as the expected proportion of trading rules incorrectly identified as generating abnormal performance among all the selected rules. More precisely, we employ the $\text{FDR}^{+/-}$, developed by Barras, Scaillet and Wermers (2007) on the basis of the FDR. The $\text{FDR}^{+/-}$ gives the proportion of false discoveries separately among rules delivering respectively positive and negative returns. Compared to statistical methods used in previous studies, the advantage of the FDR approach is that it allows to construct a 'portfolio' of rules, in order to diversify risk in much the same way as adding companies to a stock portfolio. Our study confronts our FDR approach with the BRC—the data snooping measure used in STW, and with RW method. The BRC only indicates whether the rule that performs best in the sample indeed beats the benchmark, after accounting for data snooping. It provides no information on the other strategies. Though potentially able to detect further outperforming rules, RW method controls a conservative error measure, which prevents it from detecting more than a few rules. However, even with the increased power of our FDR approach, we are not able to detect enough trading rules generating significant positive returns.

The second contribution concerns how an investor could have possibly determined the best trading rules prior to investing his money. Although it may be the case that we are able to find technical rules that performed well historically, there is no indication that it is possible to select ex ante the trading rules that will generate superior performance in the future. We tackle this issue in

[1]See Lo and MacKinley (1990), White (2000), and the references therein.

two ways, by performing persistence tests, and by investigating whether certain categories of trading rules perform better under particular economic conditions.

To run the persistence analysis, we form portfolios of technical rules using the FDR approach on a two-year trailing window of data. We hold the portfolios for one year, then re-form them, sliding the training window one year forward. This yields a time series of portfolios returns obtained by exploiting only historically available information. Then, we measure the performance of the resulting portfolios and compare it to the performance of the rules that performed best ex post. While persistence analysis has been applied to mutual funds, e.g., Carhart (1997), to our knowledge this is the first time this type of persistence tests are performed on technical trading rules. STW only mention that the best rule selected by the end of 1986 "did not continue to generate valuable economic signal in the subsequent 10-year period". This result needs further investigation, as during that time period even in-sample performance is poor[2]. Jacquier and Yao (2002) implement another approach to persistence analysis also inspired by the mutual fund literature. They follow Brown and Goetzmann (1995), and estimate for example the probability that a trading rule beats the benchmark over two consecutive periods. Their study is limited to the ten moving average rules of Brock, Lakonishok and LeBaron (1992), and shows that the performance is not persistent at horizons shorter than five years. Allen and Karjalainen (1999) also point out that the rules need to be chosen using price data available before the start of the testing period. However, their study is different from ours as its principal aim is to tackle the problem induced by the ex post specification of rules[3]. Our tests show that the performance of a portfolio of trading rules does not persist out-of-sample.

In order to check whether an investor could improve his performance by switching to specific trading rules depending on the state of the economy, we condition on the business cycle when selecting the rules. Our analysis based on the dates of the National Bureau of Economic Research (NBER) shows that this approach does not yield better returns. Hence, it is hard to imagine how an investor could have picked the future outperforming strategies.

[2]STW also report the performance of what they call the cumulative wealth rule. This strategy generates a trading signal based solely on past information. However, it is unlikely that an investor would have followed this strategy, and it does not help in forming a portfolio.

[3]Whereas most of the studies seek to test whether particular kinds of technical analysis rules have forecasting ability, Allen and Karjalainen (1999) use a genetic algorithm to compose the rules, which avoids much of the arbitrariness involved in choosing which rules to test.

The third contribution of this paper is the in-depth analysis of the impact of transaction costs. The rules selected before transaction costs produce many trading signals and their performance is likely to be wiped out once we take into account these costs. Transaction costs are difficult to measure precisely as they include unobservable components such as the price impact of a trade, and they have been declining over time. To circumvent this difficulty, we examine how the performance of technical rules evolves across a whole range of transaction costs and lending fees levels (in the case of short selling). We explain why break-even transaction costs are not a satisfactory indicator, and why it is important to consider transaction costs already during the selection process, e.g., the training period of the persistence tests. We find that the initial performance fully disappears after the inclusion of transaction costs.

Overall, all three points that we investigate lead to the conclusion that technical trading rules—at least those from STW universe—could not have been used to generate significant profits. Though we are using a more powerful approach to select outperforming rules, the apparent historical success is wiped out by the inclusion of transaction costs and the absence of persistence. Our findings are in line with the Efficient Market Hypothesis (EMH) advocated by academics, and run against supporters of technical analysis.

Section 2 describes the universe of technical trading rules and the performance measures. Section 3 reviews existing methods to account for data snooping and presents our FDR based approach. Section 4 presents our empirical findings when transaction costs are omitted. In Section 5, we analyze the impact of transaction costs and the influence of the state of the business cycle.

2 Universe of trading rules and performance measures

2.1 Technical analysis

The conflict between the level of resources dedicated to technical analysis in the investment industry and academic theories of market efficiency has yet to be resolved. If the Efficient Market Hypothesis is correct, technical analysis should not work at all. The prevailing market price should reflect all information, including past price movements. Numerous studies provide results consistent with this traditional academic argument. For example, Allen and Karjalainen

(1999) find little support for the technical strategies they examine, and Fama and Blume (1966) and Bessembinder and Chan (1998) show that transaction costs could offset the benefits of technical analysis. Furthermore, Sullivan et al. (1999) and Jegadeesh (2000) warn to the danger of data snooping and survivorship biases when evaluating technical rules. Other studies have found results consistent with the practitioner view by providing evidence that technical analysis can predict price movements. For example, Neftci (1991), Brock et al. (1992), Taylor and Allen (1992), Blume, Easley and O'Hara (1994), Osler and Chang (1995), and Lo, Mamaysky and Wang (2000) test different technical trading rules and find evidence that technical analysis provides information beyond that already incorporated in the current price. More recently, Kavajecz and Odders-White (2004) explore the relation of technical analysis to liquidity provision. An explanation why technical analysis might work can be found in the field of behavioral finance, which suggests that investors may not be completely rational and that their psychological biases could cause prices to deviate from their 'correct' level (e.g., DeBondt and Thaler (1985), and Barberis, Shleifer and Vishny (1998)). Therefore, sophisticated speculators will not trade purely on consideration of the economic fundamentals, but will also aim to exploit market movements generated by less sophisticated, 'noise traders'. For example, some technical predictions may be self-fulfilling. If everyone believes a technical analysis signal saying that stock X will rebound at $30, they will buy as the price approaches that level. Schmidt (2002) shows using a simple agent-based market dynamics model that if the technical traders are able to affect the market liquidity, their concerted actions can move the market price in the direction favorable to their strategy. Therefore, knowledge of chart signals can be essential as they have a bearing on the action of many market participants.

2.2 Universe of trading rules

To investigate whether technical trading rules generate superior performance, we need to specify a universe of technical rules from which investors could have drawn their strategies. When applied to a series of past prices, a trading rule indicates whether a long position, a neutral position (i.e., out of the market), or a short position should be taken in the next time period. Formally, we consider that to the k-th rule corresponds a signal function $s_{k,t-1}$, based on the information up to time $t-1$, which returns the value 1 for a long position, 0 for a neutral position, and -1 for a short position. In order to allow for comparison with their results, we stick to STW universe, which consists of $l = 7'846$ rules divided into five categories: filter rules, moving averages, support and resistance rules, channel breakouts, and on-balance volume averages. See Brock et al. (1992), and STW for an explanation of the rules and their exact parameterizations.

In order to be included in the universe, a rule has to fulfill a certain number of requirements. For example, STW explain that it is important that the strategy was known at the time its performance is evaluated, and that it was possible for an investor to implement it. Suppose that some technical trading rules can be found to unambiguously outperform the benchmark over the sample period, but that these are based on technology (e.g., neural networks) that only became available after the end of the sample. As investors would not have been able to apply these rules, such evidence cannot lead to any meaningful conclusions.

2.3 Performance measures

Each rule k, $1 \leq k \leq l$, generates an investment signal $s_{k,t-1}$ for each prediction period t, $R \leq t \leq T$. For each rule, we compute a test statistic φ_k, which measures the performance of the rule relative to a benchmark. The statistic is defined in such a way that $\varphi_k = 0$ under the null hypothesis that rule k does not generate abnormal performance relative to the benchmark.

Following STW, we focus our analysis on two simple performance criteria: the mean return, which measures the absolute performance, and the Sharpe ratio, which measures the average excess return per unit of total risk. For the mean return criterion, we set the rule of a neutral position at all times (zero return, always out of the market) as the benchmark. Hence, the test statistic of rule k is simply its mean return. In the case of the Sharpe ratio criterion, we follow standard practice and compute the return in excess of the risk-free rate. This implies that trading rules earn the risk-free rate on days where a neutral signal is generated.

Let y_t be the (arithmetic) period t return on the price series on which the strategies are applied. We use the same notation as STW and denote by $f_{k,t} = \ln(1 + s_{k,t-1}y_t)$ the (logarithmic) period t return generated by rule k. Then, the test statistic for the mean return criterion can be written as $\varphi_k = \bar{f}_k = \frac{1}{N}\sum_{t=R}^{T} f_{k,t+1}$, where $N = T - R + 1$ is the number of prediction periods.

In order to obtain the expression for the Sharpe ratio, let $f^e_{k,t} = s_{k,t-1}y_t - r_{f,t}$

denote the (arithmetic) period t excess return of rule k, where $r_{f,t}$ is the risk-free rate. The mean excess return can be written as $\bar{f}_k^e = \frac{1}{N}\sum_{t=R}^{T} f_{k,t+1}^e$, and the standard deviation as $\sigma_k^e = \frac{1}{N-1}\sum_{t=R}^{T}\left(f_{k,t+1}^e - \bar{f}_k^e\right)^2$. Then, the test statistic for the Sharpe ratio is simply $\varphi_k = \mathrm{SR}_k = \frac{\bar{f}_k^e}{\sigma_k^e}$.

2.4 Data

The nearly eight thousand parameterizations of trading rules are applied to the same data set as in STW, namely daily closing prices on the Dow Jones Industrial Average (DJIA) index. It can be argued that until the recent introduction of exchange-traded funds (ETFs), it was impossible to trade stock indices frequently without incurring significant transaction costs. ETFs are open-ended collective investment schemes, traded as shares on most global stock exchanges. Typically, ETFs try to replicate a stock market index, a market sector, or a commodity such as gold or petroleum. Investors can sell short ETFs, use a limit order, use a stop-loss order, buy on margin, and invest as much or as little money as they wish. In addition to these stock-like features, most ETFs have a lower expense ratio than comparable mutual funds. The DIAMONDS Trust is an ETF designed to "correspond to the price and yield the performance of the DJIA"[4]. For the period after the fund inception date in January 1998, we applied the technical rules both directly to the DJIA index series, and to the DIAMONDS ETF. The very similar results obtained indicate that it has become realistic to assume that investors apply technical rules directly to a stock market index.

The data in STW finish in 1996. We consider the same subperiods as STW, and add one period for the new data between January 1997 and July 2007 (see Table 1). For comparison, we also run the strategies on the 100-year period from the inspection of the DJIA index. Results for this latter sample should be viewed with caution, as market conditions have evolved dramatically in the last 100 years. Furthermore, managers never get to trade for 100 years before their performance is evaluated. Finally, when computing the Sharpe ratio, we use the same risk-free rate as STW[5].

[Table 1]

[4]See the DIAMONDS Trust prospectus available at the American Stock Exchange webpage.

[5]We are grateful to A. Timmermann for providing us the DJIA index and risk-free rate series.

3 Data snooping measures

3.1 Problem formulation and existing methods

In an attempt to settle the conflict between academic theories and the methods employed by practitioners, we investigate whether it is possible to find technical trading rules able to generate significant positive performance[6]. For each rule k, $1 \leq k \leq l$, we test the null hypothesis H_{0k} of no abnormal performance, versus the alternative H_{Ak} of the presence of abnormal performance, positive or negative:

$$H_{0k} : \varphi_k = 0, \quad H_{Ak} : \varphi_k > 0 \text{ or } \varphi_k < 0. \tag{1}$$

In an ideal world, we would reject H_{0k} exactly for those rules that generate abnormal performance. In reality, this usually cannot be achieved with certainty. In applications where several tests are carried out at once, the danger of data snooping is great. The consequence is that trading rules are selected, which do not possess genuine predictive power.

In the finance literature, a first solution to account for data snooping is provided by White (2000) bootstrap reality check. The BRC tests the null hypothesis that the performance of the best technical trading rule is no better than the performance of the benchmark:

$$H_0 : \max_{k=1,\ldots,l} \varphi_k \leq 0.$$

Hence, it implicitly accounts for the dependence structure of the individual test statistics. The BRC addresses the question whether the rule that appears best in the sample really beats the benchmark. It is the data snooping measure used in the study of STW. However, it is not able to identify further strategies that beat the benchmark. A first attempt to tackle this issue is the stepwise multiple testing method of Romano and Wolf (2005). RW algorithm uses a modified BRC as a first step, and can potentially detect further outperforming strategies in subsequent steps, which makes it more powerful than the BRC. The method of RW controls the familywise error rate (FWER), which is defined as the probability of erroneously selecting one or more trading rules as significant, when in reality they are simply lucky. The FWER is a conservative criterion, resulting in a low power to detect superior performance, especially when the

[6]The present section focuses on trading rules discovery. However, the methods presented can be applied to any multiple testing problem.

universe of rules is large. It is not very appropriate in our case, where the goal is to find a large number of significantly positive strategies[7].

3.2 False Discovery Rate

We now present our new approach to data snooping. Benjamini and Hochberg (1995) argue that the control of the FWER is not always necessary, especially if the goal is to find several outperforming trading rules. They propose a more tolerant error measure, the false discovery rate (FDR). The FDR is the expected ratio of the F trading rules erroneously selected as generating abnormal performance—the false discoveries, over the total number R of selected rules (see Table 2). We can write the estimator of the FDR as

$$\widehat{\mathrm{FDR}} = \frac{\widehat{F}}{\widehat{R}},$$

where $\widehat{F}$ and $\widehat{R}$ are estimators of F and R. See the appendix for the detailed formulas. An FDR of 100% means that no rule is able to deliver significant returns and that the apparent performance is purely due to luck, i.e., data snooping. On the other extreme, an FDR of 0% indicates that all selected strategies do genuinely generate significant performance.

[Table 2]

The FDR offers a much less conservative criterion over the FWER and, therefore, leads to an increase in power[8]. Compared with RW method, the FDR approach selects a sufficient number of rules to allow an investor to construct a portfolio of trading rules. Another virtue of the FDR approach is its simplicity. Once the p-values corresponding to the individual tests have been calculated, the estimation of the FDR is straightforward. The single parameter to be estimated is the proportion π_0 of rules in the population satisfying the null

[7] Hansen (2005) offers some improvements over the BRC. Being less sensitive to the influence of poor and irrelevant strategies, his method is more powerful. However, like the BRC, Hansen method only addresses the question whether the strategy that appears best in the observed data really beats the benchmark. Hsu and Kuan (2005) utilize the test of Hansen to reexamine the profitability of technical analysis and conclude that there are no significantly profitable trading rules in mature markets (i.e., DJIA and S&P 500).

[8] The FDR approach has received much recent attention in the statistics literature, and has been extended by Storey (2002). See Abramovich, Benjamini, Donoho and Johnstone (2006) for applications of the FDR and for an extensive discussion of the advantages of using the FDR over the FWER in the field of multiple testing.

hypothesis $\varphi = 0$. We obtain the individual p-values using the same resampling technique as STW.

Technical rules for which the null hypothesis is rejected perform either better or worse than the benchmark. Barras et al. (2007) show that the proportion of false discoveries can be estimated separately for both outperforming and underperforming strategies. Let R^+ denote the number of significantly positive trading rules[9]. F^+ of them do not truly generate abnormal performance, but are simply false discoveries. The FDR among the rules yielding positive returns, denoted by FDR^+, is defined as the expected ratio of F^+ over R^+. The FDR^+ can be estimated as

$$\widehat{\text{FDR}}^+ = \frac{\widehat{F}^+}{\widehat{R}^+},$$

where $\widehat{F}^+$ and $\widehat{R}^+$ are estimators of F^+ and R^+. Similarly, an estimator of the FDR among the rules yielding negative returns, denoted by FDR^-, can be written as

$$\widehat{\text{FDR}}^- = \frac{\widehat{F}^-}{\widehat{R}^-}.$$

We can also estimate the proportions π_A^+ and π_A^- of positive and negative trading rules in the population. These are general indicators of the presence of positive or negative abnormal performance. All relevant estimation procedures to get $\widehat{\text{FDR}}^+$, $\widehat{\text{FDR}}^-$, $\widehat{\pi}_A^+$, $\widehat{\pi}_A^-$, as well as the stationary bootstrap used to obtain the individual p-values are detailed in the appendix.

The FDR provides two approaches for the construction of a portfolio of trading rules. The first approach, which we call exploratory analysis, consists in selecting strategies based on their p-values in a first step, and then computing the resulting $\widehat{\text{FDR}}^+$ [10]. The second approach is to set the $\widehat{\text{FDR}}^+$ at a predetermined rate and use the algorithm described in the appendix to pick the corresponding trading rules. In both cases, the selected strategies are weighted equally within the portfolio.

[9] We call a trading rule significantly positive if its abnormal performance is both significant (i.e., H_0 is rejected) and positive.

[10] Here we follow Storey (2002) who proposes to first fix the rejection region, before computing the corresponding proportion of false discoveries.

4 Empirical findings before transaction costs

4.1 FDR in practice / Evolution of performance over time

In this section, we investigate the performance of technical trading rules, using the FDR as a new data snooping measure, and ignoring transaction costs. As we use the same setting as STW, we can compare our findings and show the advantages of the FDR approach over the BRC and RW method. Table 3 presents performance results for the mean return criterion in each of the sample periods. The top panel shows results already reported by STW, i.e., the performance of the buy-and-hold strategy (DJIA), and the performance and BRC p-value of the best rule in the sample. It also displays the performance and size of the portfolio of trading rules constructed using the method of RW. The bottom panel displays three different indicators obtained following our new FDR based approach to data snooping. On the left-hand side is the performance of the portfolio of trading rules obtained with the simple exploratory analysis. In the middle are results for the portfolio constructed by controlling the FDR^+ at 10%. Finally, the last columns show $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$, the estimated proportions of rules in the population with positive and negative performance.

[Table 3]

Over the 100-year period (1897–1996), the best rule generates a yearly mean return of almost 16%, whereas the DJIA delivers only 4.5%. As explained by STW, this performance stands up to the effects of data snooping since the BRC p-value is zero. This is encouraging but provides no information on other trading rules than the best in the sample. In theory, the method of RW provides a solution to this issue as it can possibly detect further outperforming rules. As shown in Table 3, over the 100-year period, RW technique selects 199 rules from the universe of nearly eight thousand. The first three columns of the bottom panel present results for the FDR exploratory analysis. In the exploratory analysis, first we select the rules having a positive return and a p-value inferior to 5%, and put them into an equally weighted portfolio. Then we estimate the amount of false discoveries within this portfolio and measure its performance. For the 100-year period, 1'216 rules are selected during the exploratory analysis—significantly more than by the RW approach, and the corresponding

FDR is very low at 1.3%[11]. This difference is even more pronounced during the individual subperiods, where the average RW portfolio size does not exceed 20. Such a low number cannot guarantee a real diversification effect and highlights the low power of RW method. The increased power of the FDR method is a clear advantage when the aim is to form a portfolio of trading rules.

Globally, the conclusions from the FDR approach are in line with the BRC. During the 100-year period until 1996 and before transaction costs, technical analysis seems to be a useful tool to generate performance, even after accounting for data snooping. As already reported by STW and in other studies, the historical performance of technical rules tends to disappear over time. This trend is confirmed by the new data now available for subperiod 6 (January 1997–July 2007). During that last 10-year period, the performance disappears completely. The BRC p-value equals 1 and no single rule is selected into the FDR portfolio. The FDR approach allows to detect the performance decline earlier, already in subperiod 4. Whereas the BRC p-value is still strictly zero at that time, the FDR exploratory analysis selects merely 124 rules, from which almost 30 percent are false discoveries. Furthermore, with the new indicators not limited to a single rule, we can be assured that the disappearance of performance concerns the whole universe of trading rules. This phenomenon can be explained by cheaper computing power, lower transaction costs, and increased liquidity which have helped to remove possible patterns in stock returns. Another illustration of the usefulness of our approach is when it is not obvious how to interpret the BRC, such as in subperiod 2 where the BRC p-value reaches 3%. For the same subperiod, the FDR exploratory analysis selects only 132 rules, among which almost 20 percent are false discoveries. Hence, we are confirmed that only few trading rules generate significant performance during that time period.

The remaining of the bottom panel presents results stemming from other ways of applying the FDR. 'FDR control' corresponds to the case where we control the FDR^+ at a predetermined rate (10% in our case[12]) and construct an equally-weighted portfolio of trading rules satisfying this constraint, using

[11] The corresponding asymptotic confidence interval computed following Genovese and Wasserman (2004) is [1.18% 1.38%]. In the rest of the paper we do not report confidence intervals for the FDR in order not to overload tables. The intervals are always narrow and available upon request.

[12] Controlling the FDR^+ at $\gamma = 5\%$ and $\gamma = 20\%$ leads to very similar results available upon request.

the algorithm described in the appendix. Finally, the last two columns of the bottom panel show that $\widehat{\pi}_A^+$ decreases from approximately 40% to zero, while $\widehat{\pi}_A^-$ becomes non negligible after subperiod 3 (1939–1962) and reaches more than 15% in subperiod 6 (1997–2007). Hence, the FDR provides us with three different approaches which all lead to the same conclusions. A very similar picture emerges for the Sharpe ratio criterion, as reported in Table 4.

[Table 4]

Table 5 provides summary statistics for the best-performing trading rule, for the RW portfolio, and for the 10% FDR^+ portfolio, during subperiod 3. The rules are chosen with respect to the mean return criterion and ignoring transaction costs. In the three cases, the number of short and long trades is roughly balanced out, and the winning percentage is much higher for the long than for the short trades. Long trades are also associated with average profits superior to those on the short trades.

[Table 5]

4.2 Persistence analysis

An important issue, and another contribution of our paper, is the assessment of how an investor could have possibly selected the future best-performing rules. Although it may be the case that we are able to find ex post technical rules that generate superior performance, there is no indication that it is possible to select this rules ex ante. In this section, we tackle this problem by performing persistence tests. To do so, we construct an equally weighted portfolio of technical rules selected during a training period corresponding to the last two years. The criterion to select the trading rules is the control of the FDR^+ at a predetermined level. We then measure the (out-of-sample) performance of the portfolio over a testing period corresponding to the following year, and compare it to the (in-sample) performance of the rules that performed best ex post. Every year, the composition of the portfolio is updated in order to take advantage of the new data that has become available. Each time, we move the two-year trailing window of the training period one year forward and run the selection process again (see Figure 1). Hence, we exploit only historically available information, so this approach could have been easily implemented by an investor.

[Figure 1]

[Table 6]

Table 6 displays the results of our persistence tests for subperiod 3 (1939–1962) with the mean return as the performance measure. Though we focus on this sample period as it exhibits high in-sample performance, the findings are identical for other subperiods. We consider portfolios obtained by controlling the FDR^+ at different levels (1%, 5%, 10%, and 20%). For comparison, we also report the performance of the portfolio constructed using RW methodology, of the portfolio consisting of the 200 best rules from the training period, of the best rule of the training period, and of the best rule provided the BRC p-value is below 5%. The left-hand part of the table shows the (out-of-sample) performance and size of the different portfolios rebalanced yearly as just described. The right-hand side displays in-sample results corresponding to the unrealistic situation where the trading rules are selected and evaluated over the same period. An investor earns a yearly mean return of 5.3% if he applies the FDR approach and 11.4% with RW method, which is better than the 2.4% obtained if he naively selects the best 200 rules of the training period. Hence, following the FDR or RW approach effectively allows to filter out some of the rules whose apparent performance is only a data snooping artifact. The FDR approach does not, though, allow to beat the 8.1% of the buy-and-hold strategy (DJIA). The performance of the RW portfolio is superior but relies on insufficiently few trading rules.

The out-of-sample performance corresponds to what an investor can reasonably achieve. An investor must not be lured by the prospect of achieving the high returns of the best rules selected ex post, i.e., an artificial mean return of almost 20% yearly for the FDR portfolio. Moreover, the significant trading rules are dispersed across the whole tail of the distribution, i.e., controlling the FDR^+ at 1%, 5%, 10%, or 20% yields the same portfolio return of 5.3%. Contrary to Barras et al. (2007) where the few skilled mutual funds are located in the extremity of the tail, we cannot take advantage of such a property to pick the best trading rules. The great variability of the size of the portfolios reported in Table 6 is another indicator that it is not possible to detect the top-performing rules ex ante. Unreported results (e.g., the composition of the portfolios built ex ante and ex post for each testing period) show that an investor would not have been able to pick much more than 5% of the 200 future best rules. Hence, despite the high in-sample performance during past subperiods, the strategies of STW universe could not have been used to generate significant performance.

Table 6 shows that RW portfolio is often empty which illustrates the weakness of RW method. In comparison, the FDR approach is more powerful. This is a useful property as our goal is to detect many trading rules in order to obtain a diversification effect. As reported in Table 7, results are identical for the Sharpe ratio criterion.

[Table 7]

5 Impact of transaction costs

Before transaction costs, historical in-sample performance of technical trading rules appears to be high and significant even after accounting for data snooping. We have already seen that this performance is not persistent. In this section, we show that even this in-sample historical success disappears once we take into account transaction costs and short sale constraints. Previous studies[13] call for careful consideration of this issue, especially as an important proportion of the rules that perform best before transaction costs use very short windows of data, generate very frequent trading signals, and, hence, are likely to generate substantial transaction costs. STW try to address this issue by conducting their experiment using price data on the S&P 500 index futures. When trading futures contracts, transaction costs are easy to control, and it is not a problem to take a short position. However, the futures contract started trading only in 1984 which limits the interest of this approach, considering that our study begins in 1897. In this section, we tackle this issue and investigate in depth the impact of transaction costs on the performance and characteristics of technical rules. Total transaction costs are difficult to measure precisely as they include not only the bid-ask spread but also applicable commissions, price impact costs, taxes, short sale costs, and other immediacy costs. Moreover, they have been declining over time. To circumvent this difficulty, we examine how the performance of trading rules evolves across a whole range of transaction costs and lending fees levels. Another approach, which is also taken by STW and in Bessembinder and Chan (1998), is to compute a break-even transaction cost. However, as STW mention themselves, such a number is difficult to assess since transaction costs are not constant during the sample period. We explain why break-even transaction costs are not a satisfactory measure, in particular as it is important

[13]See Brock et al. (1992), STW.

to include the costs already during the selection process. Having recognized this fact, we perform the persistence tests again, adding transaction costs in the training period. We also examine the characteristics of portfolios of trading rules that resist the inclusion of transaction costs, and investigate whether some categories of rules perform better under particular economic conditions, following the NBER classification.

5.1 Transaction costs

Transaction costs are commonly decomposed into two major components: explicit costs and implicit costs. Explicit costs are the direct costs of trading, such as broker commissions and taxes. Implicit costs, which are harder to measure, represent such indirect costs as the price impact of the trade and the opportunity cost of failing to execute the order in a timely manner. The literature provides a range of transaction cost estimation procedures[14]. The first class of measures examines transaction cost data directly. As the latter are not easily available, a second class of methods indirectly infer transaction costs based on price behavior[15].

For the period January 1991 to March 1993, Keim and Madhavan (1997) estimate that, for exchange listed stocks, the average total cost for a buy order is 0.49% (0.31% implicit costs + 0.18% explicit costs). Transaction costs were significantly greater in earlier years, particularly before commissions were deregulated in May 1975. Stoll and Whaley (1983) use published commission schedules to estimate transaction costs during the 1960 to 1975 period. For the largest decile of NYSE securities, they report an estimated one-way transaction cost of 1.35% (the commission plus half the bid-ask spread). This figure may overstate actual transaction costs during that interval, since it does not accommodate the possibility of trading within the quotes, or allow for soft-dollar payments[16]. This latter practice has grown since the late 1980s, and the true decline in commission costs is even larger than the stated numbers. Another issue is the non-proportional increase of the price impact with the volume traded[17]. However, since the emergence of ETFs, transaction costs are easier to control.

[14] See Lesmond, Schill and Zhou (2004), and Mitchell and Pulvino (2001) for a review of transaction costs estimation procedures.

[15] e.g., Glosten and Harris (1988) and Breen, Hodrick and Korajczyk (2002).

[16] See Blume (1993).

[17] See Glosten and Harris (1988), Breen et al. (2002), Korajczyk and Sadka (2004).

5.2 Short selling constraints

Selling short can be expensive. In order to sell short, we must borrow the stock from a current owner, and this stock lender charges a fee to the short seller. The fee is determined by supply and demand for the stock in the loan market. In addition to these direct costs, there are other costs and risks associated with shorting, such as the risk that the short position will have to be involuntary closed due to recall of the stock loan (short squeeze). Furthermore, legal and institutional constraints can inhibit investors from selling short. These impediments and costs are collectively referred to as short-sale constraints.

D'Avolio (2002), Duffie, Gârleanu and Pedersen (2002), Geczy, Musto and Reed (2002), and Jones and Lamont (2002) provide useful analyzes of the equity loan market. While short-sale costs might be quite low on average, they are systematically high exactly when they are critical. Practitioners refer to stocks with high fees as being "special" and those with baseline fees as "general collateral". As for transaction costs, lending fees have declined over time. The average shorting cost in Jones and Lamont (2002) sample (1926–1933) is 35 basis points per month. For the period 2000–2001, D'Avolio (2002) reports only 41 basis points *per year*. However, 9% are loan market specials, with fees averaging 4.3% per annum, but reaching spectacular heights in some rare instances.

5.3 Exploratory analysis

In this section, we explore the impact of increasing transaction costs on the performance of trading rules. We focus on subperiod 3 (1939–1962) as it exhibits high in-sample performance and perform the FDR exploratory analysis. Conclusions are identical for other subperiods. In line with the estimates presented above, we consider combinations between three different lending fees (0, 50 and 100 basis points yearly) and proportional one-way transaction costs that range from 0 to 40 basis points of the traded volume. Total one-way transaction costs during subperiod 3 were higher than 1%, meaning that even the highest value used in our simulations remains conservative. Looking at Figure 2, we observe that, under the mean return criterion, the roughly 800 rules selected before transaction costs decrease to less than 100. The corresponding proportion of false discoveries ($\widehat{\mathrm{FDR}}^{+}$) raises from less than 2% to more than 25%. In a symmetric fashion, there are almost no significant rules with negative performance when we omit transaction costs, but their number increases to more than 1'100 under the highest transaction costs considered. Hence, it is already clear from this simple analysis that the apparent historical performance disappears when transaction costs are taken into account. Moreover, this phenomenon is confirmed by other indicators, such as the BRC p-value and the estimated proportions $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$. $\widehat{\pi}_A^+$ decreases from 50% to 10%. At the same time, $\widehat{\pi}_A^-$, which is negligible at first, reaches almost 25%. Results are very similar when the rules are selected according to the Sharpe ratio criterion (see Figure 3).

[Figures 2 and 3]

5.4 Break-even transaction costs

In the preceding sections, we choose transaction costs levels based on studies specializing on this particular topic. However, they remain ad hoc values. One way to circumvent this issue is to compute break-even transaction costs. Break-even transaction costs correspond to the level of transaction costs that would just eliminate the ex post difference between cumulative returns to traders using technical rules versus those who buy and hold the DJIA stocks. Though break-even transaction costs are often reported in the literature, they are not always appropriate. First, as historical series on transaction costs are difficult to obtain, it is hard to assess this number. Second, computing break-even transaction costs ex post, i.e., after the rules have been selected, misses an important issue. The problem is that the rules selected before transaction costs tend to generate very frequent trading signals. Hence, it is natural that their performance is highly impacted once we subtract transaction costs. Market frictions should be included from the start. This way, longer-term rules would be probably selected in the portfolio, resulting in a performance less sensitive to transaction costs. To remedy this problem, we could progressively increase transaction costs until they equal the resulting break-even transaction costs of the selected rules. The persistence analysis presented in the following section incorporates transaction costs already in the training period. Therefore, it does not suffer from this weakness of break-even transaction costs.

[Tables 8 and 9]

Table 8 reports break-even transaction costs corresponding to various portfolios of technical rules, in each of the sample periods. We consider the portfolio constructed by controlling the FDR$^+$ at 10%, the portfolio of rules selected by RW method, a portfolio containing the 200 best rules, and the best rule in the sample. For example over the 100-year period (1897–1996), the best-performing trading rule for the DJIA earns a mean yearly return of 15.9 percent resulting from an average of 117 transactions per year, giving a break-even transaction cost level of 0.14 percent per transaction. Transaction costs are likely to have been higher at the beginning of the sample, but lower by the end of the sample. As it is not obvious how to compute historical series on transaction costs, this number is difficult to assess and makes it hard to draw a conclusion concerning the profitability of the best rule in the sample. However, as soon as we consider whole portfolios, the corresponding break-even transaction costs drop dramatically (e.g., 0.0001 percent for the 10% FDR$^+$ portfolio). This low value should not come as a surprise since a portfolio contains an important number of rules, which leads to a significantly increased total number of trades (see Table 9). Hence, putting aside the above comments on break-even transaction costs, we can conclude that transaction costs should have been unrealistically low for our portfolios to generate profits.

5.5 Persistence analysis

We have seen that the apparent in-sample historical performance of technical rules selected before transaction costs is wiped out by the inclusion of these costs. Contrary to the break-even transaction costs analysis, in this section we incorporate transaction costs already in the training period and perform persistence tests as in Section 4.2. We believe it is important to incorporate the frictions in the selection process, as it might help discard insignificant rules that trade too frequently. Again, we focus on subperiod 3. We set one-way transaction costs at 20 basis points and lending fees at 50 basis points per year, which are conservative figures for the subperiod under consideration. As shown in Table 10 for the mean return criterion, and Table 11 for the Sharpe ratio, the persistence results are not better than when transaction costs are omitted. The portfolios constructed following the different approaches are often empty and their mean return is negative. Unreported results show that merely 1% of the future 200 top-performing rules are selected. Again, comparison of the performance of the portfolio where the FDR$^+$ is controlled at 1%, 5%, 10%, or 20% shows that the significant rules are spread across the tail, which does not facilitate the task of selecting the best strategies. Recall that after transaction costs, in-sample performance is also poor, barely matching the buy-and-hold strategy.

[Tables 10 and 11]

5.6 Portfolio characteristics

In this section, we analyze the characteristics of the 10% FDR$^+$ portfolio when we progressively add increasing transaction costs. First, we investigate which categories of trading rules continue to generate significant performance. Figure 4 shows the results for subperiod 3 with the mean return criterion. We see that only filter rules and on-balance volume averages remain after the inclusion of transaction costs. However, unreported results show that no persistent pattern emerges across the subperiods as to which category is more resistant to transaction costs. Furthermore, the results depend on the performance criterion employed, as we observe by comparing Figure 4 and Figure 5, which presents the analysis for the Sharpe ratio. We also investigate how the proportion of slow or long-term versus fast or short-term rules composing the portfolio evolves when transaction costs increase. The balance changes progressively in favor of slow rules, meaning that once transaction costs are included, the successful rules trade on longer-term price movements. The same trend emerges from Table 12, which lists the best-performing rule for each level of transaction costs. Figure 6 and Figure 7 display boxplots of the number of trades generated by all the rules constituting the 10% FDR$^+$ portfolio under increasing transaction costs, for respectively the mean return and the Sharpe ratio criterion. As expected, the rules selected before transaction costs produce too many trading signals and their performance is wiped out once we take into account these costs. Also, Table 13 shows that once transaction costs are included, the trading rules selected for subperiod 3 generate significantly less trades, and a FDR portfolio trade averages over 300 days. This is considerably greater than the average 30.3 days per trade resulting from the FDR portfolio rules when transaction costs are ignored (see Table 5).

[Figures 4, 5, 6, and 7]

[Tables 12 and 13]

5.7 Performance under particular economic conditions

In order to see whether certain trading rules perform better within a particular economic environment, we rerun our analysis conditioning on the state of the business cycle. We use the (ex post) dates of the National Bureau of Economic Research (NBER) to determine whether the economy is in an expansion or in a contraction phase. The analysis shows that different categories of rules are selected during expansions than during contractions. For example, subperiod 3 (1939–1962) contains 5 business cycles. During that time period, Moving Averages perform better if we focus on contraction phases, and if we concentrate on expansion phases, Support and Resistance, and Channel Breakouts rules are selected. Over the whole sample (i.e., when no distinction between contractions and expansions is made), none of these three categories is able to generate constant performance. However, there does not seem to exist a stable pattern across the different subperiods, as to which category of trading rules is superior during a particular phase of the business cycle. The single general result is that more rules are selected during contractions than during expansions or over the complete sample. However, despite these differences, conditioning on the business cycle does not allow to increase performance. The rules selected over the whole sample still yield the best returns[18]. Hence, even an investor who would know ex ante the state of the business cycle could not take advantage of it.

6 Conclusion

In this paper, we reassess the apparent historical success of technical trading rules on daily prices of the Dow Jones index. The three issues we investigate lead us to the conclusion that the rules from STW universe could not have been used to generate significant profits. First, we use the False Discovery Rate as a new approach to account for data snooping. Being more powerful than statistical methods used in previous studies, the FDR approach selects more outperforming rules, which allows the construction of a diversified portfolio of strategies. Second, we perform persistence tests in order to see whether it is possible to find trading rules that generalize beyond the training sample. The absence of persistence indicates that, even if we can find technical rules that perform well historically, an investor would not have been able to select ex ante

[18]Detailed results are available upon request.

the future best-performing rules. Third, we investigate the impact of transaction costs. Since technical trading strategies require frequent transactions, return forecastability may not imply increased returns once transaction costs are considered. Our analysis shows that rules selected before transaction costs trade too frequently and that their performance is wiped out by the inclusion of these costs. Even with the increased power of our FDR approach, we are unable to detect enough trading rules that are persistent and robust to the inclusion of transaction costs.

Appendices

For the convenience of the reader, we summarize here results from Barras et al. (2007) and STW.

A Stationary Bootstrap

For each trading rule, we test the null hypothesis of no abnormal performance. In order to obtain the individual p-values, we follow STW and apply the stationary bootstrap of Politis and Romano (1994). This resampling technique is chosen due to the weak correlation in the daily returns. We describe the algorithm that generates a simulated time series of returns. The notation corresponds to that of the text and of STW. Let $\{f_t,\ t = R, \ldots, T\}$ denote the original series of returns. For $b = 1, \ldots, B$, the bootstrapped series of returns $\{f_t^b,\ t = R, \ldots, T\}$ are obtained as follows. $q \in [0, 1]$ is a smoothing parameter.

1. Set $t = R$. Draw the index $\theta(t)$ at random, independently and uniformly from $\{R, \ldots, T\}$. Set $f_t^b = f_{\theta(t)}$.

2. Set $t = t + 1$. If $t > T$, stop. Otherwise, draw a random variable U from the standard uniform distribution.

   (a) If $U < q$, draw $\theta(t)$ at random, independently and uniformly from $\{R, \ldots, T\}$.

   (b) If $U \geq q$, set $\theta(t) = \theta(t-1) + 1$. If $\theta(t) > T$, set $\theta(t) = R$.

   Set $f_t^b = f_{\theta(t)}$.

3. Repeat step 2.

The stationary bootstrap resamples blocks of varying length from the original data. The average block length equals $1/q$. The parameter q has to be chosen according to the dependence exhibited by the data. We follow STW who set the average block length to 10 (i.e., $q = 0.1$). STW show that the results are robust to the choice of q.

For each simulated series of return, we compute the corresponding performance measure φ^b, $b = 1, \ldots, B$. The p-value is obtained by comparing the original performance φ to the quantiles of φ^b, $b = 1, \ldots, B$. Whereas STW perform only 500 bootstrap iterations, we set $B = 1000$ [19].

[19] As our programs are written in C++, we can afford to perform more bootstrap iterations.

B Estimation of the FDR

Suppose that we tested the null hypothesis of no abnormal performance for each trading rule and obtained the l corresponding p-values. We call a trading rule significant (i.e., reject the null hypothesis) when its p-value is less than or equal to some threshold γ. Storey (2003) shows that an estimator of the FDR is

$$\widehat{\text{FDR}}(\gamma) = \frac{\widehat{F}}{\widehat{R}} = \frac{\widehat{\pi}_0 \; l \; \gamma}{\# \left\{ p_k \leq \gamma; \; k = 1, \ldots, l \right\}},$$

where $\widehat{\pi}_0$ is an estimate of $\pi_0 \equiv l_0/l$, the proportion of rules in the population generating no abnormal performance. Hence, measuring the FDR boils down to the estimation of π_0, which we describe in the following section.

C Estimation of π_0

In order to estimate π_0, Storey (2002) proposes a method exploiting the fact that, for a two-sided test, null p-values are uniformly distributed over $[0, 1]$, whereas p-values of alternative models tend to be close to zero. Figure 8 shows the histogram density of p-values corresponding to our $l = 7'846$ trading rules. We see that beyond 0.6, the histogram looks fairly flat, which indicates that there are mostly null p-values in this region. The height of this flat portion gives a conservative estimate of the overall proportion of null p-values:

$$\widehat{\pi}_0(\lambda) = \frac{\# \left\{ p_k > \lambda; \; k = 1, \ldots, l \right\}}{l(1 - \lambda)},$$

which involves the tuning parameter λ. It is possible to automate the selection of λ. However, as $\widehat{\pi}_0$ is not sensitive to the choice of λ when the number of rules is high, we set $\lambda = 0.6$ by visually examining the histograms[20].

[Figure 8]

D Estimation of the FDR$^+$ and the FDR$^-$

The FDR measures the proportion of false discoveries without distinction between trading rules with positive or negative performance. Since the multiple test we perform in Equation (1) is two-sided with equal tail significance $\gamma/2$, the

[20]The automated method described in Storey (2002) produces almost identical estimates of π_0.

false discoveries are spread evenly between outperforming and underperforming trading rules. Based on that observation, Barras et al. (2007) propose the following estimators for the FDR separately among the rules yielding positive and negative performance:

$$\widehat{\mathrm{FDR}}^{+}(\gamma) = \frac{\widehat{F}^{+}}{\widehat{R}^{+}} = \frac{\frac{1}{2}\,\widehat{\pi}_0\; l\;\gamma}{\#\left\{p_k \leq \gamma,\ \varphi_k > 0;\ k = 1, \ldots, l\right\}},$$

$$\widehat{\mathrm{FDR}}^{-}(\gamma) = \frac{\widehat{F}^{-}}{\widehat{R}^{-}} = \frac{\frac{1}{2}\,\widehat{\pi}_0\; l\;\gamma}{\#\left\{p_k \leq \gamma,\ \varphi_k < 0;\ k = 1, \ldots, l\right\}}.$$

E Estimation of π_A^+ and π_A^-

Appendix C shows how to estimate π_0, from which we can deduce $\pi_A = 1 - \pi_0$, the proportion of rules with abnormal (i.e., non zero) performance in the population. It is useful to split π_A into the proportions of rules with positive (π_A^+) and negative abnormal performance (π_A^-), which can be written as:

$$\pi_A^+ = \frac{T^+(\gamma) + A^+(\gamma)}{l}, \quad \pi_A^- = \frac{T^-(\gamma) + A^-(\gamma)}{l}.$$

The notations correspond to Table 2. $T^+(\gamma)$ denotes the number of alternative models with positive performance and a p-value smaller than γ. $A^+(\gamma)$ denotes the number of alternative models with positive performance which are not rejected by the hypothesis test (i.e., with a p-value greater than γ). $T^-(\gamma)$ and $A^-(\gamma)$ are defined accordingly for negative performance.

Using the same approach as in Appendix D, we estimate $T^+(\gamma)$ and $T^-(\gamma)$ with:

$$\widehat{T}^+(\gamma) = \widehat{R}^+(\gamma) - \widehat{F}^+(\gamma) = \#\left\{p_k \leq \gamma,\ \varphi_k > 0;\ k = 1, \ldots, l\right\} - \frac{1}{2}\,\widehat{\pi}_0\; l\;\gamma,$$

$$\widehat{T}^-(\gamma) = \widehat{R}^-(\gamma) - \widehat{F}^-(\gamma) = \#\left\{p_k \leq \gamma,\ \varphi_k < 0;\ k = 1, \ldots, l\right\} - \frac{1}{2}\,\widehat{\pi}_0\; l\;\gamma.$$

As we increase γ, $A^+(\gamma)$ and $A^-(\gamma)$ tend to zero, while $T^+(\gamma)$ and $T^-(\gamma)$ increase. Hence, by taking a sufficiently high value γ^*, we can estimate π_A^+ and π_A^- with:

$$\widehat{\pi}_A^+ = \frac{\widehat{T}^+(\gamma^*)}{l}, \quad \widehat{\pi}_A^- = \frac{\widehat{T}^-(\gamma^*)}{l},$$

as explained in Barras et al. (2007). We set $\gamma^* = 0.3$, which corresponds to the value for which $\widehat{\pi}_A^+$ and $\widehat{\pi}_A^-$ become constant.

F Controlling the portfolio FDR^+ level

Storey, Taylor and Siegmund (2004) show that the FDR point estimates can be used to define valid FDR controlling procedures. Hence, we can derive the following algorithm that allows the construction of a portfolio of trading rules with a FDR^+ level fixed at at predetermined rate. The algorithm starts with the rule having the smallest p-value (and a positive performance). Then, the rule corresponding to the next p-value is added and the FDR^+ recomputed. This process is repeated until we reach the desired FDR^+ rate, and we select the rules resulting in a FDR^+ not greater than the predetermined level.

References

Abramovich, F., Benjamini, Y., Donoho, D. L. and Johnstone, I. M. (2006), 'Adapting to unknown sparsity by controlling the false discovery rate', *The Annals of Statistics* **34**(2), 584–653.

Allen, F. and Karjalainen, R. (1999), 'Using genetic algorithms to find technical trading rules', *Journal of Financial Economics* **51**, 245–271.

Barberis, N., Shleifer, A. and Vishny, R. (1998), 'A model of investor sentiment', *Journal of Financial Economics* **49**, 307–343.

Barras, L., Scaillet, O. and Wermers, R. (2007), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *Working paper* .

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* **57**, 289–300.

Bessembinder, H. and Chan, K. (1998), 'Market efficiency and the returns to technical analysis', *Financial Management* **27**, 5–17.

Blume, L., Easley, D. and O'Hara, M. (1994), 'Market statistics and technical analysis: The role of volume', *Journal of Finance* **49**, 153–181.

Blume, M. (1993), 'Soft dollars and the brokerage industry', *Financial Analysts Journal* **49**(2), 36–44.

Breen, W. J., Hodrick, L. S. and Korajczyk, R. A. (2002), 'Predicting equity liquidity', *Management Science* **48**, 470–483.

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

Brown, S. J. and Goetzmann, W. N. (1995), 'Performance persistence', *The Journal of Finance* **50**(2), 679–698.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), 57–82.

D'Avolio, G. (2002), 'The market for borrowing stock', *Journal of Financial Economics* **66**, 271–306.

DeBondt, W. and Thaler, R. (1985), 'Does the stock market overreact?', *Journal of Finance* **40**, 793–805.

Duffie, D., Gârleanu, N. and Pedersen, L. H. (2002), 'Securities lending, shorting, and pricing', *Journal of Financial Economics* **66**, 307–339.

Fama, E. and Blume, M. (1966), 'Filter rules and stock-market trading', *Journal of Business* **39**, 226–241.

Geczy, C. C., Musto, D. K. and Reed, A. V. (2002), 'Stocks are special too: an analysis of the equity lending market', *Journal of Financial Economics* **66**, 241–269.

Genovese, C. and Wasserman, L. (2004), 'A stochastic processs approach to false discovery control', *The Annals of Statistics* **32**(3), 1035–1061.

Glosten, C. C. and Harris, L. E. (1988), 'Estimating the components of the bid/ask spread', *Journal of Financial Economics* **21**, 123–142.

Hansen, P. R. (2005), 'A test for superior predictive ability', *Journal of Business and Economic Statistics* **23**, 365–380.

Hsu, P.-H. and Kuan, C.-M. (2005), 'Reexamining the profitability of technical analysis with data snooping checks', *Journal of Financial Econometrics* **3**, 606–628.

Jacquier, E. and Yao, T. (2002), 'Evaluating dynamic trading strategies: The free lunch was no banquet', *Working paper* .

Jegadeesh, N. (2000), 'Discussion of foundations of technical analysis', *Journal of Finance* **55**, 1765–1770.

Jones, C. M. and Lamont, O. A. (2002), 'Short-sale constraints and stock returns', *Journal of Financial Economics* **66**, 207–239.

Kavajecz, K. A. and Odders-White, E. R. (2004), 'Technical analysis and liquidity provision', *The Review of Financial Studies* **17**, 1043–1071.

Keim, D. B. and Madhavan, A. (1997), 'Transactions costs and investment style: an inter-exchange analysis of institutional equity trades', *Journal of Financial Economics* **46**, 265–292.

Korajczyk, R. A. and Sadka, R. (2004), 'Are momentum profits robust to trading costs?', *Journal of Finance* **59**, 1039–1082.

Lesmond, D. A., Schill, M. J. and Zhou, C. (2004), 'The illusory nature of momentum profits', *Journal of Financial Economics* **71**, 349–380.

Lo, A. and MacKinley, C. (1990), 'Data snooping biases in tests of financial asset pricing models', *The Review of Financial Studies* **3**, 431–468.

Lo, A., Mamaysky, H. and Wang, J. (2000), 'Foundations of technical analysis: computations algorithms, statistical inference, and empirical implementation', *The Journal of Finance* **55**, 1705–1765.

Mitchell, M. and Pulvino, T. (2001), 'Characteristics of risk and return in risk arbitrage', *The Journal of Finance* **56**, 2135–2175.

Neftci, S. (1991), 'Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of technical analysis', *Journal of Business* **64**, 549–571.

Osler, C. L. and Chang, P. H. K. (1995), 'Head and shoulders: Not just a flaky pattern', *Federal Reserve Bank of New York Staff Report* **4**, 1–65.

Politis, D. and Romano, J. (1994), 'The stationary bootstrap', *Journal of the American Statistical Association* **89**, 1303–1313.

Romano, J. and Wolf, M. (2005), 'Stepwise multiple testing as formalized data snooping', *Econometrica* **73**, 1237–1282.

Schmidt, A. B. (2002), 'Why technical tradingmay be successful? a lesson from the agent-based modeling', *Physica A* **303**, 185–188.

Stoll, H. R. and Whaley, R. E. (1983), 'Transaction costs and the small firm effect', *Journal of Financial Economics* **12**, 57–79.

Storey, J. (2002), 'A direct approach to false discovery rates', *Journal of the Royal Statistical Society: Series B* **64**, 479–498.

Storey, J. (2003), 'The positive false discovery rate: a bayesian interpretation and the q-value', *The Annals of Statistics* **31**, 2013–2035.

Storey, J., Taylor, J. E. and Siegmund, D. (2004), 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach', *Journal of the Royal Statistical Society: Series B* **66**, 187–205.

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

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

White, H. (2000), 'A reality check for data snooping', *Econometrica* **68**, 1097–1126.

Subperiod 1:	January 1897 – December 1914
Subperiod 2:	January 1915 – December 1938
Subperiod 3:	January 1939 – June 1962
Subperiod 4:	July 1962 – December 1986
Subperiod 5:	January 1987 – December 1996
Subperiod 6:	January 1997 – July 2007

Table 1: Sample periods

	H_0 rejected	H_0 not rejected	Total
Null H_0 true	F	N	l_0
Alternative H_A true	T	A	l_A
Total	R	$l - R$	l

Table 2: Possible outcomes resulting from l hypotheses tests

Sample	RW portfolio		Best rule		DJIA
	Mean return (%)	Port. size	Mean return (%)	BRC p-value	Mean return (%)
Subperiod 1 (1897–1914)	15.2	33	19.0	0.00	0.6
Subperiod 2 (1915–1938)	16.7	2	16.8	0.03	1.8
Subperiod 3 (1939–1962)	12.0	41	25.3	0.00	5.3
Subperiod 4 (1962–1986)	15.9	26	24.1	0.00	4.2
Subperiod 5 (1987–1996)	-	0	13.8	0.12	13.4
Subperiod 6 (1997–2007)	-	0	8.9	0.97	4.2
100 years (1897–1996)	7.7	201	15.9	0.00	4.5

Sample	FDR exploratory			FDR control (10%)			
	Mean return (%)	Port. size	$\widehat{FDR}^+$ (%)	Mean return (%)	Port. size	$\widehat{\pi}_A^+$ (%)	$\widehat{\pi}_A^-$ (%)
Subperiod 1 (1897–1914)	10.0	579	3	7.5	1848	39	0
Subperiod 2 (1915–1938)	8.4	114	22	10.2	31	24	0
Subperiod 3 (1939–1962)	6.0	749	2	3.7	4235	48	0
Subperiod 4 (1962–1986)	8.4	115	31	9.7	82	1	0
Subperiod 5 (1987–1996)	10.6	100	21	13.8	1	19	6
Subperiod 6 (1997–2007)	-	0	-	-	0	0	16
100 years (1897–1996)	4.8	1226	1	2.9	4342	50	0

Table 3: **Performance over time under the mean return criterion, ignoring transaction costs.** This table reports performance results, in each of the sample periods. The top panel presents results for the portfolio constructed using RW method, for the best rule, and for the Dow Jones Industrial index. The bottom panel presents results for the FDR exploratory analysis, for the portfolio obtained by controlling the FDR^+ at 10%, and the proportion of rules generating positive ($\widehat{\pi}_A^+$) and negative ($\widehat{\pi}_A^-$) performance.

	RW portfolio		Best rule		DJIA
Sample	Sharpe ratio	Port. size	Sharpe ratio	BRC p-value	Sharpe ratio
Subperiod 1 (1897–1914)	1.24	39	1.18	0.00	-0.12
Subperiod 2 (1915–1938)	-	0	0.73	0.10	0.06
Subperiod 3 (1939–1962)	1.50	59	2.34	0.00	0.41
Subperiod 4 (1962–1986)	1.43	15	1.45	0.00	-0.16
Subperiod 5 (1987–1996)	-	0	0.84	0.92	0.66
Subperiod 6 (1997–2007)	-	0	0.60	1.00	0.13
100 years (1897–1996)	0.70	85	0.82	0.00	0.12

	FDR exploratory			FDR control (10%)			
Sample	Sharpe ratio	Port. size	$\widehat{FDR}^+$ (%)	Sharpe ratio	Port. size	$\widehat{\pi}_A^+$ (%)	$\widehat{\pi}_A^-$ (%)
Subperiod 1 (1897–1914)	0.86	558	5	0.77	913	24	0
Subperiod 2 (1915–1938)	0.76	128	20	0.77	57	24	0
Subperiod 3 (1939–1962)	0.99	751	2	0.67	3577	45	0
Subperiod 4 (1962–1986)	1.14	71	18	1.17	58	0	50
Subperiod 5 (1987–1996)	1.32	11	100	0.87	3	0	17
Subperiod 6 (1997–2007)	0.70	10	100	0.59	3	0	34
100 years (1897–1996)	0.50	571	5	0.45	847	18	0

Table 4: **Performance over time under the Sharpe ratio criterion, ignoring transaction costs.** This table reports performance results, in each of the sample periods. The top panel presents results for the portfolio constructed using RW method, for the best rule, and for the Dow Jones Industrial index. The bottom panel presents results for the FDR exploratory analysis, for the portfolio obtained by controlling the FDR^+ at 10%, and the proportion of rules generating positive ($\widehat{\pi}_A^+$) and negative ($\widehat{\pi}_A^-$) performance.

Summary statistics	FDR portfolio	RW portfolio	Best rule
Annualized average return (%)	3.7	12.0	25.3
Portfolio size	4’235	41	1
Total number of trades	546’735	38’686	2’722
Number of winning trades	198’764	10’229	645
Number of losing trades	347’971	28’457	2’077
Average number of days per trade	30.3	5.5	2.3
Average number of days per winning trade	54.4	12.8	4.4
Average number of days per losing trade	16.5	2.9	1.6
Annualized average return per trade (%)	0.6	0.2	0.1
Number of long trades	301’573	20’176	1’361
Number of long winning trades	128’144	6’254	376
Number of long losing trades	173’429	13’922	985
Average number of days per long trade	37.7	6.3	2.4
Annualized average return per long trade (%)	1.2	0.3	0.1
Number of short trades	245’162	18’510	1’361
Number of short winning trades	70’620	3’975	269
Number of short losing trades	174’542	14’535	1’092
Average number of days per short trade	21.2	4.6	2.1
Annualized average return per short trade (%)	-0.2	0.0	0.0
Proportion of winning signals (%)	34	45	56
Proportion of neutral signals (%)	36	16	0
Proportion of losing signals (%)	31	40	44

Table 5: **Summary statistics: subperiod 3 (1939–1962) with the mean return criterion and no transaction costs.** This table provides summary statistics for the portfolio obtained by controlling the FDR^+ at 10%, for the RW portfolio, and for the best rule in the sample.

	Out-of-sample				In-sample			
		Portfolio size				Portfolio size		
	Mean return (%)	1^{st} quart.	median	3^{rd} quart.	Mean return (%)	1^{st} quart.	median	3^{rd} quart.
FDR (1%) portfolio	5.5	1	5	28	19.6	0	3	13
FDR (5%) portfolio	5.3	1	5	28	19.6	0	4	16
FDR (10%) portfolio	5.3	1	5	47	18.9	0	4	28
FDR (20%) portfolio	5.3	1	5	3'352	18.1	0	4	53
RW portfolio	11.4	0	2	3	8.1	0	0	0
Best portfolio	2.4	200	200	200	19.1	200	200	200
Best rule	12.7	-	-	-	31.1	-	-	-
BRC rule	11.3	-	-	-	8.3	-	-	-
DJIA	8.1	-	-	-	8.1	-	-	-

Table 6: **Persistence tests under the mean return criterion, subperiod 3 (1939–1962), ignoring transaction costs.** This table compares in-sample and out-of sample performance. It also provides the first, second, and third quartiles of the distribution of the size of the portfolios. Results are reported for portfolios where the FDR^+ is controlled at different levels, for the portfolio constructed using RW method, for the portfolio consisting of the 200 best rules, for the best rule in the sample, for the BRC rule, and for the Dow Jones Industrial Average index. The left-hand side contains the results when the performance is tested out-of-sample (i.e., when the trading rules are selected ex ante). In-sample results are displayed on the right-hand side.

	Out-of-sample				In-sample			
	Sharpe ratio	Portfolio size			Sharpe ratio	Portfolio size		
		1st quart.	median	3rd quart.		1st quart.	median	3rd quart.
FDR (1%) portfolio	0.75	12	22	50	2.82	6	22	48
FDR (5%) portfolio	0.67	12	22	55	2.76	6	22	50
FDR (10%) portfolio	0.41	12	22	253	2.47	6	22	221
FDR (20%) portfolio	0.45	12	51	3'184	2.32	6	28	289
RW portfolio	1.02	0	0	2	1.80	0	0	0
Best portfolio	0.31	200	200	200	2.43	200	200	200
Best rule	1.32	-	-	-	2.99	-	-	-
BRC rule	1.01	-	-	-	1.80	-	-	-
DJIA	0.68	-	-	-	0.68	-	-	-

Table 7: **Persistence tests under the Sharpe ratio criterion, subperiod 3 (1939–1962), ignoring transaction costs.** This table compares in-sample and out-of sample performance. It also provides the first, second, and third quartiles of the distribution of the size of the portfolios. Results are reported for portfolios where the FDR^+ is controlled at different levels, for the portfolio constructed using RW method, for the portfolio consisting of the 200 best rules, for the best rule in the sample, for the BRC rule, and for the Dow Jones Industrial Average index. The left-hand side contains the results when the performance is tested out-of-sample (i.e., when the trading rules are selected ex ante). In-sample results are displayed on the right-hand side.

Sample	FDR portfolio	RW portfolio	200 best rules	Best rule
Subperiod 1 (1897–1914)	0.0003	0.0090	0.0016	0.3997
Subperiod 2 (1915–1938)	0.0071	0.0716	0.0026	0.1444
Subperiod 3 (1939–1962)	0.0001	0.0039	0.0013	0.1140
Subperiod 4 (1962–1986)	0.0017	0.0058	0.0010	0.1051
Subperiod 5 (1987–1996)	31.2210	-	0.0272	31.2210
Subperiod 6 (1997–2007)	-	-	0.0030	1.4137
100 years (1897–1996)	0.0001	0.0010	0.0010	0.1357

Table 8: **Break-even transaction costs (%).** This table reports the level of transaction costs that would just eliminate the profits to traders using technical rules, in each of the sample periods. Results are provided for the portfolio obtained by controlling the FDR$^+$ at 10%, for the RW portfolio, for the portfolio consisting of the 200 best rules, and for the best rule in the sample.

(a) Average yearly return (%)

Sample	FDR portfolio	RW portfolio	200 best rules	Best rule
Subperiod 1 (1897–1914)	7.5	15.2	12.8	19.0
Subperiod 2 (1915–1938)	10.2	16.7	10.1	16.8
Subperiod 3 (1939–1962)	3.7	12.0	8.4	25.3
Subperiod 4 (1962–1986)	9.7	15.9	7.6	24.1
Subperiod 5 (1987–1996)	13.8	-	10.0	13.8
Subperiod 6 (1997–2007)	-	-	3.5	9.8
100 years (1897–1996)	2.9	7.7	7.7	15.9

(b) Average yearly number of transactions

Sample	FDR portfolio	RW portfolio	200 best rules	Best rule
Subperiod 1 (1897–1914)	27’108	1’690	8’205	47
Subperiod 2 (1915–1938)	1’432	233	3’929	116
Subperiod 3 (1939–1962)	27’711	3’065	6’400	222
Subperiod 4 (1962–1986)	5’719	2’749	7’653	229
Subperiod 5 (1987–1996)	0	-	370	0
Subperiod 6 (1997–2007)	-	-	1’143	7
100 years (1897–1996)	31’783	7’929	7’911	117

Table 9: **Average yearly return (%) and average yearly number of transactions.** This table reports the average yearly return (%) (Panel (a)) and the average yearly number of transactions (Panel (b)), in each of the sample periods. Results are provided for the portfolio obtained by controlling the FDR$^+$ at 10%, for the RW portfolio, for the portfolio consisting of the 200 best rules, and for the best rule in the sample.

	Out-of-sample				In-sample			
	Mean return (%)	Portfolio size			Mean return (%)	Portfolio size		
		1st quart.	median	3rd quart.		1st quart.	median	3rd quart.
FDR (1%) portfolio	-2.2	0	0	1	7.4	0	0	1
FDR (5%) portfolio	-2.5	0	0	1	7.6	0	0	1
FDR (10%) portfolio	-2.5	0	0	1	7.2	0	0	1
FDR (20%) portfolio	-1.3	0	0	1	6.4	0	0	1
RW portfolio	0.0	0	0	0	1.6	0	0	0
Best portfolio	-0.2	200	200	200	16.5	200	200	200
Best rule	-3.0	-	-	-	24.1	-	-	-
BRC rule	0.0	-	-	-	1.6	-	-	-
DJIA	8.1	-	-	-	8.1	-	-	-

Table 10: **Persistence tests under the mean return criterion, subperiod 3 (1939–1962), 20 basis points one-way transaction costs, 50 basis points yearly lending fees.** This table compares in-sample and out-of sample performance. It also provides the first, second, and third quartiles of the distribution of the size of the portfolios. Results are reported for portfolios where the FDR$^+$ is controlled at different levels, for the portfolio constructed using RW method, for the portfolio consisting of the 200 best rules, for the best rule in the sample, for the BRC rule, and for the Dow Jones Industrial Average index. The left-hand side contains the results when the performance is tested out-of-sample (i.e., when the trading rules are selected ex ante). In-sample results are displayed on the right-hand side.

	Out-of-sample				In-sample			
		Portfolio size				Portfolio size		
	Sharpe ratio	1st quart.	median	3rd quart.	Sharpe ratio	1st quart.	median	3rd quart.
FDR (1%) portfolio	-0.42	0	2	8	1.77	0	1	3
FDR (5%) portfolio	-0.31	0	2	8	1.73	0	1	3
FDR (10%) portfolio	-0.22	0	2	8	1.71	0	1	3
FDR (20%) portfolio	-0.36	0	2	8	1.59	0	1	3
RW portfolio	0.00	0	0	0	0.00	0	0	0
Best portfolio	-0.13	200	200	200	1.96	200	200	200
Best rule	-0.69	-	-	-	2.29	-	-	-
BRC rule	0.00	-	-	-	0.00	-	-	-
DJIA	0.68	-	-	-	0.68	-	-	-

Table 11: **Persistence tests under the Sharpe ratio criterion, subperiod 3 (1939–1962), 20 basis points one-way transaction costs, 50 basis points yearly lending fees.** This table compares in-sample and out-of sample performance. It also provides the first, second, and third quartiles of the distribution of the size of the portfolios. Results are reported for portfolios where the FDR^+ is controlled at different levels, for the portfolio constructed using RW method, for the portfolio consisting of the 200 best rules, for the best rule in the sample, for the BRC rule, and for the Dow Jones Industrial Average index. The left-hand side contains the results when the performance is tested out-of-sample (i.e., when the trading rules are selected ex ante). In-sample results are displayed on the right-hand side.

Trans. costs (bps)	Lending fees (bps)	Mean return criterion	Sharpe ratio criterion
0	0	2-days on-balance volume*	2-days moving average, 0.1% band*
0	50	2-days on-balance volume*	2-days moving average, 0.1% band*
0	100	2-days on-balance volume*	2-days moving average, 0.1% band*
10	0	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
10	50	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
10	100	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
20	0	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
20	50	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
20	100	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
30	0	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
30	50	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
30	100	75- and 100-days on-balance volume, 50 days holding period	40- and 150-days on-balance volume, 5% band
40	0	Filter rule, 14% position initiation, 10 days lag for extrema definition	40- and 150-days on-balance volume, 5% band
40	50	Filter rule, 14% position initiation, 10 days lag for extrema definition	40- and 150-days on-balance volume, 5% band
40	100	Filter rule, 14% position initiation, 10 days lag for extrema definition	40- and 150-days on-balance volume, 5% band

Table 12: **Best technical trading rule under increasing transaction costs, subperiod 3 (1939–1962).** This table reports the historically best-performing trading rule, chosen with respect to the mean return and the Sharpe ratio criterion, for increasing levels of one-way transaction costs and lending fees. An asterisk (i.e., *) indicates that, according to the BRC, the performance of the rule remains significant after accounting for data snooping.

Summary statistics	FDR portfolio	RW portfolio	Best rule
Annualized average return (%)	5.5	0.0	7.4
Portfolio size	14	0	1
Total number of trades	239	0	37
Number of winning trades	118	0	15
Number of losing trades	121	0	22
Average number of days per trade	189.0	0.0	166.2
Average number of days per winning trade	327.5	0.0	273.3
Average number of days per losing trade	54.0	0.0	93.2
Annualized average return per trade (%)	7.7	0.0	5.2
Number of long trades	179	0	19
Number of long winning trades	99	0	7
Number of long losing trades	80	0	12
Average number of days per long trade	230.2	0.0	223.7
Annualized average return per long trade (%)	10.4	0.0	8.9
Number of short trades	60	0	18
Number of short winning trades	19	0	8
Number of short losing trades	41	0	10
Average number of days per short trade	66.3	0.0	105.6
Annualized average return per short trade (%)	-0.3	0.0	1.3
Proportion of winning signals (%)	29	0	52
Proportion of neutral signals (%)	47	0	0
Proportion of losing signals (%)	24	0	48

Table 13: **Summary statistics: subperiod 3 (1939–1962), mean return criterion, 20 basis points one-way transaction costs, 50 basis points yearly lending fees.** This table provides summary statistics for the portfolio obtained by controlling the FDR^+ at 10%, for the RW portfolio, and for the best rule in the sample.

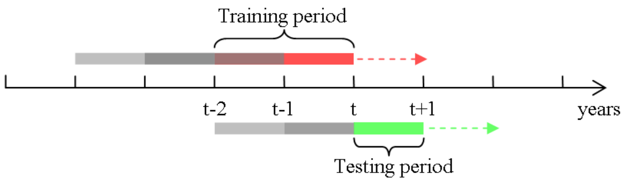

Figure 1: **Persistence test.** At time t, the portfolio is re-formed in order to include the rules that performed best in a training period corresponding to the last two years. The performance is then evaluated out-of-sample during a testing period of one year. This process is repeated every year.

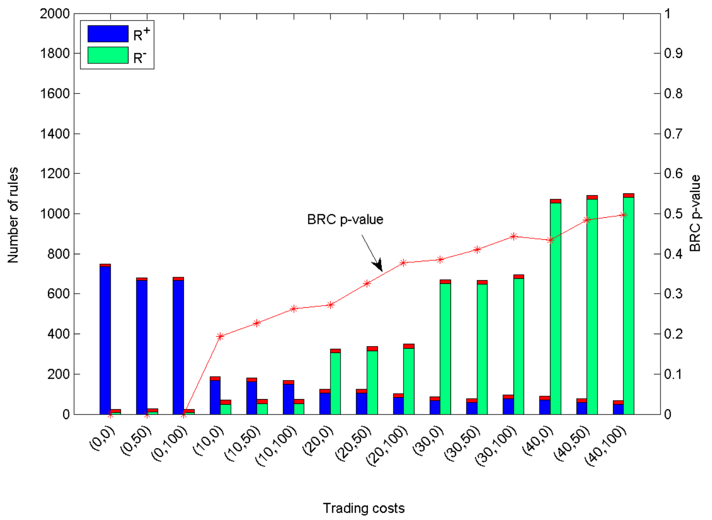

Figure 2: **FDR exploratory analysis over increasing transaction costs: subperiod 3 (1939–1962) with the mean return criterion.** This figure displays the number of rules detected as significantly positive ($\widehat{R}^+$) and significantly negative ($\widehat{R}^-$), and the corresponding false discoveries (displayed in red). On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points). The solid line represents the BRC p-value. The p-value can be read from the right-hand axis.

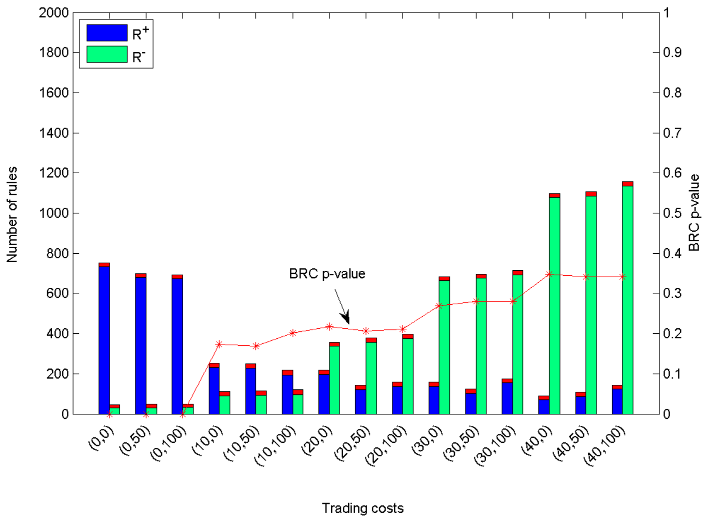

Figure 3: **FDR exploratory analysis over increasing transaction costs: subperiod 3 (1939–1962) with the Sharpe ratio criterion.** This figure displays the number of rules detected as significantly positive ($\widehat{R}^{+}$) and significantly negative ($\widehat{R}^{-}$), and the corresponding false discoveries (displayed in red). On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points). The solid line represents the BRC p-value. The p-value can be read from the right-hand axis.

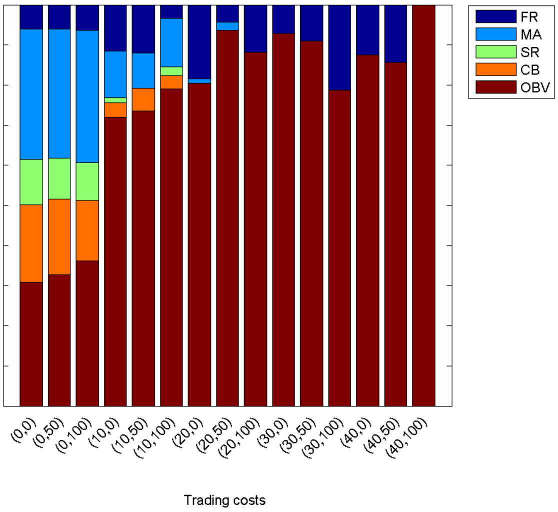

Figure 4: **Repartition by technical rules categories over increasing transaction costs: subperiod 3 (1939–1962) with the mean return criterion.** This figure displays the evolution of the composition of the 10% FDR^+ portfolio when transaction costs increase. The different categories are Filter Rules (FR), Moving Averages (MA), Support and Resistance rules (SR), Channel Breakouts (CB), and On-Balance Volume averages (OBV). On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points).

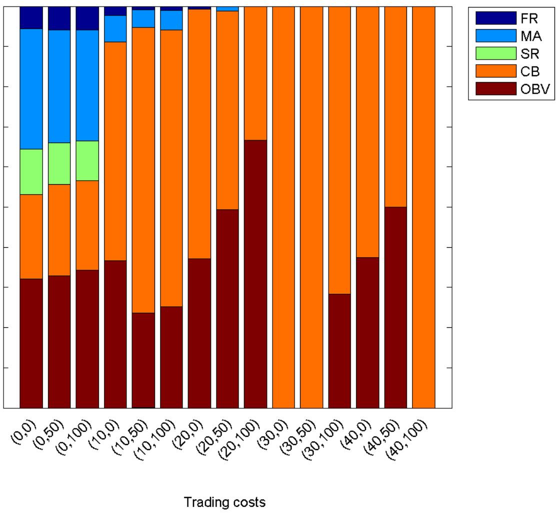

Figure 5: **Repartition by technical rules categories over increasing transaction costs: subperiod 3 (1939–1962) with the Sharpe ratio criterion.** This figure displays the evolution of the composition of the 10% FDR^+ portfolio when transaction costs increase. The different categories are Filter Rules (FR), Moving Averages (MA), Support and Resistance rules (SR), Channel Breakouts (CB), and On-Balance Volume averages (OBV). On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points).

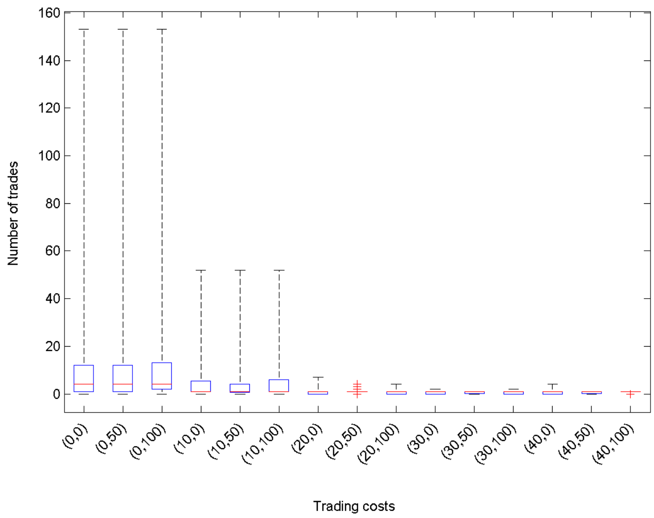

Figure 6: **Number of trades over increasing transaction costs: subperiod 3 (1939–1962) with the mean return criterion.** This figure displays boxplots of the number of trades generated by the rules composing the 10% FDR^+ portfolio when transaction costs increase. On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points).

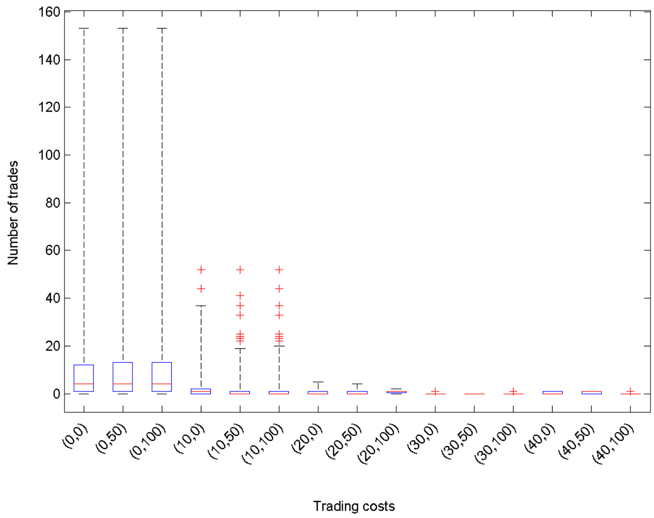

Figure 7: **Number of trades over increasing transaction costs: subperiod 3 (1939–1962) with the Sharpe ratio criterion.** This figure displays boxplots of the number of trades generated by the rules composing the 10% FDR^+ portfolio when transaction costs increase. On the horizontal axis, each bracket (x, y) corresponds to a pair of one-way transaction costs x and annualized lending fees y (in basis points).

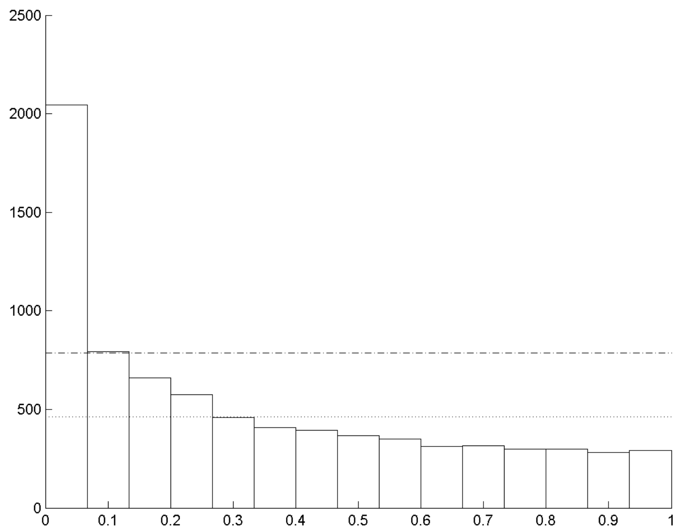

Figure 8: A density histogram of the 7'846 p-values (subperiod 3 (1939–1962), 10bps one-way transaction costs, 50bps lending fees). The dashed line is the density histogram we would expect if all rules were truly null (i.e., did not generate abnormal performance). The dotted line is at the height of our estimate of the proportion of rules that do not generate abnormal performance (i.e., π_0).