Technical Analysis Profitability and Persistence: A Discrete False Discovery Approach on MSCI Indices

Georgios Sermpinis*, Arman Hassanniakalager†, Charalampos Stasinakis‡, Ioannis Psaradellis§

Abstract

We investigate the performance of more than 21,000 technical trading rules on 12 categorical and country-specific markets over the 2004-2015 study period. For this purpose, we apply a discrete false discovery rate approach in more than 240,000 hypotheses and examine the profitability, persistence and robustness of technical analysis. In terms of our results, technical analysis has short-term value and its profitability is mainly driven by short-term momentum. Financial stress seems to have a strong negative effect in technical analysis profitability for US markets and a strong positive effect for emerging and other advanced markets.

Keywords: False Discovery Rate; Technical Analysis, Trading; Bootstrap/resampling;

JEL codes: C12; C15; C53; G11; G15; G17

* Adam Smith Business School, University of Glasgow, Glasgow, G12 8QQ, UK
† School of Management, University of Bath, Bath, UK
‡ Corresponding author: Adam Smith Business School, University of Glasgow, Glasgow, G12 8QQ, UK, email: Charalampos.Stasinakis@glasgow.ac.uk , tel.: +44 (0)141 330 7591
§ School of Economics & Finance, University of St Andrews, The Scores, St Andrews KY16 9AR, UK,

1. Introduction

In this study, we re-evaluate the profitability of Technical Analysis (TA) and we examine its persistence through a novel Discrete False Discovery Rate ($DFDR^{+/-}$) framework. More specifically, we investigate the daily data of nine individual Morgan Stanley Capital International (MSCI) indices and three general MSCI indices replicating the performance of advanced, emerging and frontier markets covering the period from 2004 to 2015. MSCI indices are important references for institutional investors (see among others, Hsu, Hsu and Kuan, 2010; Bena *et al.*, 2017) as 99% of the top global investment managers apply MSCI indices on their trading activity. They include large and mid-cap segments of the benchmark markets and thus mitigate liquidity and tradability issues, an important aspect especially for emerging and frontier markets. As for the technical trading rules considered, we employ the expanded universe of 21,195 rules of Hsu, Taylor and Wang (2016) incorporating a variety of classes of technical indicators and oscillators.

Testing a large set of technical trading rules on a given data set risks false rejections of the null ofzero performance, as some rules in a large set will, by chance, prove to be profitable ex post. A researcher first needs to adopt a Multiple Hypothesis Testing (MHT) framework in order to identify whether truly profitable rules exist. The framework should be powerful, adaptive and computationally efficient. In our study, we extend the False Discovery Rate ($FDR^{+/-}$) approach of Barras, Scaillet and Wermers (2010) and generate the Discrete False Discovery Rate ($DFDR^{+/-}$). $DFDR^{+/-}$ constructs a large-scale homogeneous discrete *p*-values framework where its main parameters are being dynamically estimated. This includes the estimation of the tuning parameter, which maps the null *p*-values in the empirical distribution. A selection that is crucial for the power of $FDR^{+/-}$ (Harvey and Liu, 2018). We employ the $DFDR^{+/-}$ in 12 different markets and try to answer the questions raised earlier related to the performance, profitability and robustness of TA. Specifically, we investigate the out-of-sample performance of the rules in a rolling-forward structure, just as fund managers do in practice, and we explore their performance persistence.

The contribution of this study is fivefold. Firstly, for the first time, the performance of such a considerable universe of more than 240,000 series of trading strategies returns is assessed. We examine and compare the trading power of TA in a series of advanced, emerging and frontier financial markets.

The larger universe of technical rules and the study of different markets in terms of financial development allows us to study trading specifications that were ignored from the past literature, provide robust empirical evidence and examine whether market development influences TA's profitability. Secondly, our work is the first that examines the persistence of technical rules profitability on different horizons. An element overlooked by the related studies but of outmost importance to practitioners. Short-term profitability might explain the discrepancies of the previous literature, which is focusing mainly on the study of longer periods. Thirdly, the effect of financial stress market conditions on TA is assessed. This element might also explain the time-varying profitability of technical rules observed to previous studies. Last but not least, we extend the work of Barras, Scaillet and Wermers (2010) and introduce the $DFDR^{+/-}$, an adaptive large-scale homogeneous discrete *p*-values framework. Although in large studies, such as the one we use, the effect of discretization can be mitigated, an ad-hoc selection of the tuning parameter induces bias and can lead to misleading results as it affects the power of the method. $DFDR^{+/-}$ has a similar power with $FDR^{+/-}$ (when the optimal $FDR^{+/-}$ tuning parameter is being set) but has a reduced computational time, an advantageous feature in real-world financial applications.

Regarding our findings, we find evidence of technical rules profitability after transaction costs in all markets studied, with trend-following families (e.g. moving averages and channel breakouts) dominating the contrarian ones. The profitability seems stronger in emerging and frontier markets compared to the advanced ones. The persistence of this profitability varies over the years, but it is weak in general. In addition, we note that financial stress has a varying effect on TA's performance depending on year and market under study. On average, in periods of low stress we observe profitability only in US market. This profitability is around six and one times higher when compared to other advanced and emerging markets respectively. It is interesting to see, though, that in periods of high financial stress the picture reverses, namely TA provides clearly more value for emerging markets. For example, on average the profitability in emerging markets appears more than five and four times higher than US and other advanced markets respectively. These findings are consistent with the empirical results of Smith *et al.* (2016).

The rest of this work is laid out as follows. Section 2 provides a literature review in TA and the motivation of our study. Section 3 reports the examined data set and demonstrates the technical trading rules as well as the performance metrics employed in this study. Section 4 presents the methodology under study. Section 5 presents the main empirical results related to OOS profitability, the performance persistence and the effect of stressed market conditions. Finally, Section 6 provides our concluding remarks. In the Appendix, we present a series of Monte Carlo simulations on $DFDR^{+/-}$.

2. Literature Review and Motivation

Section 2.1 presents a literature review on the profitability of TA in stocks indices. In Section 2.2 we discuss how the contradictions in the empirical literature motivate our study.

2.1 Literature Review

TA constitutes the application of simple mathematical formulations or graphical representations of financial assets' time series to explore trading opportunities. In its algorithmic, and thus more quantitative form, it utilizes the analysis of the asset's price history, volume data and summary statistics, through mathematical tools, usually referred to as technical indicators and oscillators. Even though this form of analysis has been widely exploited by both investors and academics over the years, there is a long and ongoing discussion about whether it truly has predictive power and can generate significant profitability in equities markets. Previous literature is split into studies highlighting the genuine profitability of TA (see among others, Brock, Lakonishok and LeBaron, 1992; Hsu, Hsu and Kuan, 2010; Batten *et al.*, 2018; Jamali and Yamani, 2019) and those arguing against it (see among others, Sullivan, Timmermann and White, 1999; Bajgrowicz and Scaillet, 2012; Urquhart, Bartosz and Hudson, 2015). For example, Brock, Lakonishok and LeBaron (1992) apply two trading strategies to the Dow Jones Industrial Average and report strong support for TA. Hsu, Hsu and Kuan (2010) study stocks, exchange traded funds, and MSCI indices around the world and argue that technical trading rules have significant predictive power for these markets over certain market development stages. On the other hand, Bajgrowicz and Scaillet (2012) examine the profitability of TA by studying a pool of 7,864 trading rules and conclude that TA offers no 'economic value'.

In the studies that argue in favour of TA, success is associated with asset pricing anomalies, such as momentum (Jegadeesh and Titman, 1993; Asness, Moskowitz and Pedersen, 2013) and reversal (DeBondt and Thaler, 1985; Jegadeesh, 1990). Some recent studies have tried to use this information and employ technical trading rules for the creation of universal trend factors and the optimization of asset allocation in portfolio construction (see Zhu and Zhou, 2009; Han, Zhou and Zhu, 2016). From the point of view of practitioners, the use of TA is not debatable, and many successful hedge funds and portfolio managers make substantial use of technical trading rules (see Fung and Hsieh, 2001). On the opposite side of the argument, there is related research arguing that the findings of the aforementioned authors are due to inadequate estimation of transaction costs or extensive experimentation (i.e. data snooping). Goldbaum (2003) argues that variations in the TA's popularity results in unsustainable periods of positive profits coupled with long-term losses. Thus, any profitability will be limited in small periods and the performance of TA is not sustainable. Sullivan, Timmermann and White (1999) and Bajgrowicz and Scaillet (2012) control for luck with the Reality Check and the $FDR^{+/-}$ approaches, respectively, and argue that TA has no value after considering transaction costs.

Additionally, the literature provides insights regarding the link of TA with market development and fluctuations related to business cycles, periods of recessions or high levels of financial stress. Although most studies focus on developed markets (usually US), the literature towards investigating the profitability of technical rules in frontier and emerging markets is growing (McKenzie, 2007; Zaremba, 2020). Marshall, Nguyen and Visaltanachoti (2013) explain that frontier markets are attracting increased attention from investors as they have low world-market integration and offer high diversification benefits. On the other hand, emerging markets are more liquid and can also attract institutional investors from developed markets. Hatgioannides and Mesomeris (2007) focus their TA approach on emerging markets and they find that top performing rules are generating profits before and after the crisis, as simple trading strategies are able to capture bull phases easier than bear ones. Neely *et al.*, (2014) show that technical indicators have higher predictive power than fundamental analysis over the US equity risk premium in periods close to market troughs. Smith *et al.* (2016) also link the profitability of TA with periods high- and low-sentiment. The authors associate TA's profitability with high-sentiment periods. During such periods, financial markets are associated with increased volatility and strong upward or downward trends, which is what technical indicators are known to successfully capture. Finally, Dai, Zhu and Kang (2021) explain that stock returns predictability is associated with recessions/expansions of stock markets and they test the utility of their technical indicator in such a framework. Their results show that in period of high financial stress technical indicators have the highest incremental effect in terms of predictability.

2.2 Research Aims and Motivation

Previous studies in TA tend to be narrowly focused on specific aspects of equity indices. For example, they only look at a single market index, make use of a restricted number and classes of technical trading rules, perform no disaggregated analysis of the profitable classes, or focus on a “sterilized” exercise of TA (e.g., no transaction costs involved), which is totally different from the way traders operate in practice. Other issues, such as the size of the in-sample/out-of-sample (IS/OOS) horizons or the level of financial stress and market development on technical trading performance are also frequently ignored. In addition, no previous study has consistently measured the persistence of TA, the role of frequency in portfolio rebalancing or the main drivers of its outperformance. Researchers and practitioners in the field of finance and economics are aware of the past behaviour of financial markets. The only true out-of-sample is the live trading experience (Arnott, Harvey and Markowitz, 2019). Knowledge of the past events can lead the research design and the outcomes of a study. This paper answers to these debates and offers a comprehensive empirical study that covers advanced, emerging and frontier markets.

Specifically, this paper attempts to examine where and if TA can add economic value, while adjusting for false discoveries. We expand the discussion of Hsu, Taylor and Wang (2016) and Batten *et al.* (2018) and we pose some interesting empirical questions that will complement the puzzling picture previously outlined. *What level of transaction cost can repel all market participants? Does technical analysis exhibit short-term profitability? If yes, how long does this profitability persist? Does the level of market development and financial stress play any role in technical analysis profitability? What is the optimal IS to OOS ratio for achieving the best performance?* The answers to these questions are given by studying twelve market indices and the application of the proposed $DFDR^{+/-}$ method.

3. Data set, Technical Trading Rules and Performance Metrics

This section presents the details of the setting at which the DFDR$^{+/-}$ is applied. In Section 3.1 the dataset of the examined markets is introduced along with information regarding the technical trading rules utilized in the study. Section 3.2 covers the performance measures used to compare the trading rules for different markets.

3.1 Data set and Technical Trading Rules

In this paper, we study nine MSCI indices that replicate the performance of markets in the United States, United Kingdom, Japan, Brazil, China, Russia, Estonia, Jordan and Morocco and the three general MSCI indices that replicate the World (developed), emerging and frontier markets in total (i.e., 12 indices in total)[1]. The MSCI indices are market cap weighted indices that reflect the holding returns of US investors in different markets. They are denoted in US dollars and include large and mid-cap segments of the benchmark markets. We conduct our exercise from the perspective of a US investor. Except from using the nine country-based MSCI indices, using the three MSCI categorial indices (i.e., World, Emerging and Frontier markets) creates the framework to build portfolios that help to avoid unintended bets and risks that may appear when investing separately in the individual indices. This is because the classification of each of the three generic indices is done by including information from at least 20 countries from each category. We use return indices with dividend adjustment for our application. The sample period for all-time series starts on 1 January 2004 and ends on 31 December 2015 (most MSCI indexes under study start in 2004)[2]. The summary statistics of the log returns of the twelve series and of the risk-free rates series are presented in Table 1.

[Table 1 here]

[1] We include representative countries from each market classification and continent based on their cap weights in the MSCI World index, MSCI emerging market index and MSCI frontier market index (https://www.msci.com/market-cap-weighted-indexes), (e.g., one from America, Europe, Pacific etc.). We avoid countries that belong to the West African Economic and Monetary Union (as they include classified securities) and those with inconsistent or shorter historical data.

[2] The MSCI indices do not consider public holidays. In case of market closure or if a security does not trade on a specific day or a specific period, MSCI carries forward previous day prices (or latest available closing price) to calculate its indices. Their index calculation methodology is presented at: https://www.msci.com › methodology › meth_docs.

All indices are leptokurtic while the risk-free rate series behaves very close to the normal curve. The UK, Brazil and Russia display quite high kurtosis. All time series except for the frontier index and the risk-free rate exhibit negative skewness (with the UK having the least). The positive autocorrelation coefficient is seen for all times series except for Japan and the US, but the reported coefficient is not statistically significant for Japan.

The purpose of technical trading strategies is to generate long (short) positions for the coming period based on historical quotes for open, high, low and closing prices along with other characteristics such as previous trends/momentums and directional movements. As already mentioned, their efficacy mainly lies with popular market price anomalies (i.e. momentum and reversals). In this study, 21,195 technical trading rules are utilized following the work of Hsu, Taylor and Wang, (2016) for each of the twelve MSCI indices under study. This comes down to more than 240,000 return series to be evaluated in terms of performance and so, an equal number of null hypotheses to be tested in terms of statistical significance. A summary of these trading rules is provided in online Appendix A.

3.2 Performance Measurement and Transaction Costs

In terms of performance metrics, we provide the annualized *mean excess return* and *Sharpe ratio*. In this way, we consider an absolute measure based on each technical trading rule's returns, such as the mean excess return, and a relative performance measure reporting the ratio of the mean excess return to the total risk of the investment in terms of excess returns' standard deviation (the Sharpe ratio). Denoting the trading signal triggered from a trading rule $j, 1 \leq j \leq \ l$ (where $= 21{,}195$) at the end of each prediction period $t-1$ $(\tau \leq t \leq T)$ as $s_{j,t-1}$, where $s_{j,t-1} = 1, 0,\ or -1$ represents a long, neutral or short position taken at time t, the mean excess return criterion $\overline{f}_{j,t}$ for the trading rule j is given by:

$$\overline{f}_{j,t} = \frac{1}{N}\sum_{t=\tau}^{T} \ln\,[1 + s_{j,t-1}(r_t - r_{f,t}) - TC_{j,t}], \qquad j = 1, \dots, l$$

where $N = T - \tau + 1$ is the number of days examined, $r_{f,t}$ is the risk-free rate at time t and $TC_{j,t}$ is the one-way transaction cost applied at time t for trading rule j. We treat transaction costs "*endogenously*" to the trading process. For instance, we deduct one-way transaction costs every time a long or short position is taken on the index. The τ is the activation period, since some of the technical trading rules use lagged values of indices up to one year (i.e., 260 trading days).

Then, the Sharpe ratio metric expression SR_j for trading rule j at time t is defined by:

$$SR_{j,t} = \frac{\overline{f}_j}{\widehat{\sigma_j}},\ j = 1, \dots, l,$$

where $\overline{f}_{j,t}$ is the mean excess return and $\widehat{\sigma_{j,t}}$ the estimated standard deviation of the mean excess return. Another important feature of the Sharpe ratio metric is its direct link with the actual *t*-statistic of the empirical distribution of a rule's returns (Harvey and Liu, 2015). Thus, such a property makes the Sharpe ratio the most appropriate criterion for our proposed multiple hypothesis testing framework. In other words, the test statistic (i.e. φ_j) is the Sharpe ratio[3].

Through the specifications described, the performance of each rule is calculated and tested for significant positive difference compared to a benchmark. Following Sullivan, Timmermann and White, (1999) and Bajgrowicz and Scaillet (2012), our benchmark is the risk-free rate that corresponds to abstaining from the market when no profitable opportunity is expected. For the risk-free rate, we use the effective federal funds rate reported by the Federal Reserve in the US. Since the quotes for the risk-free rate are reported on an annual basis, we transform the rates into the daily values by $r_{f,t} = (1 + S_t)^{\frac{1}{260}} - 1$, where $r_{f,t}$ is the estimated daily rate and S_t is the quoted federal funds rate. Alternatively, the benchmark can be defined as the buy and hold strategy on the MSCI World index or further to a combination of bonds and stock indices.

In this study, we consider a one-way proportional transaction cost of 25 bps for advanced markets (US, UK, Japan, and Developed) and 50 bps for the other markets. These costs correspond to the total

[3] We observe similar trends in our results with the Sortino ratio, the manipulation ratio and the annualized return as performance metrics.

costs that a large institutional investor should be expecting. Industry-based factsheets, traders, retail online brokers, hedge fund managers and the academic literature recommend a transaction cost of 25-50 bps for trading MSCI indices (see for example, Cesari and Cremonini, 2003; Eurex, 2018). MSCI (2013) suggests transaction costs up to 50 bps for their indices.

4. Data Snooping Bias and Proposed Methodology

This section initially focuses on the traditional FDR procedure used for capturing data snooping bias. We also present our methodology extension, the DFDR+/-, and discuss the application of our technique in portfolios construction.

4.1 The False Discovery Rate Procedure

Introduced by Benjamini and Hochberg (1995) as a tolerant error metric, the False Discovery Rate (FDR) measures the proportion of false discoveries among true rejections of the null hypothesis. If F and R are the number of the total Type I errors (false discoveries) and the rejected null hypotheses (total discoveries) respectively, then the FDR is estimated as $FDR = E\left(\frac{F}{R}\right)$ ratio. The FDR corresponds to the expected False Discovery Proportion (FDP). It controls the FDR at a level γ as $FDR = E(FDP) \leq \gamma$ where γ is a user defined parameter and it should not be confused with the significance level.

Storey (2002) and Storey (2003) based on the assumption that, for a two-tailed test, the true null *p*-values are uniformly distributed over the interval [0,1], whereas the *p*-values of alternative models lie close to zero, introduce a method to control the FDR. His approach utilizes information from the centre of the distribution of *t*-statistics (i.e. φ_j), which is mainly represented by non-outperforming rules. A key point regarding this direction is the precise estimation of the proportion of rules satisfying the null hypothesis, $\varphi_j = 0$, (i.e. π_0) in the entire population. A conservative estimator of the π_0 parameter is given by:

$$\widehat{\pi_0}(\lambda) = \frac{\#\{p_j > \lambda;\ \ j = 1, \dots, l\}}{l(1-\lambda)}$$

where $\lambda \in [0,1)$ is a tuning parameter indicating above which specific level the null p-values exist and # is the number of identified p-$values$ for the given criterion. The required inputs for the FDR approach are mainly the (two-sided) corresponding p-values of the performance metrics (φ_j) of each individual rule associated with null hypothesis of non-abnormal performance ($H_{0j}: \varphi_j = 0$) against the alternative of abnormal performance ($H_{Aj}: \varphi_j > 0$ or $H_{Aj}: \varphi_j < 0$). There is no need for a priori knowledge of the p-values distribution which are obtained through the stationary bootstrap resampling technique of Politis and Romano (1994).

The FDR approach holds for "weak dependence" conditions. As the number of tests increases to infinity, the dependence effects diminishing to zero due to asymptotics (Benjamini and Yekutiely, 2001; Storey, 2003; Storey, Taylor and Siegmund, 2004). Technical rules exhibit weak dependence within specific classes (i.e. moving averages), while each class is independent of each other (i.e. different families of rules) (Brajgowicz and Scaillet, 2012).

In financial applications, Barras, Scaillet and Wermers (2010) extend the work of Storey (2002) and Storey (2003) and introduce a method ($FDR^{+/-}$) that focus on both the significant and positive hypotheses. Barras, Scaillet and Wermers (2010) measure the proportion of false discoveries among mutual funds generating alphas, while trying to identify those displaying significant positive performance [4]. In our context, we are interested in identifying the technical rules that have a significant and positive performance (i.e. $\varphi_j > 0$). The FDR^+ is described as the expected value of the proportion of erroneous selections, F^+, over the significant and positive rules, R^+ , (i.e. $\frac{F^+}{R^+}$). The number of F^+ represents the rules, whose p-values falsely reject the true null (i.e. $H_{0j}: \varphi_j = 0$) in favour of the alternative and exist among R^+. On the other hand, R^+ portrays the number of rules rejecting the H_{0j}, in a two-tailed test, and their performance metric φ_j is positive. The estimate of FDR^+ is given by

[4] A critique on the FDR approach in mutual funds' performance is provided in Andrikogiannopoulou and Papakonstantinou (2019). The authors in a series of simulations demonstrate that FDR (under certain conditions) has low power in assessing the percentage of mutual funds' non-zero alphas. Barras, Scaillet and Wermers (2019) replied that with reasonable adjustments to the parameter choices of Andrikogiannopoulou and Papakonstantinou (2019), their $FDR^{+/-}$ approach retains its power. We also confirm the power of our $DFDR^{+/-}$ in our dataset through a series of Monte Carlo simulations in the Appendix. These results are available upon request.

$\widehat{FDR}^{+} = \hat{F}^{+}/_{\hat{R}^{+}}$ where $\hat{F}^{+}$ and $\hat{R}^{+}$ are the estimators of $F^{+}$and R^{+}, respectively[5]. For example, an FDR^{+} of 100% conveys that no trading strategy genuinely outperforms the benchmark, while any existing performance can be purely attributed to chance. In general, the FDR$^{+/-}$ produces a sensible trade-off between true positives and false selections. Its less conservative nature, the FDR method has the advantage of selecting outperforming rules even if the best rule is not significant in terms of performance.

We can estimate the frequency of false discoveries or the number of lucky rules, F^{+}, in the right tail of the distribution of performance metrics, φ_j at a given significance level γ as:

$$\hat{F}^{+} = \pi_0 * l * \gamma/2$$

where π_0 is the proportion of rules satisfying the null hypothesis, $\varphi_j = 0$, in the entire population, l is the number of the entire population and $\gamma/2$ is the probability of a positive non-genuine rule exhibiting luck due to symmetry conditions.

4.2 DFDR$^{+/-}$

Our DFDR$^{+/-}$ approach accounts for homogeneous discrete *p*-values, while provides different estimates of the proportion of false selections for outperforming and underperforming rules. Since our aim is to identify the significant outperforming rules only, we will focus on the estimation of DFDR^{+}. In our study, we utilize more than 21,000 technical trading rules over an IS horizon of two years (i.e. 504 observations). Performing a resampling procedure on each trading rule though, generates *p*-values which are discrete rather than continuous due to the finite number of bootstraps employed. Following Kulinskaya and Lewin (2009), we acquire discrete *p*-values which satisfy a uniform condition while sharing the same discrete support V. Let us consider as $N = \{n_1, \dots, n_{v+1}\}$ the number of occurrences of each element in V, i.e. $n_i = \#\{p_j = \gamma_i\}$ for $i = 1, \dots, s+1$ in order to express the empirical distribution of the computed *p*-values. Then, the empirical distribution of homogeneous discrete *p*-values with common support points can be explained by (V, N). Coming to the FDR approach

[5] Similarly, the $\hat{F}^{-}$ and $\hat{R}^{-}$ correspond to the estimators of the proportion of false discoveries (F^{-}) and alternative rules (R^{-}) generating negative performance (i.e., $\varphi_j < 0$).

calculation, the F^{+}, R^{+} and FDR^{+} also represent step functions with possible change points at the support points. Then it is sufficient simply to acquire their values at the specific support points to control the FDR[6]. Given that, the distribution function of the null discrete *p*-values on every support point is identical with that of continuous *p*-values, which is the key evidence in developing a parallel method for discrete *p*-values based on similar dynamic set-ups for continuous *p*-values.

In a two-sided test and for continuous true null *p*-values uniformly distributed, it holds $\Pr(p - \text{value} \leq \gamma) = \gamma$ for all $\gamma \in [0,1]$. On the other hand, for the discrete ones we observe only a certain number of support points for the *p*-values (i.e. $V = \{\gamma_1, \dots, \gamma_v, \gamma_{v+1}\}$ with $0 < \gamma_1 < \cdots < \gamma_v, < \gamma_{v+1} \equiv 1$) with potentially many ties, which satisfy a *discrete uniform condition* such that $\Pr(p - \text{value} \leq \gamma) = \gamma$, for $\gamma \in (0,1]$ and only for $\gamma \in V$. Using Politis and Romano (1994) bootstrap to compute the associated *p*-values for each rule, we end up with *p*-values satisfying the above condition. More specifically, every *p*-value is calculated by comparing the value of each performance metric with the value of its corresponding quantiles of bootstrapped metrics (Sullivan, Timmermann and White, 1999). This means that large values of observed test statistics provide evidence against the null and the corresponding *p*-values are given as $p_j = \frac{1}{B}\sum_{i=1}^{B}(\varphi_{jb} \geq \varphi_j)$, where B is the number of bootstrap replications, while φ_{jb} is the test statistic calculated for the bth bootstrap for the jth rule and φ_j is the realized test statistic. *P*-values computed this way are attached with support points in the form: $V = \left\{\frac{1}{B}, \frac{2}{B}, \dots, \frac{B-1}{B}, 1\right\}$, which also verify a discrete nature. Thus, providing an FDR framework, which takes into account larger discrete *p*-values (opposite to smaller, continuous ones) might help improve the existing FDR$^{+/-}$ method.

In order to minimize any undesired conservativeness of the estimators, any bias in the parametrization and make the algorithm more computationally efficient, we dynamically select λ under a stopping time rule. We define this stopping time condition as the point which holds $E[\widehat{\pi_0}(\gamma_q)] \geq \pi_0$, while q is the exact stopping time with respect to n_i (for $i = 0, \dots, s$), which is the history of *p*-values up to γ_q, when $q = i$. We also determine the whole procedure up to the stopping time q, as

[6] Suppose that γ is a time-running parameter from zero to one, then the continuous time processes F^{+}, R^{+} and FDR^{+} relax to discrete stochastic process on the support points.

$\{0 \equiv n_0, \dots, n_i\}$. Then we just set λ equal to γ_q. We check every support point instead of checking every single *p*-value for the stopping condition. If q is an appropriate stopping time, it must also hold $E[\widehat{FDR}(\gamma_q)] \geq FDR$, where $\widehat{FDR}$ is the estimation of the actual FDR provided by our method. The rationale for this approach is related to the idea of discovering the smallest support point, in which the number of appearances of *p*-values, n_i, to each right-hand side is almost equal. However, the stopping time condition is very general since we can construct numerous stopping time rules fulfilling the above criteria, while the actual right-hand side counts are unobservable, setting hurdles in the computation of the stopping time approach. It is already known though that employing a right boundary procedure, such as the one introduced by Liang and Nettleton (2012) for continuous *p*-values solves this issue by taking into account only the average of the remaining counts.

In general, the right-boundary specification guarantees conservative estimators for π_0 and FDR depending on a grid of candidate points for λ in line with data characteristics and a stopping time condition, at least for a continuous framework (Liang and Nettleton, 2012). We adopt the same procedure for discrete *p*-values. In addition to this, the right boundary procedure performs effectively for both independent and weakly dependent *p*-values, as observed in our case. Liang and Nettleton (2012) and Liang (2016) provide evidence of computing an FDR estimator using the right boundary procedure, while certain limits exist. Their results clearly satisfy a special case of the weak dependence condition of Storey, Taylor and Siegmund, (2004). The aim of the right boundary procedure is to find the first λ, at which the values of $\widehat{\pi_0}(\lambda)$ stop decreasing, satisfying in that way the stopping time condition. Poor selection of λ can cause unnecessary conservativeness in $\widehat{\pi_0}$ and $\widehat{FDR}$. For example, if λ lies in between two support points (e.g. $\lambda_i < \lambda < \lambda_{i+1}$)), then choosing $\widehat{\pi_0}(\lambda) > \widehat{\pi_0}(\lambda_i)$ can lead to an extra and unnecessary conservativeness in the estimation of proportion of rules with no abnormal performance (Harvey and Liu, 2018). Additionally, a small value of λ can lead to estimators with a large positive bias, while a high value of λ leaves only a small number of *p*-values on its right-hand side to estimate π_0, yielding an increase in the variance of estimators.

Previous literature follows a common approach to choosing λ under a continuous setup, visually examining the histogram of all *p*-values and setting the λ parameter equal to the support point above which the number of occurrences of p-values becomes fairly flat or selecting an arbitrary value. Brajgowicz and Scaillet (2012) explored three values, λ=0.4, 0.6 and 0.8. The authors found the value 0.6 to be optimal and they note that their results are not sensitive on the choice of λ and their exercise on the DJIA index. On the other hand, Harvey and Liu (2018) with a similar application that explores the same set of values for λ (0.4, 0.6 and 0.8), produce results that are sensitive to the choice of λ on Standard and Poor's Capital IQ database. On their database the optimal value of λ is 0.8. In our application where we study multiple series, we also note in our Monte Carlo simulations that $\widehat{\pi_0}(\lambda)$ is sensitive to some extent to the choice of λ, a finding that is contrary to previous evidence produced by Brajgowicz and Scaillet (2012) and in agreement with the simulations of Harvey and Liu (2018).

We consider a candidate set for λ, $\Lambda = \{\lambda_1, .., \lambda_n\}$, in which we place its components in ascending order, $0 \equiv \lambda_0 < \lambda_1 <\ldots< \lambda_n < \lambda_{n+1} \equiv 1$ (and $\lambda \subseteq \Lambda$). Then we select the best λ, as the minimum λ_q, which fulfils the condition $\widehat{\pi_0}(\lambda_i) \geq \widehat{\pi_0}(\lambda_{i-1})$, (i.e. $q = \min\{1 \leq i \leq n-1 : \widehat{\pi_0}(\lambda_i) \geq \widehat{\pi_0}(\lambda_{i-1})\}$). Specifically, we use the set Λ to separate the interval between zero and one, $(0,1]$, into $n+1$ bins with the i-th bin being $(\lambda_{i-1}, \lambda_i]$ for $i \in \{1, \ldots, n+1\}$ and $w_i = \#\{p_j \in (\lambda_{i-1}, \lambda_i]\}$ being the number of p-values in the i-th bin. Assuming equal intervals between λs, this approach practically chooses the right boundary of the first bin whose number of p-values is no larger than the average of the corresponding number to its right. In this way, we achieve the stopping condition when the downward trend of the number of p-values in each bin is neutralized, as we move forward, to a level where the random variants of rest p-values are fairly equal. Finally, acquiring the optimal λ in this way, we can easily calculate a conservative estimator for π_0 based on Storey's formula (2002) as has already been mentioned in previous sections.

The rest of the steps for the selection of outperforming rules remain similar to those of Barras, Scaillet and Wermers (2010) in the FDR specification. In terms of bootstrapping, we generate 1,000 sequence replications, and we retain the same bootstrap draw of the time series sample period for each trading rule's returns. In this way, we actually bootstrap the cross-section of trading rules returns through time in order to preserve the cross-sectional dependencies (Kosowski *et al.*, 2006; Fama and French, 2010). The application of stationary bootstrap also allows us to preserve the autocorrelations in returns structures. We then use the "point estimates" procedure of Storey, Taylor and Siegmund, (2004) on generated p-values, under weak dependence to select the outperforming rules, while setting a target for false discoveries. We can also extrapolate the proportion of trading rules displaying non-zero performance as $\pi_A = 1 - \pi_0$ in the entire universe of technical trading rules by using the FDR approach. This may be useful for an investor who wants to divide π_A into the proportions of positive, π_A^+, and negative, π_A^-, rules in the population[7]. The former includes both alternative rules generating positive performance and rejecting the null ($p - value < \gamma$) as well as those with positive performance but not rejecting the null ($p - value > \gamma$). The latter include those relevant for rules showing negative performance.

In our Monte Carlo simulation, also presented in Appendix A, we provide evidence that our discrete right boundary FDR procedure achieves a good trade-off between the bias and variance in various weakly dependent settings. We also compare the performance of the proposed procedure relative to the established FDR procedure of Storey, Taylor and Siegmund, (2004) as well as the StepM test (RW) of Romano and Wolf (2005).

4.3 DFDR$^{+/-}$ portfolio construction and decomposition of selected rules

We construct a portfolio of technical trading rules by setting the $\widehat{DFDR}^+$ target (the estimated DFDR for outperforming rules only) to 10%, which achieves a good trade-off between wrongly chosen rules and truly outperforming ones[8]. For our 10%-DFDR$^+$ portfolio, only 10% of the rules selected do not have genuine profitability among the outperforming rules, while 90% possess significant predictability. Moreover, we use the forecast averaging technique and allocate equal weight to the signals pooled from the chosen rules at each time-step in order to construct and calculate the portfolio returns. Since each trading rule might generate a long, short or neutral signal at a single time-step, we invest an equal

---

[7] For the estimation of π_A^+ and π_A^- we refer the reader to Barras, Scaillet and Wermers (2010).

[8] We experiment with $\widehat{DFDR}^+$ levels range from 5% to 30%. The series of Monte Carlo simulations indicate that our results are robust in this range.

proportion of our wealth in the signals and their corresponding returns generated by each individual rule, similar to calculating their equally-weighted cross-sectional mean.[9]

Following previous studies (see among others, Brock, Lakonishok and LeBaron, 1992; Bajgrowicz and Scaillet, 2012), a trading position is opened when a long or short signal is produced and liquidated when the signal is either reversed or neutralized. Should a neutral position be raised, the proportional wealth is assumed to be invested in the risk-free asset or the saving account. The gross daily return is calculated by the change in the closing value of the underlying index. A one-way transaction cost is deducted from the gross return when a position is terminated. The excess return is then estimated, to compare the profitability of the trading rules with the risk-free rate.

Table 2 presents the the percentage and standard deviations of the surviving rules identified by our 10%-DFDR$^{+/-}$ portfolios. The positions of all portfolios are re-adjusted on a monthly basis while the lookback period is two years[10].

[Table 2 here]

We note that the percentage of identified rules varies over the years and from market to market. The higher number of identified rules are found in the UK, Russia and the frontier markets indices. There is no obvious trend in the percentage of identified rules. There are peaks for the years 2006, 2009 and 2010 while the low is in year 2012. It is interesting to note that our procedure selects profitable rules for all indices and all years. A robustness exercise that presents the IS performance where the look-back period is set to one year is presented in Online Appendix C. These results confirm the analysis of this Section.

Table 3 reports a disaggregated analysis of the classes of trading rules that are profitable (i.e., relative strength indicators, filter rules, moving averages, support and resistance and channel breakouts) in every single market based. More specifically, in Table 3 we present the average percentages of the families of rules that are profitable across all years (i.e. 2006-2015) when the lookback period is set to two years and the portfolios are readjusted on a monthly basis.

[9] We have also tried a weighted average of the cross-section of signals, as portfolio construction technique, weighted according to the level of each rules' IS mean return. However, the OOS performance is similar to the equally-weighted portfolio.
[10] The relevant IS profitability of the portfolios is presented in Online Appendix B.1. In Online Appendix B.2. we present a break-even transaction cost analysis. These results are not presented in main text for the sake of space.

[Table 3 here]

It is obvious that moving averages dominate their counterparts across all markets. The second largest family of genuine profitable rules is that of support and resistance. We note that contrarian rules (relative strength indicators) present considerably lower percentages for all markets under study (with the notable exception of Jordan and UK). It seems that all markets under study are characterized by momentum and strong trends. Our results extend the findings of Jegadeesh (1990) and Jegadeesh and Titman (1993) who find profitability on momentum trading strategies on US and European stocks.

5. OOS Analysis

This Section provides a comprehensive analysis of the portfolios constructed on the surviving rules of the $DFDR^+$ procedure with an ex-ante approach where the look-back period is set to two years. Section 5.1 presents the excess profitability of the $DFDR^+$ portfolios over the OOS. Section 5.2 studies performance persistence by measuring the number of periods a $DFDR^+$ portfolio can generate a return above the risk-free rate. Finally, in Section 5.3 we investigate the role of financial stress on portfolios' performance.

5.1 Profitability

In this Section, we explore the profitability of the technical rules in the OOS, that we have identify as significantly profitable in the IS. We select the significantly positive rules based on their performance over the previous two years and under the 10%-$DFDR^+$ approach as a portfolio construction tool (see Section 4.3). Then the daily portfolios' performance is evaluated in the following one-month. We rebalance the daily positions of our $DFDR^{+/-}$ portfolio on a monthly basis and in a rolling-forward structure over a year. For example, after evaluating the performance of our portfolio in the first OOS period of one month, the remaining eleven OOS returns are calculated by rolling forward the IS by one month. We repeat the same procedure for every year separately in our sample (i.e. 2006-2015). In this way we dynamically build and evaluate our portfolios just as an active investor would in practice.

Table 4 reports the average excess annualized mean return and Sharpe ratio (in parenthesis)[11] for every index, using a two-year look-back period as IS and one month as OOS over the full sample of ten years after transaction costs. Both annualized mean excess returns and Sharpe ratios presented are calculated as the corresponding OOS averages of the twelve portfolios of significantly profitable outperforming rules built for every index after rolling forward the IS by one month during a year.

[Table 4 here]

OOS evidence shows that technical trading rules perform quite well for the majority of the markets considered during the earlier periods (i.e., 2006, 2007), with the performance of emerging markets being more profound. Focusing now on the following years and especially 2008 and 2009, almost all markets present extraordinary performances. Those years correspond to the global financial crisis period, during which most of the markets faced extreme downward trends and so severe losses. This environment seems beneficial for our technical trading rules' universe, as it consists mainly of momentum rules exploiting such big trends[12]. In 2010 the financial markets rebound but as our portfolios are constructed in the IS, where considerable different market conditions were prevailing, our portfolios are unable to exploit this effect. The limited profitability of our DFDR$^{+/-}$ portfolios continues to a certain extent for the following years, and then it recovers for the more recent periods. The level of momentum and reversal, the two processes that TA tries to capture, fluctuate and this is reflected to portfolios. We note that profitability over recent years (i.e. 2013-2015) is evident mostly in some advanced and emerging markets. When it comes to the comparison of the overall performance among markets, frontier markets provide a more solid investment option over the full sample period, generating positive mean returns and Sharpe ratios for the majority of the years, with second and third best being emerging and advanced markets respectively. Our general finding, though, is that the profitability of technical trading rules is variable across the years examined and TA can offer time-varying short-term profits.

---

[11] We observe similar trends (time-varying profitability) with the Sortino ratio and the manipulation ratio. These results are available upon request. The results presented in Table 4 are the averages of 12 consecutive experiments for each year. This procedure and the fact that our OOS is short (even in the 3- and 6-month cases presented in the online Appendix D) validates the discussion that follows, irrespective of the performance metric.

[12] Griffin, Ji and Martin (2003) state that momentum profits around the world are economically large and statistically reliable in both good and bad economic states.

In Online Appendices D and E, an extensive robustness exercise is presented where the look-back period is set to 1 year and the OOS period is set to 3 and 6 months. The findings of this exercise reveal similar trends but a lower profitability when the OOS is increased from 1 month to 3 and 6 months. These findings can explain parts of the past literature in the field of TA and their profitability. Previous researchers have tended to examine long periods (see among others, Sullivan, Timmermann and White, 1999; Bajgrowicz and Scaillet, 2012; Hsu, Taylor and Wang, 2016) and report considerably lower TA performance. Our empirical exercise also highlights, that the profitability of specific technical rules is short-term and varies across periods.

5.2 Performance Persistence

In the previous Section, we focused on identifying profitable technical rules with the $DFDR^{+/-}$ and examining the trading performance of the generated portfolios. As discussed, it seems that technical rules and our approach are able to generate short-term profits. It will be interesting to see how fast this profitability decays and whether there are differences between different markets and periods. This element is overlooked in the related literature, where the empirical evaluation is static and limited to specific periods. In real-world trading environments, practitioners are adaptive and rebalance their portfolios on a frequent basis. This exercise will provide an insight for the first time, how adaptive portfolio managers need to be with TA.

Table 5 presents the persistence of our generated monthly- rebalanced portfolios over the OOS. We measure persistence as the consecutive OOS months for which our portfolios have a trading performance above the relevant risk-free rate.

[Table 5 here]

We note that in the vast majority of cases, traders need to rebalance portfolios on more than a monthly basis. These results are expected, considering the trading performance of our trading rules as outlined in the previous Section. However, it should be noted that there are a few cases where the portfolios might have negative profitability in the first month of the OOS, but bounce back in the following periods (see for example, persistence for 2008 in Tables 5 and the one reported for the three- and six-month case in Tables D.3 and D.4 in online Appendix D[13]). The market with the higher persistence is the US, the largest and most liquid index under study, with China to closely follow. On the other hand, the markets with the lowest performance persistence are Russia and Japan. Emerging markets seems to offer no "*safe haven*" for static portfolios. Interestingly, in the cases where persistence is high, adaptiveness does not always lead to increased profits, while patience is rewarded. Finally, we apply a cross-validation experiment following Harvey and Liu (2015) to make our OOS analysis more robust. Our results show that there are technical rules profitable in the OOS that are also significant in both the IS and OOS. These findings, presented in online Appendix G for the sake of space, further augment the main message of this study, that technical rules are genuinely valuable in trading.

5.3 Financial Stress

In the previous Sections, we noted a peak in the performance of technical rules for the years 2008 and 2009 that correspond to the recent global financial crisis. The performance of our portfolios deteriorates in the following years, but there are still cases where excess profitability after transaction costs is present even in advanced markets. Our results contradict the previous recent literature that finds no recent excess profitability of TA after transaction costs (see among others, Hsu, Hsu and Kuan, 2010; Bajgrowicz and Scaillet, 2012 and Taylor, 2014)[14]. These authors argue that the popularity of exchange-traded funds, algorithmic trading, market liquidity, derivatives or the effect of other macroeconomic factors have eliminated excess profitability in recent years. Although the effect of these factors cannot be rejected, our results guide us to explore the effect of financial stress on our portfolios.

In order to explore the effect of financial stress on our portfolios, we need first to consider how to measure it. The related literature is rich in financial stress indices. The difference between them is based

---

[13] We further validate our results by exploring the same OOS exercises but using only one year as IS. These findings are presented in online Appendix E. For the case of US, we have investigated the potential drivers of the OOS performance of $DFDR^{+}$ index-based portfolios by running monthly time series regressions of each returns series on momentum (Mom), short-term reversal (STRev) and long-term reversal (LTRev) factors over the 2006-2015 period. The corresponding coefficients are not significant and available upon request.

[14] As discussed before, none of the previous literature presents such an extensive empirical application as the one presented here. Previous studies also applied more conservative and time-consuming MHT frameworks and limited their empirical applications to specific years or periods. For example, let us consider the case of the MSCI US index, the 2-year IS and 1-month OOS. If our application was limited only to 2014 or 2015, our interpretation would be anti-diametrical. The flexibility and adaptiveness of $DFDR^{+/-}$ along with recent developments in computational power allowed us to conduct an empirical analysis that unveils previously unknown patterns.

on the components that are used to construct them, their frequency and the market to which they are applied. In our study, we apply the Office of Financial Research (OFR) stress indices. They are constructed based on 33 market financial variables and they are available for the US, other advanced economies and emerging markets[15]. We match our findings for the US, advanced and emerging markets indices from Section 5.1 with their corresponding stress levels as reported by the related US, other advanced[16] and emerging markets OFR stress indices. In table 6 below we present the performance of our portfolios under high and low financial market stress[17].

[Table 6 here]

We note that the trading performance of technical rules is on average considerably better when financial stress is high in emerging and other advanced markets. The annualized returns are five to six times higher in the high than the low period. This finding is consistent with the evidence found in Smith *et al*. (2016) where the authors demonstrate that TA is relatively more profitable in high-sentiment periods. During such periods, financial markets exhibit stronger trends which can be captured by momentum based technical trading rules. The OFR stress index is based on series of indicators that capture financial market sentiment.

For US, we observe the opposite trend. The average profitability of technical rules in the US over the 10-year period of study is positive when stress levels are low, and negative when stress levels are high. One possible explanation of these results can be the decomposition of the technical rules families that consist our $DFDR^{+}$ portfolios. We note from Table 3, where we report the percentage decomposition of the technical trading rules, that the RSI's percentage (i.e., the mean reverting technical rules family in our TA pool) is considerably lower compared to the ones of the other financial markets under study. Mean reverting rules can capture the turning points in the underlying trends (that,

[15] Other indices (such as the St. Louis Fed Financial Stress Index and the Kansas City Financial Stress Index) focus on a specific index while others stop before our sample (such as the International Monetary Fund stress index). Our criteria to select the index are to cover as many markets as possible from those under study and to ensure that all indices have been constructed with the same methodology. For a list of the OFR financial stress index sources see, https://www.financialresearch.gov/financial-stress-index/files/indicators/index.html.

[16] We note that the MSCI and the OFR indices do not correspond to each other perfectly. For example, the "other advanced" OFR index does not include the US. However, amongst the highest cited financial stress indices, OFR is the closest to our study.

[17] The OFR financial stress indices have a daily frequency and are volatile in certain periods. We examine only the 1-month OOS case as we are interested in measuring the effect of previous stress levels at the highest possible frequency.

occur in periods of high financial stress) and assist the performance of our portfolios. In the US portfolios where mean reverting rules are to some extent absent, profitability considerably deteriorates in high financial stress periods.

6. Conclusions

Previous research on TA's performance on equity markets is scarcely focused on certain features such as, study of a single, usually advanced, market index; a capped number and families of technical trading rules with no decomposition of their genuine profitable families; a non-realistic exercise of technical trading, which ignores transaction costs. In addition, other important questions, such as the length of the in-sample IS/OOS horizon for exercising TA; the effect of financial stress on its performance; a detailed estimation of the persistence of TA's profitability; the importance of the frequency a portfolio of technical trading rules is rebalanced have rarely been answered or not answered at all. In finance, researchers and practitioners are aware of the past behaviour of the financial markets and how macroeconomic events have affected their series. The real OOS is the live trading experience. A technical rule might be profitable either by chance or by overfitting and data mining in the OOS.

In this study we exercise a large universe of 21,195 technical rules on twelve MSCI advanced, emerging and frontier indices covering a period from 2004 to 2015. Towards this direction, we examine the IS and OOS profitability, its persistence and the role of financial stress. We also explore the role of the size of the IS and the OOS on TA's performance. In addition, we extend the work of Barras, Scaillet and Wermers (2010) and introduce $DFDR^{+/-}$ which adopts a large-scale homogeneous discrete *p*-values framework while dynamically performing the estimation of its parameters. $DFDR^{+/-}$ is an adaptive computationally and efficient approach for data snooping testing that can assist academics and investors dealing with large data sets and a high number of competing models. A Monte Carlo simulation proves its accuracy and demonstrates its superiority in terms of power relative to a FDR benchmark and the most common MHT framework in finance, the RW test. Compared to the $FDR^{+/-}$ of Barras, Scaillet and Wermers (2010), $DFDR^{+/-}$ has the same power when the tuning parameter of the $FDR^{+/-}$ is correctly set and is more computationally efficient.

The DFDR$^{+/-}$ is applied in more than 240,000 hypotheses and identifies subsets of the technical rules that are genuine and profitable in the IS and the OOS. Among the families that outperform in all markets are mostly those of moving averages, channel breakouts and support and resistance, which belong to "momentum" classes of technical trading. The profitability and persistence of TA vary in the OOS between the indices and the years, but we observe a peak for 2008 and 2009. In the following period, the profitability of TA diminishes, only to recover again in the most recent period. Concerning the performance persistence of TA, it appears to be higher on the US index. Finally, we also note that for the US market TA gains value when financial stress levels are low, while we get the opposite picture for the emerging and other advanced markets. Our robustness exercises, provided as an online Appendix, reveal that the performance of TA is negatively affected, when the IS is decreased to one year or the OOS is increased to three or six months. They also present that, when a cross-validation exercise is employed, a small number of technical rules are significantly profitable in both the IS and the IS-OOS periods.

Our results demonstrate that TA has value and can generate short-term profits even in advanced markets. Its profitability varies over the years and indices, while it seems to be a factor of several parameters. Our empirical exercise reveals four factors, the underlying market conditions, the level of financial market development, the level of financial stress and the choice of the IS and OOS. Our results also explain to some extent the contradicting findings of the previous literature. Previous research in the field tries to identify whether genuine technical rules exist for long periods (Brock, Lakonishok and LeBaron, 1992; Hsu, Hsu and Kuan, 2010; Hsu, Taylor and Wang, 2016). Our study discovers that TA has short persistence. The fact that a technical rule is not profitable for extended periods does not mean that TA has no value. Similarly, other research focuses only on individual indices and/or specific years for OOS testing (Brock, Lakonishok and LeBaron, 1992; Sullivan, Timmermann and White, 1999; Bajgrowicz and Scaillet, 2012). We reveal that TA profitability varies over the years and the equity markets investigated. Focusing on a specific index or year can lead to misleading results, as the effectiveness of technical rules appears to be highly volatile.

References

Andrikogiannopoulou, A. and Papakonstantinou, F. 2019. Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? *Journal of Finance*, 74 (5): 2667-2688.

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): 64-74.

Asness, C.S., Moskowitz, T.J. and Pedersen, L.H. 2013. Value and Momentum Everywhere. The *Journal of Finance* 68(3): 929-985.

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

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

Barras, L., Scaillet, O., Wermers, R. 2019. Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply. *Journal of Finance, Replications and Corrigenda* 74 (5): 2667-2688.

Batten, J.A., Lucey, B.M., McGroarty, F., Peat, M. and Urquhart, A. 2018. Does intraday technical trading have predictive power in precious metal markets? *Journal of International Financial Markets, Institutions and Money* 52: 102-113.

Bena, J., Ferreira, M.A., Matos, P. and Pires, P. 2017. Are foreign investors locusts? The long-term effects of foreign institutional ownership. *Journal of Financial Economics* 126(1): 122-146.

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): 289-300.

Benjamini, Y. and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4): 1165-1188.

Brock, W., Lakonishok, J. and LeBaron, B. 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *Journal of Finance* 47(5): 1731-1764.

Cesari, R. and Cremonini, D., 2003. Benchmarking, portfolio insurance and technical analysis: a Monte Carlo comparison of dynamic strategies of asset allocation. *Journal of Economic Dynamics and Control* 27(6): 987-1011.

Dai, Z., Zhu, H. and Kang, J., 2021. New technical indicators and stock returns predictability. *International Review of Economics & Finance*, 71:127-142.

DeBondt, W. F. M. and Thaler, R. 1985. Does the stock market overreact? *Journal of Finance* 40(3): 783–805.

Eurex, 2018. Contract Specifications for Futures Contracts and Options Contracts at Eurex Deutschland and Eurex Zürich.

Fama, E.F. and French, K.R 2010. Luck Versus Skill in the Cross-Section of Mutual Fund Returns. *Journal of Finance* 65 (5): 1915–1947.

Fung, W. and Hsieh, D.A. 2001. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies* 14(2): 313–341.

Goldbaum, D., 2003. Profitable technical trading rules as a source of price instability. *Quantitative Finance*, 3: 220-229.

Griffin, J.M., Ji, X. and Martin, J.S., 2003. Momentum investing and business cycle risk: Evidence from pole to pole. *The Journal of Finance*, 58(6): 2515-2547.

Han, Y., Zhou., G. and Zhu, Y. 2016. A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics* 122(2): 352-375.

Harvey, C.R. and Liu, Y. 2015. Backtesting. *The Journal of Portfolio Management* 42(1): 13-28.

Harvey, C.R. and Liu, Y. 2018. Detecting repeatable performance. *The Review of Financial Studies* 31(7): 2499-2552.

Hatgioannides, J. and Mesomeris, S., 2007. On the returns generating process and the profitability of trading rules in emerging capital markets. *Journal of International Money and Finance*, 26(6): 948-973.

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

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

Jamali, I. and Yamani, E. 2019. Out-of-sample exchange rate predictability in emerging markets: Fundamentals versus technical analysis. *Journal of International Financial Markets, Institutions and Money* 61: 241-263.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *Journal of Finance* 45(3): 881–898.

Jegadeesh, N. and Titman, S. 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* 48(1): 65–91.

Kosowski, R., Timmermann, A.,Wermers, R. and White, H. 2006. Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis. *Journal of Finance* 61(6): 2551–2595.

Kulinskaya, E. and Lewin, A. 2009. On fuzzy familywise error rate and false discovery rate procedures for discrete distributions. *Biometrika* 96(1): 201-211.

Liang, K. 2016. False discovery rate estimation for large-scale homogeneous discrete p-values. *Biometrics* 72(2): 639-648.

Liang, K. and Nettleton, D. 2012. Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 74(1): 163-182.

Marshall, B.R., Nguyen, N.H. and Visaltanachoti, N., 2013. Liquidity measurement in frontier markets. *Journal of International Financial Markets, Institutions and Money*, 27:1-12.

McKenzie, M.D., 2007. Technical trading rules in emerging markets and the 1997 Asian currency crises. *Emerging Markets Finance and Trade*, 43(4): 46-73.

MSCI 2013. Deploying Multi-Factor Index Allocations in Institutional Portfolios. [Online] https://www.msci.com/documents/1296102/1336482/Deploying_Multi_Factor_Index_Allocations_in_Institutional_Portfolios.pdf/857d431b-d289-47ac-a644-b2ed70cbfd59

Neely, C.J., Rapach, D.E., Tu, J. and Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. *Management science*, 60(7): 1772-1791.

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

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

Smith, M.D., Wang, N., Wang, Y., Zychowicz, J.E. 2016. Sentiment and the effectiveness of technical analysis: Evidence from the hedge fund industry. *Journal of Financial and Quantitative Analysis* 5: 1991–2013.

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

Storey, J.D. 2003. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics* 31(6): 2013-2035.

Storey, J.D., 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 (Statistical Methodology)* 66(1): 187-205.

Sullivan, R., Timmermann, A. and White, H. 1999. Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *Journal of Finance* 54(5): 1647–1691.

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

Zaremba, A., 2020. Performance persistence in anomaly returns: Evidence from frontier markets. *Emerging Markets Finance and Trade*, 56(12): 2852-2873.

Zhu, Y. and Zhou, G. 2009. Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics* 92(3): 519-544.

Appendix A. Monte Carlo simulations

In this appendix, we present supporting evidence of the finite sample performance of the DFDR$^{+/-}$ test using a Monte Carlo experiment. Our main goal is the exploration of the empirical level and power of the test in accurately estimating the proportions of outperforming, underperforming and neutral trading rules. Even though we mainly focus on the FDR rate and its power on the rejection frequency of rules with significant returns (either positive or negative), we also compare it with the power of the FDR procedure presented in Storey, Taylor and Siegmund, (2004) and the RW test of Romano and Wolf (2005) using the familywise error rate (FWER) as selection criterion instead of the FDR, which by definition is a more conservative approach.

Before we start running our Monte Carlo simulation, we need to ensure that our experiment correctly embodies the empirical properties of the technical trading strategies employed, such as their time series and cross-sectional dependencies (see also Barras, Scaillet and Wermers, 2010, Hsu, Hsu and Kuan, 2010; Bajgrowicz and Scaillet, 2012). We have previously demonstrated that our technical trading rules are fully characterized by a weak form of dependence, and this holds especially for those belonging in the same family (e.g. moving averages). This is the main property, and we need to take that into consideration when constructing our experiment. In this way, we can also examine whether our DFDR$^{+/-}$ does indeed have a good response to weak dependence conditions. In order to work in this direction we simultaneously resample matrices of $\mathscr{b} \times l$ returns, where $\mathscr{b}$ the random block size is consecutive time series observations ($\bar{\mathscr{b}} = 10$) under the stationary bootstrap and $l = 21{,}195$ denotes the trading rules universe as in the empirical exercise. This approach also allows us to preserve the cross-sectional dependencies among the strategies of the same class, while we also preserve autocorrelation every time we apply the same bootstrap replication to all trading rules. For the Monte Carlo experiment, we randomly select a 155-day sample (i.e. seven months) from July 1, 2013 to February 1, 2014, to simulate our trajectories and we generate the 155-day trajectories for the $l = 21{,}195$ trading rules as in the empirical exercise. In particular, we employ the stationary bootstrap to create every realized trajectory similar to calculating the *p*-values of the empirical study. We generate 1,000 bootstrap replications of returns, where each replication has similar statistical properties. In order to obtain the true power of the DFDR$^{+/-}$ test in selecting the proportions of outperforming, underperforming and neutral rules, we need to control these proportions beforehand. We can then compare them with their corresponding estimations based on the DFDR$^{+/-}$. We adjust 20% of the simulated strategies to outperform the benchmark, 50% to deliver "neutral" returns with no significant performance and 30% to underperform the benchmark during the simulation process. The selected outperforming (underperforming) strategies consist only of a group of neighbouring rules, ranked in terms of highest (lowest) returns in our empirical sample. In this way we avoid having in our groups rules with slightly different parameters, which at the same time possibly belong to both outperforming and underperforming classes.

In terms of the specific procedure followed, we achieve the control of outperforming, "neutral" and underperforming rules by re-centring the generated returns of each trading rule with its own mean and we utilize that across all five families of rules. This actually leads to all trajectories having almost zero-mean properties while retaining their corresponding, unique standard deviations. We then shift the paths of the outperforming and underperforming rules by some positive and negative value respectively, while keeping each rule's corresponding standard deviation the same. Such a parallel transition does not, however, affect the empirical properties of the paths, other than the mean. The notion is to construct the trajectories of different strategies in such a way as to exactly acquire the same, positive Sharpe ratio for all outperforming rules and the same negative Sharpe ratio for all underperforming rules.

As for the target Sharpe ratios employed for shifting the paths of outperforming and underperforming strategies, we follow the study of Bajgrowicz and Scaillet (2012) and select Sharpe ratios closely related to those obtained in our empirical exercise. In particular, we set three specific targets of positive Sharpe ratios for our outperforming rules, i.e., 2, 3, 4; and three specific targets of negative Sharpe ratios for our underperforming rules, i.e., -2, -3, -4. All of them correspond to annualized Sharpe ratios, just like those calculated from the daily returns of each strategy. We then consider pairs of the Sharpe ratios above in order to adjust the outperforming and underperforming rules, while shifting their trajectories towards the target. Take the (2, -2) pair for example. We design 20% of the rules to yield an equal Sharpe ratio of 2 (i.e. outperforming) and likewise all 30% of the rules share an equal Sharpe ratio of -2 (i.e. underperforming). The remaining 50% of our rules' universe show zero performance. This results in nine possible combinations of positive and negative Sharpe ratio pairs representing fixed alternative hypotheses against the null of a Sharpe ratio being equal to zero. The above levels seem to match our historical sample results since we obtain positive annualized Sharpe ratios up to 4 for the best-performing strategies and negative annualized Sharpe ratios down to -4 for the worst-performing ones.

Focusing on the estimation power of the DFDR$^{+/-}$ approach, Table A.1 presents the estimates for the proportions of outperforming ($\widehat{\pi_A^+}$), underperformming ($\widehat{\pi_A^-}$) and neutral ($\widehat{\pi_0}$) strategies under the Sharpe ratio metric and for the nine possible Sharpe ratio pairs. It also reports the success of the estimators in tracking the actual proportions of outperforming ($\pi_A^+ = 20\%$), underperforming ($\pi_A^- = 30\%$), and neutral ($\pi_0 = 50\%$) trading rules. Once again, we apply the "point estimates method" of Storey, Taylor and Siegmund, (2004) to the DFDR $^{+/-}$ test to obtain the estimators of these proportions based on the Monte Carlo results. This time we keep the cut-off threshold fixed to $\gamma^* = 0.4$, as at this point $\widehat{\pi_A^+}$ and $\widehat{\pi_A^-}$ become constant. In other words, as γ increases up to an adequate enough value, $\widehat{\pi_A^+}$ and $\widehat{\pi_A^-}$ include both genuine and false selections of trading rules representing the total number of outperforming and underperforming rules respectively.

[Table A.1 here]

Our DFDR$^{+/-}$ approach seems to provide quite robust estimators for the outperforming, underperforming and neutral proportions of technical trading rules, with only small deviations from their true corresponding levels. For instance, looking at the (3, -3) Sharpe ratios pair, the estimator for the outperforming rules (i.e. $\widehat{\pi_A^+}$), is 15.23%, the relevant estimator for underperforming rules (i.e. $\widehat{\pi_A^-}$) is 27.68% and that for neutral rules (i.e. $\widehat{\pi_0}$) is 57.09%, which are quite close to their true levels of 20%, 30% and 50% respectively. This clearly highlights the power of our method in accurately identifying the true proportions of outperforming, underperforming and neutral rules in the entire population.

Finally, we present in Table A.2 the performance of constructed portfolios of outperforming rules under the DFDR$^+$ approach based on the Monte Carlo simulation and for each of the nine Sharpe ratio combinations. We control the DFDR$^+$ at a prespecified level similar to our empirical exercise. For instance, we build two different types of DFDR$^+$ portfolios by setting the targets of erroneous selections at 10% and 20% respectively. In terms of performance and power, the table reports the actual false discovery rate achieved (FDR$^+$) in comparison with its fixed level adjusted in advance (i.e. 10% and 20% respectively), the proportions of genuinely best-performing rules over the total number of outperforming rules denoted as "power", and the absolute number of genuinely best-performing trading rules as "portfolio size"[18]. To reflect the contribution of the proposed method, we compare our results

[18] We compute the actual false discovery rate (FDR$^+$) by replacing the actual proportion of neutral trading rules (i.e., $\pi_0 = 50\%$) instead of the estimated one (i.e., $\widehat{\pi_0}$) in $FDR^+ = \frac{\pi_0 * l * \gamma/2}{\#\{p_k \leq \gamma,\ \varphi_k > 0; k=1,\dots,l\}}$).

with the FDR procedure of Storey, Taylor and Siegmund, (2004), while controlling the FDR at the target levels of 10% and 20% respectively. We also benchmark our procedure with RW tests at the 5% and 20% levels.

[Table A.2 here]

The findings of Table A.2 reveal that the $DFDR^{+/-}$ approach is superior in terms of finite sample power to the FDR of Storey, Taylor and Siegmund, (2004) and the RW test. Specifically, the $DFDR^{+/-}$ reports robust power in rules selection and portfolio size, while it closely tracks the actual false discovery rate across all conditions and Sharpe ratio pairs. For example, consider again the (3, -3) Sharpe ratios pair, where the 10%-$DFDR^{+}$ portfolio efficiently converges to its FDR rate at 8% and successfully discovers on average 64.74% of the best-performing rules. The corresponding 10%-FDR portfolio discovers only 30.27% of the best-performing rules on average, while it meets its target rate only at 7.3%. When it comes to the size of portfolios, the 10%-$DFDR^{+}$ outstandingly outperforms the 10%-FDR approach by sufficiently selecting 3,048 rules, while the latter method detects only 1907. Increasing the target rate of the FDR to 20% does not affect the patterns since it improves the power of selection to 39.86% while the portfolio size is not affected (1,907). The 20%-$DFDR^{+}$ though, performs even better by detecting on average 66.3% of the outperforming rules, and forms a portfolio of 3,322 trading rules. In terms of target rate, the 20%-$DFDR^{+}$ portfolio falls below 20% and achieves an FDR^{+} of 10.59%. Asymptotic theory is the most possible reason for this outcome, but the 20%-$DFDR^{+}$ portfolio is still able to successfully deal with data snooping bias as seen above. The RW test clearly underperforms the FDR procedures based on all metrics retained. Overall, our Monte Carlo experiments undoubtedly reveal that the $DFDR^{+/-}$ method has greater power when compared with conservative FDR methods, such as the Storey, Taylor and Siegmund, (2004) procedure.

Tables

Table 1. Summary statistics of the daily return series under study (12 MSCI indices and the federal funds rate).

Market	Mean (bp)	Max (%)	Min (%)	Std. dev. (%)	Kurtosis	Skewness	First AC (significance)
Advanced	1.55	9.10	-7.33	1.02	12.86	-0.50	0.12 (*)
US	1.45	11.04	-9.51	1.18	15.11	-0.36	-0.10 (*)
UK	1.60	17.32	-36.26	1.29	212.22	-6.62	0.01
Japan	2.66	12.77	-20.75	1.27	62.62	-2.12	-0.07
Emerging	1.91	10.07	-9.99	1.27	11.38	-0.49	0.22 (*)
Russia	1.83	42.37	-58.10	2.35	172.93	-2.26	0.02
China	3.32	14.05	-12.84	1.74	10.10	-0.04	0.03 (*)
Brazil	3.59	37.69	-46.23	2.19	109.54	-0.39	0.02
Frontier	1.95	12.54	-9.32	1.62	8.74	0.20	0.06 (*)
Estonia	2.96	5.50	-7.70	1.06	6.47	-0.13	0.16 (*)
Morocco	0.99	5.69	-9.07	0.83	15.56	-1.38	0.26 (*)
Jordan	1.21	7.82	-9.08	1.10	13.03	-0.71	0.07 (*)
Federal funds rate	0.53	0.02	0.00	0.01	2.78	1.18	1.00 (*)

The mean daily returns are reported in basis points (bp). Maximum, minimum and standard deviation are presented in percentages (%). The last column reports the first-order autocorrelation coefficients. Coefficients notated with (*) are significant at 1% (*) level for the Ljung-Box Q statistic. The study period for all time series is 01/01/2004 to 31/12/2016.

Table 2. Percentage and standard deviation of the DFDR$^{+/-}$ procedure survivors (IS 2 years).

Market	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Average
Advanced	0.34	1.16	0.43	2.39	2.84	3.52	0.24	3.50	16.24	0.33	**3.10**
	(0.97)	(1.87)	(1.03)	(1.78)	(1.56)	(7.03)	(0.06)	(7.02)	(11.37)	(0.15)	**(3.28)**
US	0.01	0.02	0.16	0.91	1.54	8.33	0.28	7.75	31.49	19.58	**7.01**
	(0.00)	(0.01)	(0.32)	(1.12)	(0.94)	(9.49)	(0.07)	(10.55)	(3.55)	(13.81)	**(3.99)**
UK	20.08	15.52	0.32	12.97	20.89	9.41	0.18	0.15	8.72	0.29	**8.85**
	(9.08)	(10.26)	(0.38)	(10.21)	(6.96)	(9.88)	(0.04)	(0.06)	(11.71)	(0.09)	**(5.87)**
Japan	3.59	0.11	0.69	3.20	2.58	1.09	0.29	0.27	0.18	0.16	**1.22**
	(6.56)	(0.04)	(1.49)	(1.32)	(0.29)	(1.40)	(0.05)	(0.07)	(0.03)	(0.04)	**(1.13)**
Emerging	2.87	1.26	1.12	3.35	6.04	9.01	0.36	0.39	0.19	0.24	**2.48**
	(5.39)	(1.19)	(1.46)	(0.97)	(4.46)	(10.21)	(0.09)	(0.10)	(0.08)	(0.12)	**(2.41)**
Russia	14.39	16.22	1.77	25.64	17.61	10.90	0.28	0.44	0.58	0.90	**8.87**
	(7.39)	(9.55)	(3.48)	(7.84)	(4.78)	(10.65)	(0.10)	(0.11)	(0.14)	(0.30)	**(4.43)**
China	2.59	32.32	5.14	0.93	3.72	3.59	0.28	0.27	0.20	0.81	**4.99**
	(3.38)	(15.03)	(9.68)	(0.56)	(6.77)	(7.16)	(0.01)	(0.07)	(0.09)	(0.52)	**(4.33)**
Brazil	22.52	8.62	8.84	17.79	20.84	8.71	0.09	0.27	1.02	0.32	**8.90**
	(13.6)	(7.28)	(8.44)	(5.10)	(5.19)	(9.74)	(0.07)	(0.35)	(0.38)	(0.14)	**(5.03)**
Frontier	14.14	1.10	3.85	29.22	25.26	7.24	0.37	0.47	4.39	1.49	**8.75**
	(12.40)	(0.45)	(7.50)	(7.27)	(9.00)	(9.64)	(0.19)	(0.23)	(7.36)	(1.26)	**(5.53)**
Estonia	17.37	0.35	4.34	7.91	7.52	10.23	0.19	4.35	8.86	0.66	**6.18**
	(16.94)	(0.71)	(7.12)	(2.73)	(4.13)	(10.14)	(0.05)	(7.18)	(6.51)	(1.26)	**(5.68)**
Morocco	7.36	26.96	17.24	4.82	0.64	0.15	0.34	0.65	0.15	0.22	**5.85**
	(8.27)	(8.76)	(9.58)	(2.90)	(0.63)	(0.06)	(0.34)	(0.62)	(0.05)	(0.10)	**(3.13)**
Jordan	20.26	1.57	1.77	4.27	1.52	0.67	0.21	0.09	0.11	0.18	**3.06**
	(11.4)	(2.22)	(1.67)	(1.34)	(0.62)	(0.84)	(0.05)	(0.03)	(0.03)	(0.07)	**(1.83)**
Average	**10.46**	**8.77**	**3.81**	**9.45**	**9.25**	**6.07**	**0.26**	**1.55**	**6.01**	**2.10**	5.77
	(7.95)	**(4.78)**	**(4.35)**	**(3.60)**	**(3.78)**	**(7.19)**	**(0.09)**	**(2.20)**	**(3.44)**	**(1.49)**	(3.89)

This table reports the percentage and standard deviations of the survivor rules adjusted by the total number of rules. For example, in 2006 for the advanced market, the average number of surviving rules is 72 (0.0034*21195) and their standard deviation is 206 (0.0097*21195). The average is estimated from the twelve portfolios whose OOS is in 2006. The first portfolio's IS runs from 01/01/2004-31/12/2005 and the remaining eleven are calculated by rolling forward the IS by one month.

Table 3. Percentage decomposition of the DFDR$^{+/-}$ procedure survivors in classes of technical trading rules (IS 2 Years)

Market	Average of RSI%	Average of FR%	Average of MA%	Average of SR%	Average of CB%
Advanced	4.97%	8.54%	65.78%	7.30%	13.40%
US	0.85%	10.84%	61.46%	10.07%	16.78%
UK	10.09%	11.32%	60.81%	5.79%	12.00%
Japan	8.31%	8.93%	39.10%	22.42%	21.25%
Emerging	6.07%	21.97%	42.98%	13.98%	15.00%
Russia	3.57%	18.28%	48.54%	19.32%	10.29%
China	3.69%	8.71%	63.61%	16.77%	7.22%
Brazil	7.09%	10.96%	56.41%	13.31%	12.23%
Frontier	3.19%	16.79%	52.88%	14.23%	12.90%
Estonia	1.27%	7.65%	69.34%	12.77%	8.96%
Morocco	9.81%	6.12%	59.34%	18.38%	6.35%
Jordan	16.93%	9.26%	54.13%	14.15%	5.54%
Grand Total	**6.32%**	**11.61%**	**56.20%**	**14.04%**	**11.83%**

This table reports the decomposition (in percentage terms) of the DFDR$^{+/-}$ procedure survivors in every single family of technical trading rules (Filter Rules (FRs), Moving Average rules (MAs), Support-Resistance rules (SRs) and Channel Breakout rules (CBs) and Relative Strength Indicators (RSIs)) as an average across all years examined. For example, for the advanced market, 4.97% of the survivors belongs to RSI rules, 8.54% to filter rules, 65.78% to moving averages, 7.3% to support and resistance and 13.4% to channel breakouts on average for the period 2006-2015. For every year, the IS portfolio runs for two years and the remaining eleven are calculated by rolling forward the IS by one month.

Table 4. Annualized Returns and Sharpe Ratios after Transaction Costs (IS 2 Years and OOS 1 Month)

Market	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Average
Advanced	-3.66%	-4.72%	**10.63%**	**4.14%**	-17.79%	-1.92%	-3.23%	**0.35%**	-2.07%	-1.43%	-1.97%
	(-0.54)	(-0.43)	**(0.41)**	**(0.25)**	(-2.91)	(-0.50)	(-1.27)	**(0.13)**	(-0.42)	(-0.64)	(-0.59)
US	**1.81%**	-6.16%	**22.72%**	-3.64%	-14.07%	-1.65%	-1.37%	**7.02%**	**5.73%**	-5.28%	0.51%
	(0.20)	(-0.45)	**(0.77)**	(-0.19)	(-1.49)	(-0.31)	(-0.50)	**(1.31)**	**(0.77)**	(-0.71)	(-0.06)
UK	**15.94%**	-3.14%	**16.5%**	**19.58%**	-12.73%	-4.12%	-3.83%	**1.67%**	-7.09%	-2.80%	2.00%
	(1.55)	(-0.27)	**(0.41)**	**(1.16)**	(-1.47)	(-0.58)	(-3.70)	**(0.67)**	(-0.97)	(-0.45)	(-0.36)
Japan	-11.33%	-8.68%	**41.53%**	**11.47%**	-2.11%	-7.87%	-3.30%	-3.95%	-5.75%	-2.98%	0.70%
	(-0.76)	(-1.80)	**(1.40)**	**(0.51)**	(-0.37)	(-1.62)	(-1.31)	(-1.03)	(-2.27)	(-1.13)	(-0.84)
Emerging	**1.75%**	-1.93%	**43.52%**	**4.25%**	-7.22%	-3.14%	-0.93%	-1.71%	-3.71%	-1.16%	2.97%
	(0.15)	(-0.15)	**(1.10)**	**(0.29)**	(-1.09)	(-0.52)	(-0.23)	(-0.55)	(-1.28)	(-0.31)	(-0.26)
Russia	**45.77%**	-9.00%	**45.64%**	**13.09%**	-20.64%	-9.90%	-2.42%	-4.63%	**6.92%**	-12.48%	5.24%
	(1.25)	(-1.33)	**(1.01)**	**(0.47)**	(-2.10)	(-1.02)	(-0.44)	(-0.92)	**(0.44)**	(-0.98)	(-0.36)
China	**59.59%**	**29.16%**	-2.10%	-19.48%	-8.93%	-10.76%	-0.29%	-3.54%	-4.51%	**11.46%**	5.06%
	(2.62)	**(1.17)**	(-0.05)	(-1.04)	(-1.09)	(-1.40)	(-0.64)	(-1.42)	(-0.66)	**(0.82)**	(-0.17)
Brazil	**7.53%**	**67.85%**	**48.86%**	**15.4%**	-15.81%	-5.12%	-7.98%	-3.03%	-7.35%	-1.07%	9.93%
	(0.28)	**(1.20)**	**(0.90)**	**(0.68)**	(-1.67)	(-0.73)	(-2.42)	(-0.49)	(-0.60)	(-0.07)	(-0.29)
Frontier	-11.64%	**14.24%**	**64.67%**	**11.12%**	**0.26%**	-12.4%	-0.27%	**4.20%**	**7.75%**	**2.33%**	**8.03%**
	(-2.04)	**(1.44)**	**(2.24)**	**(1.04)**	**(0.06)**	(-3.55)	(-0.09)	**(0.98)**	**(1.39)**	**(0.32)**	**(0.18)**
Estonia	-4.93%	-7.76%	**65.08%**	**2.83%**	**6.06%**	-24.12%	**6.26%**	-4.29%	**12.92%**	-13.23%	3.88%
	(-0.62)	(-0.45)	**(1.46)**	**(0.10)**	**(0.30)**	(-2.60)	**(0.98)**	(-0.69)	**(1.23)**	(-2.11)	(-0.24)
Morocco	**34.00%**	**21.97%**	**25.32%**	**1.50%**	-10.22%	-2.49%	**1.19%**	-4.90%	-0.64%	-0.36%	**6.54%**
	(1.87)	**(1.65)**	**(1.20)**	**(0.10)**	(-2.17)	(-0.87)	**(0.10)**	(-0.48)	(-0.61)	(-0.17)	**(0.06)**
Jordan	-0.62%	-3.47%	**37.89%**	-9.46%	-0.12%	-0.79%	-4.59%	-2.27%	-4.63%	-1.32%	1.06%
	(-0.04)	(-0.43)	**(1.32)**	(-0.81)	(-0.02)	(-0.13)	(-1.55)	(-1.03)	(-1.23)	(-0.37)	(-0.43)
Average	**11.18%**	**7.36%**	**35.02%**	**4.23%**	-8.61%	-7.02%	-1.73%	-1.26%	-0.20%	-2.36%	3.66%
	(0.33)	**(0.01)**	**(1.02)**	**(0.21)**	(-1.17)	(-1.15)	(-0.92)	(-0.29)	(-0.35)	(-0.48)	(-0.28)

This table reports the average OOS annualized returns and Sharpe ratios of twelve portfolios for two years of IS and one month of OOS after transaction costs (rolling forward by one month). For example, the -3.66% annualized return of the advanced markets (2006) is calculated as the average OOS annualized return of twelve portfolios. The first portfolio's OOS return is calculated over January 2006 using as IS the period 01/01//2004-31/12/2005. The remaining eleven OOS returns are calculated by rolling forward the IS by one month. The same logic applies for the Sharpe ratios. The last column and row present the average performance per market across all years and per year respectively.

Table 5. Monthly Performance Persistence for IS 2 Years (1 month rolling forward)

Market	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Average
Advanced	1.00	0.58	0.75	0.83	0.42	0.75	0.50	0.67	0.83	0.75	**0.71**
US	1.17	0.42	1.08	1.08	1.33	1.17	1.00	1.92	0.83	0.25	**1.03**
UK	0.92	0.58	0.50	1.08	0.42	0.58	0.08	0.75	0.42	1.00	**0.63**
Japan	0.17	0.50	0.58	0.92	1.58	0.58	0.25	0.17	0.08	0.42	**0.53**
Emerging	0.50	0.92	1.17	1.33	0.33	0.75	0.58	0.17	0.17	0.83	**0.68**
Russia	0.42	0.25	0.17	0.33	0.33	0.67	0.67	0.33	0.92	0.33	**0.44**
China	2.17	2.92	0.42	0.58	0.42	0.42	0.83	0.42	0.67	1.08	**0.99**
Brazil	0.67	0.58	0.67	0.75	0.33	0.42	0.33	0.58	0.92	0.50	**0.58**
Frontier	0.42	1.67	2.25	0.50	0.58	0.17	0.75	0.67	1.33	1.25	**0.96**
Estonia	0.50	0.42	0.75	0.58	0.75	0.17	0.67	0.42	1.08	0.25	**0.56**
Morocco	1.92	2.00	1.25	0.42	0.25	0.33	0.58	0.92	0.42	0.58	**0.87**
Jordan	0.42	0.75	2.08	0.50	0.83	0.83	0.25	0.75	0.58	0.58	**0.76**
Average	**0.85**	**0.97**	**0.97**	**0.74**	**0.63**	**0.57**	**0.54**	**0.65**	**0.69**	**0.65**	**0.73**

This table reports the average number of consecutive months that the monthly OOS returns of the twelve portfolio returns are above the risk-free rate. This average is calculated by generating the monthly OOS in consecutive months for each of the twelve portfolios mentioned in Table 4. For example, in advanced markets for the first portfolio, we calculate the OOS returns for 2006 (January, February, etc.). If the OOS returns over the first month are below the relevant risk-free rate, we assign a value of 0. If the OOS returns remain above the risk-free rate during the first month e.g. in January but not for February, we assign the value of 1. Otherwise, we assign a value of 2 or more. This process is repeated for the remaining eleven portfolios of the year. The analysis is done using a maximum 18 months of OOS calculations for each portfolio. The last column and row present the average monthly performance persistence per market across all years and per year respectively.

Table 6. Financial Stress Performance

Market	Period	Financial Stress	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Average
US	**IS of 2 Years - OOS 1 Month**	**High**	-	-9.96% (-0.59)	22.72% (0.77)	-3.64% (-0.19)	-18.27% (-2.18)	-0.57% (-0.17)	-5.08% (-2.16)	-	-	-3.86% (-0.75)	-2.67% (-0.75)
		Low	1.81% (0.20)	-3.36% (-0.32)	-	-	-0.28% (-0.02)	-2.42% (-0.39)	0.53% (0.18)	7.02% (1.31)	5.73% (0.77)	-5.41% (-0.71)	0.45% (0.13)
Other Advanced		**High**	-	-0.72% (-0.06)	10.63% (0.41)	4.14% (0.25)	-16.24% (-2.88)	-2.15% (-0.55)	-4.86% (-1.81)	-	-	-0.37% (-0.07)	-1.37% (-0.59)
		Low	-3.66% (-0.54)	-6.02% (-0.57)	-	-	-33.23% (-3.59)	0.61% (0.16)	0.11% (0.05)	0.35% (0.13)	-2.07% (-0.42)	-1.53% (-0.85)	-5.68% (-0.70)
Emerging		**High**	-	-13.42% (-4.45)	43.52% (1.10)	1.78% (0.11)	-20.32% (-3.13)	-4.89% (-0.69)	3.00% (0.74)	26.12% (4.45)	-4.83% (-4.01)	-1.48% (-0.37)	3.28% (-0.69)
		Low	1.75% (0.15)	-0.82% (-0.06)	-	7.8% (0.70)	-4.37% (-0.66)	-1.87% (-0.36)	-3.65% (-0.94)	-3.94% (-1.45)	-3.61% (-1.21)	0.48% (0.29)	-0.91% (-0.39)

This table reports the average OOS annualized returns and Sharpe ratios of the portfolios generated in Section 5.1. High and low correspond to high and low financial stress conditions as reported by the OFR stress indices. "–" indicates that for this year and market there were no periods with high (or low) financial stress.

Table.A.1 . Estimation of neutral, positive, and negative proportions by the DFDR$^{+/-}$ procedure versus the actual ones.

Outperforming SR	Proportion	Underperforming SR -2	-3	-4
2	$\pi_0 = 50\%$	73.08	62.93	60.97
	$\pi_A^+ = 20\%$	9.39	11.91	11.17
	$\pi_A^- = 30\%$	17.53	25.17	27.86
3	$\pi_0 = 50\%$	66.43	57.09	53.44
	$\pi_A^+ = 20\%$	14.35	15.23	16.09
	$\pi_A^- = 30\%$	19.22	27.68	30.47
4	$\pi_0 = 50\%$	64.35	53.55	50.21
	$\pi_A^+ = 20\%$	16.28	18.08	18.89
	$\pi_A^- = 30\%$	19.37	28.36	30.91

The quantities presented correspond to the average values estimated over 1000 Monte Carlo simulations. The proportion of rules that are neutrally performing (π_0), outperforming (π_A^+) and underperforming (π_A^-) are set to 50%, 20% and 30% respectively. The table provides the estimates when the annualized Sharpe ratio for out- and under-performing rules is set to 2, 3, 4 and -2, -3, -4 respectively.

Table A.2. True false discovery rate, accuracy and the positive-performing portfolio size through different methods.

Outperforming SR	Portfolio Type	Underperforming SR								
		-2			-3			-4		
		FDR$^+$	Power	Portfolio size	FDR$^+$	Power	Portfolio size	FDR$^+$	Power	Portfolio size
2	10%-DFDR$^+$	12.92	39.46	1995.71	12.67	42.01	2118.86	12.05	40.31	2014.43
	10%-FDR	16.36	11.75	896.75	14.14	11.98	941.47	14.61	11.91	883.09
	20%-DFDR$^+$	13.79	39.92	2110.78	15.03	43.3	2369.41	14.65	41.66	2281.92
	20%-FDR	16.12	16.67	896.75	13.95	17.69	941.47	14.63	16.15	883.09
	5%-RW	0.86	0.01	0.53	0.90	0.01	0.52	0.71	0.01	0.48
	20%-RW	8.42	0.05	3.03	8.28	0.06	3.33	7.11	0.05	2.90
3	10%-DFDR $^+$	8.44	64.74	3076.41	8.00	64.74	3048.45	8.78	62.89	3039.27
	10%-FDR	8.29	29.61	1856.71	7.30	30.27	1906.89	7.59	31.46	1945.02
	20%-DFDR$^+$	9.54	65.29	3212.88	10.59	66.30	3321.58	11.55	64.62	3343.04
	20%-FDR	8.89	38.26	1856.71	7.90	39.86	1906.89	8.25	40.11	1945.02
	5%-RW	0.31	0.01	0.52	0.04	0.01	0.60	0.08	0.01	0.57
	20%-RW	3.21	0.07	3.47	2.30	0.07	3.41	2.97	0.07	3.49
4	10%-DFDR$^+$	6.45	82.39	3790.07	6.62	83.57	3879.56	7.83	83.01	3945.84
	10%-FDR	4.85	56.8	3187.43	4.56	54.55	3012.02	4.50	55.99	3138.75
	20%-DFDR$^+$	7.80	83.35	3948.29	9.53	85.49	4194.29	10.5	84.89	4244.14
	20%-FDR	6.63	68.06	3187.43	5.60	65.59	3012.02	5.96	67.98	3138.75
	5%-RW	0.00	0.02	0.65	0.01	0.02	0.73	0.00	0.02	0.68
	20%-RW	0.53	0.09	3.87	0.31	0.10	4.38	0.34	0.09	3.73

This table reports the FDR$^+$ and accuracy in percentages and the portfolio size (out of 21195). Accuracy is estimated by the ratio of actual outperformers discovered by the underlying procedure. We consider confidence target levels of 10% and 20% for the DFDR$^+$ and benchmark it against the FDR procedure in Storey, Taylor and Siegmund, (2004) for the same target levels. The second benchmark is the FWER procedure by RW tested at the levels 5% and 20%. The quantities refer to average values over 1000 Monte Carlo simulations for different combinations pairs, when annualized Sharpe ratio for out- and under-performing rules is set to 2, 3, 4 and -2, -3, -4 respectively.