‘MHT-hacking’: Implementing Multiple Hypothesis Tests in FX markets

Minyou Fan[a], Fearghal Kearney[a,*], Youwei Li[b], Jiadong Liu[a]

[a] *Queen’s Management School, Queen’s University Belfast, UK*
[b] *Hull University Business School, University of Hull, UK*

Abstract

This study presents an important guide for navigating the growing number of multiple hypothesis testing techniques proposed in the literature. It also serves to inform an investor of how best to implement such methods when seeking to identify truly profitable trading strategies. We employ over one million technical trading rules across 48 popular currencies to evaluate the test power of leading multiple hypothesis testing approaches in the FX market. After comparing a comprehensive range of algorithms and control approaches we find that controlling the false discovery proportion with Hsu et al. (2014)’s superior predictive ability test yields the most desirable performance.

Keywords: Multiple hypothesis testing; Big data; Data snooping bias; False discovery; Foreign exchange; Trading rule selection.

JEL: C12, C52, C58, F31

*Corresponding author. Queen’s Management School, Queen’s University Belfast, Riddel Hall, 185 Stranmillis Road, BT9 5EE, UK. email: f.kearney@qub.ac.uk.

1. Introduction

Market participants conduct many performance appraisals across several trading strategies to determine those that are profitable. However, when evaluating a large number of trading rules using Single Hypothesis Testing (SHT), some can appear to significantly outperform benchmark strategies, even though they are not truly superior. These false positives are known as Type I errors and are important for investors because they may lead to large losses in practice and can increase opportunity costs.[1] More specifically, assuming fixed funds, having invested in an incorrectly identified strategy, a financial practitioner will miss the opportunity to invest in other trading rules that would have ultimately led to significant profits. For these reasons, it is vitally important that researchers and professionals implement Multiple Hypothesis Testing (MHT) when a statistical analysis involves simultaneous tests. MHT can effectively manage Type I errors since it adjusts the critical value of rejections taking into account confidence levels for the whole family of simultaneous tests conducted. The benefits of adopting MHT have been proven by previous studies, for example, Chordia et al. (2020), who assess over 2 million stock signals, finding that 62% to 95% of strategies uncovered by SHT are not deemed to be truly significant when using MHT controls.

However, due to the plethora of frameworks proposed in the literature, investors are now faced with a paucity of guidance on what MHT approach to select and how to appropriately implement it. Extensively used MHT approaches include, the Bonferroni correction of Bonferroni (1936), the Holm correction of Holm (1979), the false discovery rate controlling methods of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001), the bootstrap reality check of White (2000), the superior predictive ability of Hansen (2005), the Stepwise and generalised reality check of Romano et al. (2007) and Romano et al. (2010), the Stepwise superior predictive ability check of Hsu et al. (2010), the model confidence set of Hansen et al. (2011), the generalised superior predictive ability check of Hsu et al. (2014), and the double-bootstrap approach of Harvey and Liu (2020). While investors naturally do not wish to leave profitable opportunities unexploited, they likely have an even greater aversion to following strategies that ultimately lead to losses. An MHT approach is less likely to yield false rejections than following SHT, but a strict MHT approach may uncover

[1] Type II errors refer to the failure to discover truly outperforming profits. When these errors occur, investors miss potential opportunities but do not suffer from the increased costs of investment or potentially large losses. Therefore, Type II errors are less likely to be the focus of market participants.

no significant rejections even when truly significant outperforming trading rules exist in reality. An overly conservative MHT approach is meaningless for investors if it fails to identify any profitable strategies. For this reason a delicate balance is required for investors seeking the most appropriate MHT approach for their context. Most, previous empirical studies rely on one or two MHT approaches, for example, Sullivan et al. (1999), Hsu and Kuan (2005), Qi and Wu (2006), Neuhierl and Schlusche (2011), Hsu et al. (2018), and Hudson and Urquhart (2021). As a result, another potential risk is that with different MHT implementations resulting in diverging conclusions, investors could, either deliberately or inadvertently, game MHT methods by selectively applying particular frameworks or parameter sets that align with their viewpoint. We term this phenomenon '*MHT-hacking*', building on the term p-hacking, which highlights the need to move from SHT to MHT (Harvey, 2017, Chen, 2021).

Our study aims to assess the test power of different MHT approaches in the FX market and provide guidance on which MHT approach FX practitioners should apply. We achieve this by conducting a comprehensive comparison of MHT approaches and identifying how FX investors should best implement them. We focus on the FX market for a number of reasons. First, FX is the world's largest financial market, with over $1.5 trillion traded each day. Second, the FX market trades 24 hours a day, thus generating more observations and data to sift through than many other markets. Third, the FX market is a more challenging market in which to identify profitable opportunities in comparison with equity or commodity markets, with Moskowitz et al. (2012), Asness et al. (2013), Kim et al. (2016), and Papailias et al. (2021) finding that a given trading rule commonly yields higher profits in commodity or equity markets as opposed to the FX market. Fourth, seeking a desirable MHT approach could be more crucial for FX agents in comparison to investors in other asset classes. The majority of FX investors are professional institutions that have more financial knowledge than retail investors (Sager and Taylor, 2006, Menkhoff and Taylor, 2007). Their sophistication may allow them to more easily uncover potential trading signals based on economically justifiable narratives, which might in turn lead to a greater need to apply MHT given the additional number of trading opportunities to sift through. Finally, unlike stock market participants, who can use thousands of stocks to construct a diversified portfolio, FX market participants have fewer opportunities to cover the losses of applying an inappropriate MHT approach given that the limited number of correlated currencies make it more difficult to build a comprehensively diversified portfolio.

We investigate algorithms across three broad categories, namely those controlling familywise error rates (FWER), false discovery proportions (FDP), and false discovery rates (FDR). These approaches are used to evaluate a universe of 21,195 technical trading rules (TTR) across 48 currencies from 1976 to 2019. Previous studies assess the test power of MHT when evaluating portfolio performance (Hsu et al., 2018), and fund managers' skills (Hsu et al., 2010, Kearney et al., 2014). However, portfolio managers have the opportunity to diversify the risks of their investment via dynamic weighting of assets in the portfolio (Markowitz, 1952), whereas this can not be done when trading a single asset, further increasing the need to adopt the correct MHT framework. Overall, an MHT approach that uncovers a greater number of significant observations has stronger test power than an MHT approach that identifies fewer instances (Hansen, 2005). However, an MHT approach that finds too many instances of significance is too tolerant to efficiently control Type I errors (Chordia et al., 2020). Thus, the ideal MHT approach should balance the control of false rejections and the ability to uncover profitable opportunities. In other words, the most appropriate algorithm should sit between the strictest MHT approaches, which minimise Type I errors but may fail to uncover profitable rules, and SHT, which does not control for false rejections.

Consistent with with previous literature, for example, Hsu et al. (2014), we evaluate the test power of a given MHT approach by counting how many outperforming trading rules it uncovers. Our empirical findings highlight the importance of selecting the correct MHT approach in FX markets for two reasons. First, we find that the test power varies greatly across MHT approaches. For instance, when focusing on excess returns, the average number of significant outperforming rules uncovered by FWER is only 51, compared to 216 and 705 for FDP and FDR, respectively. Second, when focusing on individual currency pairs, different MHT approaches lead to contradictory conclusions. For example, when adopting FDR approaches, the Belgian Franc is deemed to be an attractive currency as more than 2,000 trading rules generate significant profits. However, by contrast, controlling for FWER leads to zero significant outperforming TTRs. This extreme variation in test powers across MHT approaches highlights the importance of establishing MHT implementation guidance for FX market participants.

Our empirical results indicate that an MHT approach that controls FDP outperforms those that control

FWER and FDR. After comparing eight MHT approaches, we find that the tests based on the Superior Predictive Ability (SPA) framework proposed by Hansen (2005), can efficiently manage Type I errors as well as, or even better than, the Reality Check (RC)-based tests.[2] Therefore, we conclude that the test power of the SPA test controlling for the FDP (FDP-SPA) proposed by Hsu et al. (2014) is superior to other MHT approaches. This differs from the findings of Chordia et al. (2020), for equity markets where they conclude that the FDP control algorithm of Romano et al. (2007) based on the bootstrap RC of White (2000) (FDP-RC) is the most desirable MHT approach. Hansen (2005) highlights that the test power of SPA-based approaches is superior to the test power of RC-based algorithms, when a model sample contains many poor-performing strategies, as is often the case when implementing technical trading rules in FX markets.[3]

We also uncover additional empirical evidence for proposing the use of SPA for FX practitioners, as we find that the RC-based tests are sensitive to the specific performance metrics adopted, whereas the SPA-based tests are not. Four popular performance metrics are implemented to measure the performance of trading rules, including the mean and Sharpe ratio of both daily excess total and Market Timing (MT) returns. The MT component is the total return minus the mean of the prior periods' returns (Hsu et al., 2016)), and is used to indicate the skill inherent in a trading strategy. We find that the SPA test can efficiently manage the Type I errors regardless of the performance metric employed, whereas the RC tests cannot. For instance, when trading the Belgian Franc, the FDP-RC approach fails to uncover any significant mean return trading rules, whereas it uncovers over 4,000 seemingly profitable Sharpe ratio based trading rules. This is in contrast with the FDP-SPA approach, which also identifies over 4,000 Belgian Franc profitable trading rules but does so regardless of the performance metric employed. An MHT approach that is sensitive to performance appraisals is undesirable, as it leads to FX investors receiving conflicting results when focusing on different performance metrics.

[2]Based on the SPA of Hansen (2005), Hsu et al. (2010) and Hsu et al. (2014) developed the stepwise SPA and FDP-SPA tests to control the FWER and FDP, respectively. The SPA-based approaches are extensively used by recent investment papers, such as Hsu et al. (2016), Hsu et al. (2018), Gu and Shen (2018), Psaradellis et al. (2019), and Xu and Wang (2021).

[3]More specifically, the RC test depends on the sample-independent null distribution, where the Bonferroni bound test (at level α) rejects the null if $p_m in < \alpha/m$, with $p_m in$ being the smallest p-value of m hypotheses (Bonferroni, 1936). Many poor-performing trading rules added to the model sample will increase m and result in a tiny critical value α/m. In contrast, the sample dependent null distribution of the SPA allows us to employ a smaller denominator $m^\star < m$ and achieve a higher critical p-value, $\alpha/m^\star$.

Our study's contribution to the literature is threefold. First, we extend the comparison across various MHT approaches to the currency market, and provide guidance to investors operating in the FX market on how to put in place such safeguards. Second, we go beyond existing studies, e.g., Chordia et al. (2020) and Harvey and Liu (2020), by incorporating SPA-based tests into the comparison and by arguing that the FDP-SPA approach of Hsu et al. (2014) has better test power than the FDP-RC when seeking to uncover outperforming trading rules. Third, the comparison across various approaches demonstrates the importance of controlling for data snooping bias in the FX market, with some MHT frameworks, that is, the non-bootstrap ones, being unable to do so effectively.

The remainder of this paper is organised as follows. We present an overview of the various MHT approaches in Section 2. Next, we detail our data universe and return calculations in Section 3. Section 4 shows the construction and performance of technical trading rules. We summarise the assessment of various MHT approaches in Section 5 with Section 6 presenting our out-of-sample evaluation. Finally, we conclude our study in Section 7. All supplementary results can be found in the Appendices.

2. Multiple hypothesis testing

Academics and professionals generally apply SHT to assess whether a trading rule produces statistically significant outperformance. We define $f_{k,t}$ ($k = 1, 2, \ldots, m$ and $t = 1, 2, \ldots, n$) as a series with n observations for the k-th considered trading rule over the sample period. φ_k refers to the performance metric of the k-th trading rule. The null hypothesis for k is:

$$H_0 : \quad \varphi_k \leq 0. \tag{1}$$

This procedure may not generate bias if the total number of tested models, m, is small. When m is large, applying SHT may falsely reject the null. The number of significant outperforming trading rules is assumed to be m_0 across m different trading rules. R is an observed variable equal to the number of trading rules rejected by SHT. Table 1 illustrates the results of this comparison procedure. V represents the number of hypotheses that are falsely rejected (i.e., Type I errors), with T representing the number of hypotheses that should be rejected but are not (i.e., Type II errors). U and S represent the number of correct rejections when the null is or is not true, respectively. Therefore, V/R refers to the realised False Discovery Proportion

(FDP). Conversely, $T/(m-R)$ is the False Non-Discovery Proportion (FNDP).

Table 1: Hypothesis testing outcomes based on m tests.

Reality	Test Conclusion		Total
	Non-significant	Significant	
True	U	V	m_0
Not true	T	S	$m-m_0$
	$m-R$	R	m

This table reports all possible results after testing m null hypothesis without controlling for error rates. R is an observed variable equal to the number of trading rules rejected by SHT. Table 1 illustrates the results of this comparison procedure. V represents the number of hypotheses that are falsely rejected (i.e., Type I errors), with T representing the number of hypotheses that should be rejected but are not (i.e., Type II errors). U and S represent the number of correct rejections when the null is or is not true, respectively. U, V, T, S are unknown variables after completing the multiple comparison procedure. R is a known variable as it is shown in the test results.

When the number of assessed trading rules is small, testing strategies one by one may not raise significant issues. However, thousands of trading rules are available for market practitioners to consider, meaning that simply using SHT can cause Type I errors that detrimentally impact investors due to biased information. Investors should be able to derive benefits from applying MHT as it manages the number of false rejections. Nevertheless, an MHT that minimises the number of false rejections but fails to uncover outperforming rules is meaningless for investors as it fails to satisfy the prior aim, uncovering profitable trading rules. For instance, when m is large, the FWER control approach may not find any significant rejections in a specific currency as it allows only one false rejection at most. Therefore, similar to Chordia et al. (2020), we propose the principle that a desirable MHT approach in the FX market should sit in the middle between the most rigid tests, such as the Bonferroni approach, and tests without controls, such as the SHT.

Consistent with existing literature, for example, Hsu et al. (2016) and Harvey and Liu (2020), our study considers the main statistical definitions controlled by MHT: the Familywise Error Rate (FWER), the False Discovery Proportion (FDP), and the False Discovery Rate (FDR). Unlike Chordia et al. (2020) and Harvey et al. (2020), we also include in our comparison two widely used MHT approaches based on Hansen (2005)'s SPA test. We therefore assess eight MHT frameworks in total and detail the procedure for each control approach below. The eight approaches are further detailed in Table 2. Consistent with previous studies, for example, Hsu and Kuan (2005) and Hsu et al. (2016), we set the benchmark as a risk-free portfolio in which the portfolio excess returns remain zero over all time horizons.

Table 2: MHT Approaches

Family	Approach	Abbreviation	Source
FWER	Bonferroni	Bonferroni	Bonferroni (1936)
	Holm	Holm	Holm (1979)
	Stepwise Boostrap Reality Check	Step-RC	Romano and Wolf (2005)
	Stepwise Superior Predictive Ability	Step-SPA	Hsu et al. (2010)
FDP	FDP-Boostrap Reality Check	FDP-RC	Romano et al. (2007)
	FDP-Superior Predictive Ability	FDP-SPA	Hsu et al. (2014)
FDR	Benjamini and Hochberg	BH	Benjamini and Hochberg (1995)
	Benjamini and Yekutieli	BY	Benjamini and Yekutieli (2001)

This table presents the MHT approaches used in our comparison. Family refers to the control targets: Familywise Error Rate (FWER), the False Discovery Proportion (FDP), and the False Discovery Rate (FDR). Source lists the original paper where each approach was proposed.

2.1. Controlling the familywise error rate

The basic control approach is to manage the FWER, which is the probability of rejecting at least one true null hypothesis. This approach, which is designed to avoid any false discoveries, is given as:

$$FWER = P(V \geq 1). \tag{2}$$

Controlling the FWER is the strictest MHT approach as it minimises the number of false discoveries to one at most, that is, the number of Type I errors is equal to or less than one (Hochberg and Tamhane, 1987). We implement four FWER control approaches: (1) the single-step approach (Bonferroni) (Bonferroni, 1936); (2) a stepwise procedure (Holm) (Holm, 1979); (3) the stepwise RC (Step-RC) (Romano and Wolf, 2005); and (4) the stepwise SPA (Step-SPA) (Hsu et al., 2010).

2.1.1. Bonferroni approach

The Bonferroni approach is a single-step procedure in which all tested statistics are compared to a single adjusted critical value (Bonferroni, 1936). If each hypothesis is tested at a given significance level α, and the probability of erroneous rejection is $\alpha^\star$, the expected number of false rejections $E(V)$ is equal to $m \times \alpha^\star$. Thus, this method rejects a specific hypothesis k if the p-value (p_k) is equal to or less than the adjusted critical value based on the total number of hypotheses, $\alpha^\star = \alpha/m$. This method is widely used due to its simplicity and inspires the FWER control methods described below.

However, the Bonferroni approach determines the critical values based exclusively on the total number of tests, and, hence, is too conservative to uncover all true significant rejections. The critical value of this method does not incorporate the cross-correlations across various trading rules that exist in most financial applications. According to this mechanism, a large m leads to a very small critical value for many financial applications, and only hypotheses that yield extremely small p-values can be rejected.

2.1.2. Holm approach

Holm (1979) developed a step-down procedure to control the FWER. This procedure first tests the most significant hypothesis, that is, the lowest p-value (highest t-statistic), and works down to the least significant one. The null hypothesis is rejected at a significance level α if $p_k \leq \alpha/(m-k+1)$ for $k = 1, ..., m$, where p_1 refers to the highest p-value and p_m represents the lowest value. The core idea is that if, when testing the hypothesis H_k, hypotheses $H_1, H_2, ..., H_{k-1}$ are not erroneously rejected, the researcher can apply the Bonferroni approach to hypothesis H_k. Holm's criterion is equivalent to that of Bonferroni but becomes more and more lenient as k increases. Therefore, the Holm approach typically rejects more null hypotheses than the Bonferroni approach, but is still regarded as relatively conservative as it ignores the dependence structure of the individual p-values.

2.1.3. Bootstrap reality check and superior predictive ability

White (2000) proposed the bootstrap RC test to determine statistical performance using an empirical distribution of the best model from a universe of models. The test power of the RC test is adversely impacted by poor-performing models. To solve this issue, Hansen (2005) developed a more powerful method known as the Superior Predictive Ability (SPA) test. Both algorithms adjust the critical values based on the bootstrap distribution.

The null hypothesis of the RC and SPA tests is that none of the involved m trading rules shows significant outperformance. This is given as follows:

$$H_0: \quad \max_{k=1,\ldots,m} \varphi_k \leq 0. \tag{3}$$

The alternative hypothesis, therefore, is that at least one of the m trading rules is significantly superior to the benchmark. Assuming a trading rule with zero excess return as the benchmark, and adopting mean excess returns as a standard performance metric, we have $\varphi_k = E(f_k)$. Therefore, the test shown in Equation 3 is naturally based on the maximum of the normalised sample mean of $f_{k,t}$, as:

$$RC_n = \max_{k=1,\dots,m} \sqrt{n}\bar{f}_k. \tag{4}$$

where $\bar{f}_k = \frac{\sum_{t-1}^{n} f_{k,t}}{n}$. Next, the RC test employs the stationary bootstrapping procedure of Politis and Romano (1994) to build a distribution of bootstrapped performance metrics. We define $f_k^*(b)$ $(b = 1, 2, \dots, B)$ as the b-th time resampled series of model k with a pre-specified block length Q, and with $\bar{f}_k^*(b)$ representing the performance of trading rule k at the b-th bootstrap. The bootstrapped normalised maximum performance, $RC_n^\star$, is:

$$RC_n^\star = \max_{k=1,\dots,m} \sqrt{n}(\bar{f}_k^*(b) - \bar{f}_k). \tag{5}$$

Then, the adjusted p-value is measured by comparing RC_n to the quantiles of the distribution of bootstrapped performance metrics, $RC_n^\star$. We reject the null hypothesis when the p-value is below the pre-determined significance level, α. Notably, White (2000) used the Least Favourable Configuration (LFC), in which the expected return of the k-th trading rule is equal to zero, that is, $E(\bar{f}_k) = 0$. When a large number of poor-performing trading rules ($E(\bar{f}_k) < 0$) are included in the model sample, the adjusted critical value will be minimal, and very few null hypotheses are rejected.

To avoid issues associated with the LFC, Hansen (2005) introduced the SPA test to minimise the impacts of poor-performing trading rules. First, in the SPA test, the normalised maximum performance of Equation 4 is converted to:

$$SPA_n = \max(\max_{k=1,\dots,m}(\sqrt{n}\bar{f}_k, 0)). \tag{6}$$

By re-centering the distribution of bootstrapped performance metrics, the SPA test no longer requires the LFC. For the k-th rule, we denote $\bar{Z_{k,t}}^\star(b)$ as the performance measure of the re-centered returns of the b-th bootstrapped series:

$$Z_{k,t}^\star(b) = f_{k,t}^*(b) - \bar{f}_k + \bar{f}_k I(\bar{f}_k \le A_k), \tag{7}$$

where $A_k = -\hat{\sigma_k}\sqrt{2loglog(n)}$, and $\hat{\sigma_k}$ refers to the standard deviation of $\sqrt{n}\bar{f}_k$. Therefore, the bootstrapped normalised maximum performance of the SPA, $SPA^{\star}_n$, is:

$$SPA^{\star}_n = \quad \max(\max_{k=1,\ldots,m} \sqrt{n}(\bar{Z}^{\star}_{k,t}(b) - \bar{f}_k), 0). \tag{8}$$

Next, the adjusted critical value of the SPA test is measured through the same procedure as the RC test.

2.1.4. Stepwise approaches

A downside of both the RC and SPA tests is that they are not able to count the total number of significant outperforming strategies. Since the null hypothesis of RC/SPA is that none of the investigated models outperforms a benchmark, rejecting the null hypothesis only implies that the best-performing model significantly beats the benchmark. To summarise the total number of significant rejections, Romano and Wolf (2005) introduced a stepwise procedure that repeats the RC in constrained sub-samples to identify as many outperforming models as possible when controlling the FWER. This approach is useful for assessing numerous trading strategies as investors are concerned with which trading rule significantly outperforms the benchmark, that is, a risk-free portfolio or a market index portfolio, rather than whether a single best-performing rule is significantly superior to the benchmark.

Much previous literature, for example, Chordia et al. (2020) and Harvey et al. (2020), focuses solely on the Step-RC test of Romano and Wolf (2005) but ignores another algorithm, the Stepwise Superior Predictive Ability (Step-SPA) test proposed by Hsu et al. (2010). The Step-SPA is similar to the Step-RC but uses the SPA test of Hansen (2005) at each step instead of the RC and tends to reject more nulls than the Step-RC test.

To implement these two stepwise procedures with a pre-specified significance level, α, we compute $\widehat{c^k_{\alpha,rc}} = (\bar{f}^*_k(b) - \bar{f}_k)$ in the Step-RC test, and $\widehat{c^k_{\alpha,spa}} = (\bar{f}^*_k(b) - \bar{f}_k + \bar{f}_k I(\bar{f}_k \leq A_{n,k}))$ in the Step-SPA test. In practice, we conduct the Stepwise approaches as follows.

1. Rank $\widehat{c^k_{\alpha,rc}}$ for the RC test and $\widehat{c^k_{\alpha,spa}}$ for the SPA test in descending order and utilise its $(1-\alpha)$-th quantile as critical value $q(\alpha)$.

2. Reject the best-performing model k if $\bar{f}_k$ is greater than the critical value, $q(\alpha)$, for the RC test, or the maximum value between zero and the critical value for the SPA test. Stop the procedure if no model is rejected; otherwise, move to the next step.

3. Create a sub-sample by removing the model k from the entire sample rand re-rank the $\widehat{c^k_{\alpha,rc}}$ for the RC test and $\widehat{c^k_{\alpha,spa}}$ for the SPA test to find the critical value, $q^\star(\alpha)$. Based on the sub-sample, further reject the model i if $\bar{f}_i$ is greater than the critical value ($q^\star(\alpha)$) for the RC test, or the maximum value between zero and the critical value for the SPA test. Stop the procedure if no model is rejected; otherwise, move to the next step.

4. Repeat step **3** to reconstruct sub-samples and re-test until the procedure stops.

In our study, we apply standardised performance measures, t-statistics, for both RC- and SPA-based tests to compare their test power consistently.

2.1.5. Generalised Step-RC/SPA tests

The aforementioned approaches control the FWER, meaning that they are static in that the procedure only stops if all the remaining nulls are rejected. To build a more flexible MHT framework, Romano et al. (2007) and Hsu et al. (2014) further developed generalisations of the Step-RC/SPA tests known as the Step-RC(K)/Step-SPA(K) tests to control the FWER at K level (FWER(K)). In other words, these tests stop when the number of rejections is less than a specified value of K. Thus, a FWER(K) test allows at most K false rejections in the MHT procedure. K can be any positive integer between one and m, and a higher K indicates a greater tolerance of false rejections. According to the size of our model sample, we set $K = 2, 3, 10, 20, 100, 200$.[4] The procedure of applying the Step-RC/SPA(K)tests is as follows.

1. Rank $\widehat{c_{\alpha,rc}}$ for the RC test and $\widehat{c_{\alpha,spa}}$ for the SPA test in descending order and utilise its $(1-\alpha)$-th quantile as the critical value $q_j(\alpha)$.

2. Reject the best-performing model k if $\bar{f}_k$ is greater than the critical value, $q_j(\alpha)$, for the RC test, or the maximum value between zero and the critical value for the SPA test. Stop the procedure if the

[4] Hsu et al. (2014) examined the test power of the Step-RC/SPA(K) test across 315 Commodity Trading Advisor (CTA) funds, and set K up to a maximum of three, i.e., 1% of m. We follow this rationale and use a maximum K of 200 as our model universe contains over 20,000 trading rules.

number of rejected models is less than a given level K ($K \geq 2$); otherwise, move to the next step.

3. Create a sub-sample by removing the model k from the entire sample. Further reject the model i if $\bar{f}_i$ is greater than the critical value, for the RC test, or the maximum value between zero and the critical value for the SPA test. Stop the procedure if the number of rejected models is less than a given K; otherwise, move to the next step.

4. Repeat step **3** to reconstruct sub-samples and re-test until the procedure stops.

The final rejection threshold is determined in the last step. A noteworthy feature of the Step-RC/SPA(K) test is that its critical value is computed once some of the hypotheses are considered as rejected. Only those hypotheses that are not rejected are tested to identify the true null at each step of this procedure. This approach increases the tolerance of false rejections and allows for up to K false rejections. However, the test results fail to explore which trading rules are falsely rejected, and therefore any combination of the $k-1$ rejected strategies are potential candidates for false rejection (Romano et al., 2008). As supplementary evidence, we report the results of the Step-RC/SPA(K) tests in Appendix B.

2.2. Controlling the false discovery proportion

The FWER control may be too stringent when the number of false rejections is large, but K is small. To further tolerate the control of Type I errors, Romano et al. (2007) introduced MHT approaches controlling the FDP, where the FDP is defined as the calculated rate of the total number of erroneously rejected null hypotheses divided by the total number of significant declarations. When no false rejection occurs, the authors define FDP as zero. When asymptotically controlling the FDP at a given significance level, α, the probability function of FDP should follow $P(FDP > \gamma) \leq \alpha$, where the number γ ranges from zero to one. Romano et al. (2007) and Hsu et al. (2014) developed algorithms to asymptotically control the FDP based on this.

The FDP control approaches are direct extensions of the above-mentioned generalised FWER(K) approaches. For instance, assuming that every outperforming trading rule is rejected at the first step with probability approaching one and $S = 10$, the FDP of MHT will be $V/(V + 10)$. Further imagining $\gamma = 0.1$, $V/(V+10)$ is greater than γ as long as $V \geq 2$. As a consequence, the FDP control approach is asymptotically equivalent to the FWER(2) control. Generally, for any known S and γ, we conclude that the asymptotic FDP control is equivalent to FWER($\lceil \frac{\gamma * S}{1-\gamma} \rceil$). The MHT approach controlling the FDP, namely FDP-RC/SPA, at the level α can be achieved by implementing the FWER(K) approach. Thus, the FDP-RC/SPA algorithm is constructed by repeating the Step-RC/SPA(K) approach with K increasing from one. In empirical tests, we set γ at 0.05. Therefore, procedure for the FDP-RC/SPA test is as follows:

1. Start with $K = 1$ and $\gamma = 0.05$.

2. With a given significance level α, R denotes the number of the hypotheses rejected by the Step-RC/SPA(K).

3. If $R < K/\gamma - 1$, stop and reject all hypotheses rejected by Step-RC/SPA(K). If not, reset $K = K + 1$ and repeat Step 2.

The FDP-RC/SPA tests are designed to ensure that the realised FDP is below a specified threshold when the same hypothesis is tested on various datasets (Harvey et al., 2020). In other words, this control approach forces the tail of the FDP distributions to behave in a certain manner, that is, $P(FDP \geq \gamma) < \alpha$. This is a desirable method to control Type I errors when MHT approaches are applied to find the significance from m trading rules based on various datasets, for example, the different currency pairs in our study.

2.3. Controlling the false discovery rate

We can also manage the number of Type I errors by controlling the false discovery rate (FDR). Multiple hypothesis testing controls the FDR at level δ if:

$$FDR = E(FDP) \leq \delta, \tag{9}$$

where δ is the tolerance level that is similar to γ and α for the FDP control. Consistent with previous literature, for example, Chordia et al. (2020) and Harvey et al. (2020), we implement the control approaches proposed by Benjamini and Hochberg (1995) (BH) and its extension constructed by Benjamini and Yekutieli (2001) (BY).

Unlike the above-mentioned step-down procedures, for example, the Holm, Step-RC/SPA, and FDP-RC/SPA, BH is a step-up procedure that sorts all p-values in ascending order. This method assesses the least significant hypothesis first and then moves upwards to find the critical value. The critical value is equal to the p-value with order $j^{\star}$, which is computed as:

$$j^{\star} = max\{j : p_j \leq \frac{j \times \delta}{m}\}. \tag{10}$$

When the critical value is p_j, $(j-1)$ hypotheses are rejected. The expected number of false rejections is $m \times p_j$, and $(m \times p_j)/j$ refers to the FDR that is always less than δ as we start from the smallest p-value.

BH assumes that each hypothesis is independent of the others, whereas BY assumes partial dependence in the data. In accordance with this adjusted assumption, the definition of $j^{\star}$ in BY is given as:

$$j^{\star} = max\{j : p_j \leq \frac{j \times \delta}{m \times C_m}\}, \tag{11}$$

where $C_m = \sum_{i=1}^{m} 1/i \approx log(m) + 0.5$ that is greater than one when $m \geq 4$. As a consequence, the critical value of BY is usually lower than that of BH, so BY is a more static MHT procedure than BH.

Both BH and BY are easy to apply, but they only control the expected FDP, $E(FDP)$, not the realised FDP. When the standard deviation of the FDP is large, the realised value can be far from the targeted tolerance level δ. Harvey et al. (2020) suggest that researchers cannot assume that they are adequately controlling the FDP when applying BH and BY, so FDP-RC/SPA may be a more desirable algorithm.

We follow previous literature, for example, Hsu et al. (2014), Hsu et al. (2016), Chordia et al. (2020), and Harvey et al. (2020), and set $\alpha = 0.05$ in every approach, $\gamma = 0.05$ in the FDP control, and $\delta = 0.05$ in the FDR control. To implement the bootstrapping procedure of the RC/SPA-based approaches, we set $Q = 4$ and $b = 500$. In addition, we skip the first 251 observations of each currency to ensure that every strategy creates trading signals, as some trading rules require up to 250 trading days to generate the first trading position.

3. Data and return calculations

Before evaluating the various MHT approaches, this section describes the sources and summary statistics of our data universe. We further present the daily raw and excess return calculations and outline the method used to measure the transaction costs of our trading strategies.

3.1. Foreign exchange rates

Our data universe consists of daily closing spot and one-month forward exchange rates between the U.S. dollar and foreign currencies from 48 regions:[5] Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Euro area, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Japan, Kuwait, Malaysia, Mexico, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Russia, Saudi Arabia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Ukraine, and the United Kingdom. Both the spot and one-month forward exchange rates are obtained from Barclays Bank International and Reuters via Datastream. The entire sample period ranges from 5th January 1976 to 18th March 2019.[6] Following Lustig et al. (2011), we further identify a *'developed'* sub-sample comprising 15 countries: Australia, Belgium, Canada, Denmark, Euro area, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, and the United Kingdom. As the Euro was introduced on 1 January 1999, the *'developed'* sub-sample consists of 10 currencies thereafter. The remaining currencies are referred to as the *'others'* sub-sample.

3.2. Gross and excess returns

We calculate daily gross returns by buying one unit of a foreign currency and holding overnight, where zero transaction costs are assumed.[7] Following previous literature, for example, Fama (1984), we work on

[5]The same data sample is also employed by Menkhoff et al. (2012*b*). The closing spot rates are calculated using a five minute fixing window around 16:00 London time. We measure the exchange rates as the U.S. dollar per unit of foreign currency, which is consistent with the measure of Hsu et al. (2016) and Boons and Prado (2019).

[6]The exchange rates versus U.S. dollar are available from 1983 in our data sources. To extend the sample period, following Burnside et al. (2010) and Menkhoff et al. (2012*b*), we construct the relevant exchange rates of 15 regions versus the U.S. Dollar based on the exchange rates against the Great British Pound before 1983. These regions are: Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden and Switzerland.

[7]Following Hsu et al. (2016), we assume zero transaction costs when calculating gross and excess returns in this section, however, we later present a discussion of transaction costs in Subsection 3.3.

the logarithms of the spot rate for ease of exposition. The daily gross return of currency k on day t is $r_t^k = s_t^k - s_{t-1}^k$, where s_t^k is the end-of-day spot exchange rate (U.S. dollar per unit of foreign currency k) in logarithmic terms. When r_t^k is positive (negative), the currency k depreciates (appreciates) against the U.S. dollar. Table 3 presents the statistics of gross returns. From this table we can see that the Slovakian Koruna appreciates the most on average (1.35bps per day), and the Russian Ruble depreciates the most on average (4.32bps per day) over our sample period. The Russian Ruble is also the most volatile currency as it reports the highest standard deviation (152bps per day). The least volatile currency is the Saudi Arabian Riyal, which has the lowest daily standard deviation (2bps per day).

Next, we compute daily excess return as $rx_t^k = r_t^k - (i_{t-1}^k - i_{t-1})$, where $i_{t-1}^k - i_{t-1}$ is the difference between the foreign interest rate of currency k and the domestic interest rate. As highlighted by Taylor (1989), Akram et al. (2008), and Menkhoff et al. (2012*a*), the interest rate differentials are equivalent to the forward discounts when covered interest parity holds. We achieve the monthly interest differential as $f_{t-1}^k - s_{t-1}^k$, where f_{t-1}^k is the one-month forward exchange rate between currency k and the U.S. dollar on day $t-1$. Next, we convert the monthly differentials to a daily frequency and compute excess returns of currency k on day t as follows:

$$rx_t^k = r_t^k - (i_{t-1}^k - i_{t-1}) \approx r_t^k - \frac{f_{t-1}^k - s_{t-1}^k}{30}, \tag{12}$$

in which all notations are as noted above.

According to the summary of excess returns shown in Table 3, the Indian Rupee generates the highest daily excess return on average (5.57bps per day) with the Ukrainian Hryvnia producing the lowest (-24.59bps per day).[8] The most volatile and stable currencies are the Indian Rupee, with a daily standard deviation of 1.4%, and the Saudi Arabian Riyal, with a daily standard deviation of 0.02%, respectively. Besides gross and excess returns, another vital dimension of currency trading is return persistence. Therefore, we report first-order autocorrelation of both the gross and excess returns for each currency. Egypt, Hong Kong, India, Russia, South Korea, and Thailand all register first-order autocorrelations of greater than 0.1

[8]The unusual change of interest rate in 2014 caused the extreme excess return in the Ukrainian Hryvnia.

regardless of the return type. However, none of the sampled currencies exhibit statistically significant first-order autocorrelation. The average gross return in the *'developed'* subsample at -0.07bps exceeds that in the *'others'* subsample at -0.89bps. This implies that the currencies in the *'developed'* subsample tend to depreciate less relative to the U.S. dollar than the *'others'* over the sample period. On the other hand, the *'others'* currencies yield lower excess daily returns on average (0.09bps) than the *'developed'* sample (0.32 bps). After excluding the Ukrainian Hryvnia outlier, the mean excess return in the *'others'* sample increases to 0.86bps, exceeding the mean in the *'developed'* sample. Finally, currencies in both samples show similar risks, with the average standard deviation of gross (excess) returns being 65 (67)bps in the *'developed'* sample and 65 (64)bps in the *'others'* sample.

3.3. Transaction costs

Transaction costs significantly impact investment revenues in the FX market. Timmermann and Granger (2004) document that transaction costs could easily eliminate the excess returns generated by exchange rate spreads. Burnside et al. (2007) and Ramadorai (2008) further indicate that this phenomenon is more likely to appear in emerging markets. In this study, following Neely and Weller (2013), who propose that the transaction cost is driven by the bid-ask spread and interest rates, we calculate the approximate one-way transaction costs on any particular trading day by using one-third of the bid-ask spread of one-month forward exchange rates. This transaction cost measurement is adopted in other technical trading rule based studies, e.g., Hsu et al. (2016) and Filippou et al. (2018). We obtain daily bid and ask quotes for the one-month forward exchange rates from Datastream. In the next section, we detail the TTRs used for our evaluation.

4. Technical trading rules

TTR are based on the analysis of past price movements, followed by the construction of trading signals through quantitative analysis of time series data (Eatwell et al., 1992). The TTR universe is ideal for assessing MHT approaches due to the large number of trading rules being considered simultaneously. Following Hsu et al. (2016), we implement TTR from five families: *oscillator, filter, moving average, support-resistance*, and *channel breakout*. These five classes of TTR are extensively used by FX traders (Taylor and Allen, 1992, Menkhoff and Taylor, 2007). *Oscillator trading rules (O)* allocate trading positions according to the

Table 3: Summary statistics in basis points

Regions	Daily gross returns					Daily excess returns					Sample period	
	Mean	Max	Min	SD	AR(1)	Mean	Max	Min	SD	AR(1)	Start	End
Developed												
Australia	-0.188	670.062	-882.792	72.631	0.004	0.910	671.177	-881.860	72.676	0.005	12/17/1984	3/18/2019
Belgium	0.072	535.486	-452.412	65.360	-0.006	0.700	535.736	-452.443	66.070	-0.004	1/5/1976	12/30/1998
Canada	-0.263	504.622	-433.752	43.536	-0.003	0.038	504.569	-433.476	43.966	-0.001	1/5/1976	3/18/2019
Denmark	-0.142	575.521	-505.296	62.008	0.007	0.438	575.652	-501.649	62.287	0.010	1/5/1976	3/18/2019
Euro Area	-0.409	461.742	-384.447	61.770	0.022	-0.286	461.802	-384.000	60.368	0.011	1/1/1999	3/18/2019
France	-0.375	575.930	-595.400	63.984	0.000	0.370	576.472	-586.879	64.607	0.004	1/5/1976	12/30/1998
Germany	0.284	578.497	-526.240	62.615	0.015	-0.130	577.109	-528.845	62.925	0.018	1/5/1976	3/18/2019
Italy	-0.909	1,322.927	-1,341.534	82.012	0.021	0.846	1,323.421	-1,341.176	91.932	0.005	1/5/1976	12/30/1998
Japan	0.868	658.176	-371.021	64.845	0.005	-0.431	656.035	-371.116	66.611	0.008	6/9/1978	3/18/2019
Netherlands	0.452	554.994	-607.189	63.998	0.015	0.107	554.111	-607.399	64.719	0.018	1/5/1976	12/30/1998
New Zealand	0.380	587.776	-848.581	76.480	0.015	1.871	588.748	-850.318	76.540	0.016	12/17/1984	3/18/2019
Norway	-0.439	645.806	-689.929	65.867	0.002	0.351	646.432	-682.448	66.228	0.003	1/5/1976	3/18/2019
Sweden	-0.713	554.743	-1,774.093	66.543	0.001	-0.096	554.840	-1,771.661	67.004	-0.001	1/5/1976	3/18/2019
Switzerland	0.808	1,141.888	-847.480	70.732	0.018	-0.183	1,141.634	-848.145	71.066	0.021	1/5/1976	3/18/2019
United Kingdom	-0.498	466.816	-831.133	60.021	0.057	0.222	556.930	-831.220	60.769	0.054	1/5/1976	3/18/2019
Others												
Austria	0.596	458.896	-538.755	66.417	-0.012	0.297	457.621	-541.997	67.131	-0.010	1/5/1976	12/30/1998
Brazil	-2.075	1,177.830	-1,080.095	96.277	0.045	2.657	741.844	-809.379	96.216	-0.016	3/30/2004	3/18/2019
Bulgaria	-0.053	461.259	-384.844	59.973	0.010	-0.179	462.941	-383.031	59.213	0.007	3/30/2004	3/18/2019
Croatia	-0.145	461.905	-382.862	61.939	0.003	0.236	466.962	-380.366	61.338	0.002	3/30/2004	3/18/2019
Cyprus	0.857	332.418	-209.819	56.538	0.001	2.002	191.820	-171.696	49.975	-0.020	3/30/2004	12/30/2007
Czech Republic	0.356	521.919	-826.379	72.036	0.025	0.542	522.457	-831.001	74.157	0.025	1/1/1997	3/18/2019
Egypt	-2.571	613.075	-4,317.734	71.159	0.226	4.043	620.068	-4,206.181	82.390	0.277	3/30/2004	3/18/2019
Finland	1.000	265.838	-221.164	56.051	-0.001	-2.757	191.772	-191.848	55.877	-0.015	1/1/1997	12/30/1998
Greece	-1.782	317.060	-792.245	60.034	-0.008	-1.280	281.232	-786.155	63.751	0.063	1/1/1997	12/30/1999
Hong Kong	0.066	185.190	-57.197	5.186	0.279	-0.017	206.904	-52.299	5.633	0.385	10/12/1983	3/18/2019
Hungary	-1.537	519.983	-630.470	79.110	0.017	1.155	523.442	-629.843	83.523	0.018	10/28/1997	3/18/2019
Iceland	-1.189	306.388	-351.175	35.839	0.066	0.641	307.687	-321.156	36.852	0.058	3/30/2004	3/18/2019
India	-2.902	2,041.611	-2,393.635	127.924	0.122	5.570	2,049.462	-2,388.801	137.587	0.144	10/28/1997	3/18/2019
Indonesia	-0.731	559.713	-904.172	64.743	0.006	0.490	395.814	-862.636	65.674	0.000	1/1/1997	3/18/2019
Ireland	-0.288	287.342	-457.492	44.587	0.032	0.651	288.426	-456.920	48.348	0.043	1/5/1976	12/30/1998
Israel	-1.502	501.557	-657.267	62.483	0.024	0.774	503.250	-647.290	63.516	0.028	3/30/2004	3/18/2019
Kuwait	-0.025	361.038	-354.169	14.780	-0.204	0.188	361.204	-354.013	14.848	-0.213	1/1/1997	3/18/2019
Malaysia	-0.584	1,337.945	-3,569.381	61.472	0.045	0.894	1,195.628	-3,549.453	63.674	0.129	12/17/1984	3/18/2019
Mexico	-2.763	1,519.222	-1,894.264	88.990	-0.045	1.090	533.539	-752.856	67.420	0.013	1/1/1997	3/18/2019
Philippines	-0.979	1,109.571	-1,263.194	49.434	0.076	0.248	1,114.414	-1,255.531	50.629	0.089	1/1/1997	3/18/2019
Poland	-0.689	669.735	-569.624	77.433	0.036	1.077	670.542	-568.148	84.768	0.025	2/12/2002	3/18/2019
Portugal	-3.107	579.456	-1,599.898	69.308	-0.051	0.948	581.824	-1,598.719	70.277	-0.048	1/5/1976	12/30/1998
Russia	-4.316	3,580.780	-4,824.951	152.509	0.109	0.271	1,563.057	-1,420.050	86.368	0.043	3/30/2004	3/18/2019
Saudi Arabia	0.000	42.421	-65.098	2.081	-0.142	0.052	42.116	-66.089	2.226	-0.123	1/1/1997	3/18/2019
Singapore	0.529	395.772	-248.060	33.494	-0.038	0.128	398.507	-247.356	33.532	-0.037	12/17/1984	3/18/2019
Slovakia	1.352	394.149	-615.376	67.711	0.044	5.215	394.555	-378.395	67.945	0.047	2/12/2002	12/30/2008
Slovenia	-0.713	333.226	-302.043	60.752	-0.004	1.099	195.746	-175.608	53.302	-0.026	3/30/2004	12/28/2006
South Africa	-2.781	1,211.157	-1,288.012	96.892	0.019	-0.085	1,253.883	-1,285.458	97.047	0.022	10/19/1983	3/18/2019
South Korea	-0.515	1,979.452	-1,364.511	82.553	0.104	0.680	1,324.225	-1,037.079	66.742	0.008	2/12/2002	3/18/2019
Spain	-1.494	530.181	-2159.427	68.280	-0.020	1.192	534.427	-2,137.734	69.364	-0.018	1/12/1976	12/30/1998
Taiwan	-0.223	262.103	-453.456	26.428	0.093	-0.546	260.725	-453.285	27.209	0.093	1/1/1997	3/18/2019
Thailand	-0.328	692.523	-913.324	51.411	0.107	0.385	700.079	-911.549	54.541	0.110	1/1/1997	3/18/2019
Ukraine	-3.746	2,139.269	-3,987.781	118.908	-0.099	-24.585	2,141.121	-3,984.062	135.758	0.003	3/30/2004	3/18/2019

This table presents summary statistics of daily gross/excess returns on holding foreign currencies versus the U.S. dollar during the sample period indicated in the final columns labelled 'Sample Period'. The statistics include the average daily log returns (Mean), the largest daily return (Max), the smallest daily return (Min), the standard deviation of daily returns (SD), and the first-order autocorrelation of daily returns (AR(1)). The statistics, except for AR(1), are shown in basis point (bps). The last two columns report the start and end dates for each currency. Following the adoption of the Euro, a number of currencies are omitted: Austria, Belgium, Finland, France, Italy, Ireland, Netherlands, Portugal, Spain in 1999; the Greek Drachma in 2000; the Slovenian Tolar in 2007; the Cypriot Pound in 2008; and the Slovakian Koruna in 2009. The sample period is 1/5/1976 to 3/18/2019. In each of the 'Developed' and 'Others' sub-groups, the regions are presented in alphabetical order. We follow Lustig et al. (2011) in classifying the 'Developed' sample.

market corrections after FX rate fluctuation; *filter rules (F)* generate buying (selling) trading signals when the spot rate of a currency increases (decreases) by a benchmark percentage; *moving average trading rules (MV)* follow market historical trends or identify breaks in trends, where the trends are identified by the relationship between spot rates and the moving average over a given period, or the comparison between the moving averages across different periods; *support-resistance rules (Sr)* depend on the assumption that the FX rate will continue to decline (grow) once it breaks through the support (resistance) level; *channel break rules (CB)* are similar to the support-resistance rules but use time-varying support and resistance levels. In each category, we further classify the TTRs into various sub-categories in response to holding periods or signal determining procedures.

As a variety of parameters are used, we build 21,195 TTRs when trading a given exchange rate pair, including 2,835 *filter rules*, 12,870 *moving average rules*, 1,890 *support-resistance signals*, 3,000 *channel-breakout rules*, and 600 *oscillator trading rules*. Our TTR sample consists of $21,195 \times 48 = 1,017,360$ trading rules in total. Appendix A presents more precise details related to the construction of the TTRs.

4.1. Performance metrics

We apply four performance appraisals to evaluate the performance of TTR after considering transaction costs. First, we measure performance by computing the mean excess return after subtracting transaction costs. For the k-th trading rule of currency j on day t, we calculate the excess return after transaction costs as follows:

$$R(Total)_{k,t} = S_{k,t-1} rx_t^j - c_t^j, \tag{13}$$

where $S_{k,t-1}$ represents the trading signal determined by the k-th trading rule on day $t-1$ and c_t^j is the transaction cost of currency j. Then, the mean excess return after transaction costs is shown as follows:

$$\bar{R}(Total)_k = \frac{1}{n}\sum_{t=1}^{n} R(Total)_{k,t}, \tag{14}$$

where n is the number of observations, which is consistent with the definition in Section 2. This is a fundamental performance measure to evaluate the profitability of a given trading rule, but its major limitations is that such a basic metric does not consider the riskiness of the trade rules. To address this, we also employ the ex-post Sharpe Ratio (SR) introduced by Sharpe (1966). We compute the SR for the k-th trading rule as follows:

$$SR(Total)_k = \bar{R}(Total)_k / \sigma(Total)_k, \tag{15}$$

where σ_k is the standard deviation of return series created by the k-th trading rule.

Furthermore, Hsu et al. (2016) demonstrate that asset managers evaluate the skill inherent in a trading strategy by excluding the specific regional risk premium from the total excess returns to avoid regional risk premium variations. In comparison with other asset classes, FX market investment risk is more directly affected by fundamental economic policies, for example, unexpected monetary policy and funding restrictions (Cornell and Dietrich, 1978, Cornell, 1982, Hodrick and Srivastava, 1986, Froot and Thaler, 1990). Hsu et al. (2016) and Dahlquist and Hasseltoft (2020) also conclude that investors require a high premium for trading currencies of unstable economies, with Sweeney (1986) finding that long-term buy-and-hold could be a method to capture country specific risk premium. We decompose the total excess returns into a 'tilting' component (the profits generated by holding an average trading position of a particular over the entire sample period multiplied by the excess return on day t, similar to a buy-and-hold strategy) and a 'Market Timing (MT)' component (the profits generated by correctly timing the entry and exit points of the trade for a particular currency). The total excess return after the transaction costs of the k-th trading rule can be converted into the sum of two components, given as:

$$R(Total)_{k,t} = R(Tilt)_{k,t} + R(MT)_{k,t}, \tag{16}$$

where $R(Tilt)_{k,t}$ and $R(MT)_{k,t}$ represent the 'tilting' and MT, respectively, defined as follows:

$$\begin{aligned} R(Tilt)_{k,t} &= [\frac{1}{n}\sum_{t=1}^{n} S_{k,t-1}]r_t, \\ R(MT)_{k,t} &= R(Total)_{k,t} - [\frac{1}{n}\sum_{t=1}^{n} S_{k,t-1}]r_t. \end{aligned} \tag{17}$$

Then, our third performance metric is the mean excess return using the MT component, given as:

$$\bar{R}(MT)_k = [\bar{R}_k - [\frac{1}{n}\sum_{t=1}^{n} S_{j,t-1}]\frac{\sum_{t=1}^{n} r_t}{n}]. \tag{18}$$

Lastly, the fourth measure is the SR of the excess return using MT as follows:

$$SR(MT)_k = \bar{R}(MT)_k / \sigma(MT)_k, \tag{19}$$

where $\sigma(MT)_k$ is the standard deviation of $R(MT)_k$.

To ensure that all the performance metrics are standardised, we measure the t-statistics for the four performance metrics, so the $\sqrt{n}\bar{f}_k$ mentioned in Equation 4 changes to:

$$\sqrt{n}\bar{f}_k = \sqrt{n}\frac{\bar{R}_k}{\sigma_k}, \tag{20}$$

where $\bar{R}_k$ can be the mean excess relative to the total or MT, σ_k refers to the corresponding standard deviation. We then follow Bailey and Lopez de Prado (2012) and calculate the t-statistics of the SR, so the above-mentioned $\sqrt{n}\bar{f}_k$ converts to:

$$\sqrt{n}\bar{f}_k = \frac{SR_k * \sqrt{n}}{\sqrt{1 - \Gamma_3 * SR_k + \frac{\Gamma_4 - 1}{4} * SR_k^2}}, \tag{21}$$

where Γ_3 and Γ_4 represent skewness and kurtosis, respectively, and SR_k is the SR of the excess return relative to both total and MT component.

Consequently, we conduct four performance appraisals for each trading rule: the mean of the total excess return ($\bar{R}(Total)_k$), the SR of the total excess return ($SR(Total)_k$), the mean of the MT excess returns ($\bar{R}(MT)_k$), and the SR relative to the MT ($SR(MT)_k$).

4.2. Strategy performance

We next report the performance of the TTRs using a sample period from 12/20/1976 to 3/18/2019. We skip the first 250 observations to ensure every TTR is able to generate its first trading signal. Table 4 summarises the annualised average return, annualised standard deviation, SR, and family name of the best-performing trading rules (with the highest mean excess return or SR). In line with Hsu et al. (2016), the performance metrics in this table are based on total excess returns without adjusting for transaction costs. The best-performing TTRs all report positive returns and SRs. In terms of mean total excess return, one of the MA4 trading rules for the Swedish Krona generates the highest average return at 9.16% per annum, in the *'developed'* sample, and an MA1 trading rule of the Ukrainian Hryvnia produces the highest mean at 30.77% per annum, in the *'others'* sample. When the performance metric is the SR, an MA5 TTR trading the Dutch Guilder (Netherlands) produces the largest SR at 0.92 per annum, across the *'developed'* countries, and an MA1 trading rule of the Ukrainian Hryvnia generates the highest SR at 3.52 per annum, in the *'others'* sample. Although a currency from the *'others'* sample yields the highest mean return, the average annualised mean across the *'developed'* currencies at 7.13%, exceeds that across the *'others'* currencies at 5.73%. When the Sharpe ratio is the performance metric, the average SR of the best-performing TTRs is 0.84 in the *'others'* sample, which exceeds that of the *'developed'* sample (0.74). Regardless of the performance metric chosen, the best-performing TTRs generate higher volatility in the *'developed'* sample. Overall, across both returns and volatility, our results align with previous literature whereby FX trading profits come mainly from *'other'* currencies, with 'developed' currencies being less profitable but more stable (Lustig et al., 2011).

5. In-sample analysis

After naively applying a comprehensive sample of TTRs as outlined in the previous section, we next demonstrate empirical evidence to illustrate the potential for MHT-hacking. We implement the in-sample evaluation as follows. We first count the number of trading rules that result in significantly positive total and MT returns ($\#profitable$), with Tables 5 and 6 reporting the results. Next, Tables 7 and 8 summarise the number of profitable trading rules in terms of SRs of total and MT returns ($\#risk\text{-}adjusted$), respectively. As outlined in Section 2, FX market practitioners are also interested in determining how many, and which, currencies should be considered when seeking to apply TTRs. For this reason, we also count how many currencies demonstrate instances of significant outperforming TTRs, and call them *"Predictable Currencies"*.

Table 4: Performance summary of the best-performing technical trading rules

Performance metric	Mean return				Sharpe ratio			
	Mean	*SD*	*SR*	*Best Rule*	*Mean*	*SD*	*SR*	*Best Rule*
Developed								
Australia	0.0612	0.1212	0.5050	F3	0.0423	0.0802	0.5275	MA5
Belgium	0.0705	0.0901	0.7827	MA4	0.0582	0.0680	0.8553	MA5
Canada	0.0413	0.0824	0.5004	MA1	0.0413	0.0824	0.5004	MA1
Denmark	0.0826	0.1132	0.7296	MA4	0.0638	0.0792	0.8053	MA5
Euro Area	0.0476	0.0774	0.6148	MA4	0.0476	0.0774	0.6148	MA4
France	0.0712	0.0881	0.8089	MA1	0.0579	0.0643	0.8998	MA5
Germany	0.0884	0.1172	0.7540	MA1	0.0633	0.0796	0.7959	MA5
Italy	0.0636	0.0837	0.7605	MA4	0.0499	0.0628	0.7950	MA5
Japan	0.0891	0.1208	0.7374	MA1	0.0612	0.0782	0.7831	MA5
Netherlands	0.0731	0.0882	0.8285	MA1	0.0674	0.0732	0.9202	MA5
New Zealand	0.0727	0.1229	0.5919	MA1	0.0727	0.1229	0.5919	MA1
Norway	0.0686	0.1242	0.5522	MA1	0.0181	0.0246	0.7341	CB2
Sweden	0.0916	0.1251	0.7319	MA4	0.0916	0.1251	0.7319	MA4
Switzerland	0.0720	0.1322	0.5448	MA1	0.0340	0.0486	0.6995	CB2
United Kingdom	0.0756	0.1139	0.6639	F3	0.0713	0.0792	0.8998	MA5
Others								
Austria	0.0657	0.0915	0.7179	MA1	0.0366	0.0455	0.8034	CB2
Brazil	0.0538	0.1062	0.5065	MA1	0.0538	0.1062	0.5065	MA1
Bulgaria	0.0353	0.0653	0.5399	MA1	0.0353	0.0653	0.5399	MA1
Croatia	0.0255	0.0584	0.4362	MA1	0.0255	0.0584	0.4362	MA1
Cyprus	0.0126	0.0276	0.4554	MA1	0.0126	0.0276	0.4554	MA1
Czech Republic	0.0473	0.0993	0.4762	MA4	0.0389	0.0685	0.5680	MA5
Egypt	0.0473	0.0910	0.5202	MA1	0.0036	0.0038	0.9644	O2
Finland	0.0149	0.0225	0.6634	MA1	0.0149	0.0225	0.6634	MA1
Greece	0.0127	0.0314	0.4046	MA4	0.0127	0.0314	0.4046	MA4
Hong Kong	0.0041	0.0061	0.6777	O1	0.0041	0.0061	0.6777	O1
Hungary	0.0474	0.1102	0.4304	O1	0.0255	0.0398	0.6410	O2
India	0.0433	0.0484	0.8942	MA1	0.0369	0.0409	0.9025	MA5
Indonesia	0.1975	0.1418	1.3933	MA5	0.1560	0.0936	1.6661	MA1
Ireland	0.0656	0.0898	0.7306	MA1	0.0588	0.0743	0.7920	MA5
Israel	0.0225	0.0371	0.6076	MA5	0.0225	0.0371	0.6076	MA5
Iceland	0.0468	0.1014	0.4612	MA1	0.0425	0.0642	0.6620	MA5
Kuwait	0.0114	0.0200	0.5711	MA3	0.0114	0.0200	0.5711	MA3
Malaysia	0.1729	0.1056	1.6371	MA4	0.0988	0.0484	2.0395	MA1
Mexico	0.0453	0.0906	0.4997	MA1	0.0453	0.0906	0.4997	MA1
Philippines	0.0408	0.0679	0.6005	MA1	0.0337	0.0547	0.6154	MA5
Poland	0.0421	0.1000	0.4213	MA1	0.0116	0.0187	0.6204	O2
Portugal	0.0656	0.0953	0.6885	MA4	0.0656	0.0953	0.6885	MA4
Russia	0.0697	0.0941	0.7403	MA4	0.0697	0.0941	0.7403	MA4
Saudi Arabia	0.0020	0.0030	0.6744	O1	0.0011	0.0011	1.0461	O1
Singapore	0.0393	0.0548	0.7177	MA4	0.0393	0.0548	0.7177	MA4
Slovakia	0.0364	0.0507	0.7174	MA4	0.0364	0.0507	0.7174	MA4
Slovenia	0.0113	0.0252	0.4476	MA1	0.0113	0.0252	0.4476	MA1
South Africa	0.1032	0.1643	0.6284	MA1	0.1032	0.1643	0.6284	MA1
South Korea	0.0317	0.0734	0.4317	O2	0.0099	0.0196	0.5033	O1
Spain	0.0639	0.0946	0.6753	MA4	0.0605	0.0767	0.7882	MA4
Taiwan	0.0356	0.0365	0.9730	MA1	0.0343	0.0291	1.1790	MA5
Thailand	0.0685	0.0732	0.9358	MA1	0.0627	0.0550	1.1417	MA5
Ukraine	0.3077	0.1514	2.0326	MA1	0.2268	0.0644	3.5235	MA1

This table summarises the annualised average return (*Mean*), annualised standard deviation (*SD*) and Sharpe Ratio (*SR*) of the best-performing technical trading rules (i.e., those with the highest mean return or SR for each currency). Best rule is populated according to the category of trading rules defined in Appendix A. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The sample period considered is 12/20/1976 to 3/18/2019. In line with Hsu et al. (2016), we do not control for the transaction costs at this stage.

The count of *"Predictable Currencies"* is given in the last line of each of the Tables 5 to 8 above.[9] Finally, we explore the difference in controlling for FDP and FDR.

First, our in-sample analysis supports the application of MHT approaches in the FX market. The results across our four adopted criteria indicate that applying SHT produces Type I errors when assessing a large number of trading rules. For instance, based on the mean of returns, Table 5 shows that the SHT indicates that 28 of 48 currencies are predictable at the 5% level, nine in the *'developed'* sample and 19 in the *'others'* sample. By contrast, after controlling for Type I errors, we find that the highest number of *'Predictable Currencies'* is 11, which is less than half that uncovered by the SHT approach. Moreover, according to the $\#profitable$ of each currency, the SHT uncovers a large number of t-statistics over the rejection threshold of 1.96, with an average of 2,612 significant rejections across all currencies. A few currencies account for a large number of the significant trading strategies uncovered. For instance, 10,787 TTRs yield significant outperforming returns when trading Indonesian Rupiah, with 8,449 strategies leading to significant outperformance when trading Spanish Pesetas, and 8,044 for Belgian Francs. The $\#profitable$ is greatly reduced when applying MHT, with the largest differences appearing for Danish Kroner (4,642 for SHT vs. zero across all MHT approaches), and French Francs (4,080 for SHT vs. zero across all MHT approaches). Changing the criterion to SR, Tables 7 and 8 also show that SHT yields far higher *#risk-adjusted* and count of *'Predictable Currencies'* than MHT. For instance, in Table 7, SHT uncovers 28 *'Predictable Currencies'* versus a maximum of 11 across all MHT approaches. Similar to the dynamic with $\#profitable$, the greatest difference in *#risk-adjusted* again appears when trading Danish Kroner (4,642 for SHT vs. zero across all MHT approaches).

We next highlight the need to appropriately select a specific MHT approach to implement. First, we find that the application of different MHT approaches provides FX market participants with conflicting information.[10] For instance, in Table 5, the results of the Bonferroni and Holm approaches show that trading Belgian Francs is not desirable as only one significant rejection is uncovered. However, if the BH and BY approaches are adopted, Belgian Francs become very attractive as more than 2,000 outperforming

[9] As the RC and SPA tests do not provide a count of how many models significantly outperform the benchmark, results for these approaches are omitted from our comparison. Furthermore, we report the critical values and K of FDP-SPA/RC in Appendix C.

[10] Hsu et al. (2016) discover differences in applying Step-RC and Step-SPA test but do not consider other MHT approaches.

trading rules are uncovered. Second, some approaches, such as controlling the FWER, may lead to potentially profitable trading signals being missed. For instance, the average $\#profitable$ across FWER approaches is only 51, whereas the average $\#profitable$ of the FDP and FDR approaches are 216 and 705, respectively. In the next section, we adopt out-of-sample testing to establish if the additional trading rules identified by the FDP and FDR approaches lead to profitable positions in reality. Tables 7 and 8 report the same patterns when focusing on reward-to-risk ratios. For instance, Table 7 reports that MHT approaches controlling the FWER still yield the lowest average $\#$*risk-adjusted* at 72, whereas the average $\#$*risk-adjusted* is for FDP and FDR is far higher at 445 and 743, respectively. When controlling the FDP, for trading most currencies, $\#$*risk-adjusted* of FDP-RC/SPA sits between BY and BH, which is consistent with Chordia et al. (2020). Also, many of the MHT approaches report diverging conclusions. In Table 7, trading Italian Lira is attractive when controlling the FDR (over 1,000 outperforming TTRs). Yet, Lira is not deemed to be a *'Predictable Currency'* when controlling the FWER and FDP, as no significant rejections are uncovered.

We now move our attention to the RC/SPA-based tests, which are also known as the data snooping bias tests (Romano et al., 2010). These approaches determine the critical values based on the bootstrapped distributions, which help to dilute the impact of extreme returns. Consistent with previous literature, for example, Hsu et al. (2010) and Hsu et al. (2018), the RC-based tests on average yield lower $\#profitable$ than the SPA-based tests. Table 5 reports that two RC-based tests fail to uncover any profitable currencies in our *'developed'* sample, whereas the remaining MHT approaches successfully find at least one predicable currency. In particular, the FDP-SPA test identifies 4,999 significant $\#profitable$ TTRs when trading the Belgian Francs, whereas the FDP-RC test uncovers zero. In the *'others'* sample, the Step-RC test uncovers only 372, 14, 6, and 188 significant outperforming rules when trading Indonesian Rupiah, Taiwanese Dollar, Thai Baht, and Ukrainian Hryvnia, respectively. However, for each of these four currencies, the Step-SPA test finds at least 268 more instances of significance than the Step-RC. We arrive at the same conclusions using the SR, with $\#$*risk-adjusted* further validating the superior test power of SPA-based tests compared to RC-based tests. For instance, six currencies in Table 7 are predictable according to the Step-RC test, whereas the Step-SPA test identifies eight *'Predictable Currencies'*. Table 7 shows that the Step-RC uncovers nine outperforming rules for Japanese Yen when controlling the FWER, whereas $\#$*risk-adjusted* increases

Table 5: The number of outperforming technical trading rules based on total returns ($\#profitable$)

Currencies	MHT control approaches								SHT
	FWER				FDP		FDR		
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Developed									
Australia	0	0	0	0	0	0	0	0	0
Belgium	1	1	0	43	0	4,999	6,317	2,006	8,044
Canada	0	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	0	0	4,642
Euro Area	0	0	0	0	0	0	0	0	4
France	0	0	0	0	0	0	0	0	4,080
Germany	0	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	5,646	1,039	8,172
Japan	0	0	0	0	0	0	5,484	0	7,632
Netherlands	0	0	0	0	0	0	0	0	3
New Zealand	0	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0	12
Sweden	0	0	0	0	0	0	0	0	2,321
Switzerland	0	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0	0
Others									
Austria	0	0	0	0	0	0	0	0	4,265
Brazil	0	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0	1,031
Egypt	0	0	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0	0	0	68
Greece	0	0	0	0	0	0	0	0	61
Hong Kong	0	0	0	0	0	0	0	0	0
Hungary	0	0	0	0	0	0	0	0	55
India	0	0	0	0	0	0	1590	0	4,231
Indonesia	1,875	1,925	372	768	518	8,844	10,097	7,440	10,787
Ireland	0	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0	0
Malaysia	14	14	0	0	0	0	556	24	1,676
Mexico	0	0	0	0	0	0	0	0	24
Philippines	0	0	0	0	0	0	0	0	1,980
Poland	0	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0	2
Russia	0	0	0	0	0	0	0	0	3,302
Saudi Arabia	0	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	4,201	0	7,774
Slovenia	0	0	0	0	0	0	0	0	80
South Africa	0	0	0	0	0	0	0	0	0
South Korea	0	0	0	0	0	0	0	0	232
Spain	0	0	0	0	0	0	6,144	0	8,449
Taiwan	247	249	14	282	14	1,966	3,234	1,290	4,564
Thailand	159	159	6	350	6	2,722	4,069	1,903	5,566
Ukraine	1,265	1,279	188	744	189	1,546	4,129	2,593	5,120
Count of Predictable Currencies	6	6	4	5	4	5	11	7	28

This table reports the number of significant outperforming trading rules in 21,195 TTRs based on eight MHT control approaches. The performance appraisal is the mean of total excess returns. The total return is equivalent to the daily excess return of each currency subtracting the transaction cost. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). Bonferroni and Holm are the results of the fundamental FWER control approaches of Bonferroni (1936) and Holm (1979), respectively. Step-RC refers to the stepwise reality check of Romano and Wolf (2005). Step-SPA represents the stepwise superiority predictive ability test documented by Hsu et al. (2010). Both stepwise procedures also control FWER. FDP-RC/SPA refers to the BRC/SPA tests controlling FDP, where $\gamma = 5\%$. BH and BY are the FDR control approaches of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) where $\delta = 5\%$. Following Chordia et al. (2020), we set the level of significance, α, at 5% for each control approach. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table 6: The number of outperforming technical trading rules based on MT returns (#*profitable*)

Currencies	MHT control approaches								SHT
	FWER				FDP		FDR		
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Developed									
Australia	0	0	0	0	0	0	0	0	0
Belgium	11	11	0	73	0	4,995	6,948	2,802	8,389
Canada	0	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	0	0	5,174
Euro Area	0	0	0	0	0	0	0	0	4
France	0	0	0	19	0	68	718	0	4,882
Germany	0	0	0	0	0	0	0	0	0
Italy	88	88	0	0	0	0	9,589	3,288	10,804
Japan	0	0	0	15	0	15	6,102	1,044	8,327
Netherlands	0	0	0	0	0	0	0	0	5
New Zealand	0	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0	17
Sweden	0	0	0	0	0	0	0	0	2,807
Switzerland	0	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0	0
Others									
Austria	0	0	0	0	0	0	0	0	4,820
Brazil	0	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0	1,110
Egypt	0	0	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0	0	0	78
Greece	0	0	0	0	0	0	0	0	79
Hong Kong	0	0	0	0	0	0	0	0	0
Hungary	0	0	0	0	0	0	0	0	59
India	0	0	0	0	0	0	2006	0	4,402
Indonesia	2,473	2,556	593	1,146	876	9,471	10,475	8,199	10,994
Ireland	0	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0	0
Malaysia	2	2	0	0	0	0	188	0	1,295
Mexico	0	0	0	0	0	0	0	0	24
Philippines	0	0	0	0	0	0	0	0	2,382
Poland	0	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0	9
Russia	0	0	0	1	0	1	0	0	3,319
Saudi Arabia	0	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	0	0	7,417
Slovenia	0	0	0	0	0	0	0	0	81
South Africa	0	0	0	0	0	0	0	0	0
South Korea	0	0	0	0	0	0	0	0	232
Spain	0	0	0	11	0	11	9,007	4,004	10,693
Taiwan	257	257	11	162	11	2,129	3,292	1,345	4,606
Thailand	174	176	15	350	15	2,338	4,145	1,994	5,628
Ukraine	718	728	136	174	144	835	2,766	1,486	4,458
Count of Predictable Currencies	7	7	4	9	4	9	11	8	28

This table reports the number of significant outperforming trading rules in 21,195 TTRs based on eight MHT control approaches. The performance appraisal is the mean of MT returns. The MT return is computed as the total return minus the tilt component. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). Bonferroni and Holm are the results of the fundamental FWER control approaches of Bonferroni (1936) and Holm (1979), respectively. Step-RC refers to the stepwise reality check of Romano and Wolf (2005). Step-SPA represents the stepwise superiority predictive ability test documented by Hsu et al. (2010). Both stepwise procedures also control FWER. FDP-RC/SPA refers to the BRC/SPA tests controlling FDP, where $\gamma = 5\%$. BH and BY are the FDR control approaches of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) where $\delta = 5\%$. Following Chordia et al. (2020), we set the level of significance, α, at 5% for each control approach. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

to 34 after applying the Step-SPA. The gap between the results of the FDP-RC and the FPD-SPA rises to 3,014 when controlling the FDP.

The weaker test power of the RC-based tests, in comparison with that of the SPA-based tests could be due to the RC algorithm assuming LFC for each trading rule, in which $E(\bar{f}_k) = 0$, as we previously outlined in Section 2. As in the calculations shown in Section 2.1.3, the LFC damages the test power of RC tests when there is a large number of poor-performing trading rules in the model sample. After applying the re-centring, the critical value reduces, and therefore the SPA test normally uncovers more significant rejections than the RC test.[11]

Based on the evidence of these four tables, we find that the RC-based tests are more sensitive to specification of different performance metrics than the SPA-based tests. For instance, in Table 5 and 6, the RC-based tests uncover four *'Predictable Currencies'*. However, when evaluating the SR, the Step-RC and FDP-RC tests uncover six (based on total returns) or seven (based on MT returns) *'Predictable Currencies'* (see Tables 7 and 8). The difference in $\#profitable$ further supports this, with Table 5 reporting that the average $\#profitable$ is 12 and 15 for Step-RC and FDP-RC, respectively. The average *#risk-adjusted* of the Step-RC and FDP-RC increases to 47 and 390, respectively, when specifying the Sharpe ratio, as shown in Table 7. Similar changes also appear when assessing MT returns, with the average *#risk-adjusted* of the Step-RC (FDP-RC) increasing from 15 (21) to 55 (410). By contrast, the SPA-based tests exhibit similar test power across the different performance metrics. As shown in Tables 5 and 7, the $\#profitable$ and *#risk-adjusted* of the Step-SPA (FDP-SPA) are 45 (60) and 418 (499), respectively. Based on these two risk-adjusted metrics, the Step-RC test uncovers nine and 42 outperforming TTRs, which might entice investment in Japanese Yen trading opportunities. However, in direct conflict with that conclusion, using the returns as a performance metric leads to the Japanese Yen failing to produce any significant profits (Tables 5 and 6).

The empirical results outlined above highlight the divergence in conclusions based on adopting different

[11] Hansen (2005) also argues that the RC test does not compare the standardised performance measure with the standardisation typically improving the test power as it assesses the performances per given unit of volatility. Romano et al. (2007) agree with this proposition and also proceed by applying the standardised performance metrics in the FDP-RC. We follow the implementation of Hansen (2005) and Romano et al. (2007) to employ the standardised performance metrics, so the increased test power of the SPA-based tests is due to LFC.

Table 7: The number of outperforming technical trading rules based on the SR of total returns (*#risk-adjusted*)

Currencies	MHT control approaches								SHT
	FWER				FDP		FDR		
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Developed									
Australia	0	0	0	0	0	0	0	0	0
Belgium	0	0	162	259	4,765	4,309	6,290	1,899	8,021
Canada	0	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	0	0	4,626
Euro Area	0	0	0	0	0	0	0	0	4
France	0	0	0	11	0	11	0	0	4,077
Germany	0	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	5,667	1,035	8,235
Japan	0	0	9	34	9	3,149	5,456	0	7,628
Netherlands	0	0	0	0	0	0	0	0	3
New Zealand	0	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0	12
Sweden	0	0	0	0	0	0	0	0	2,361
Switzerland	0	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0	0
Others									
Austria	0	0	0	0	0	0	0	0	4,252
Brazil	0	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0	1,014
Egypt	0	0	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0	0	0	67
Greece	0	0	0	0	0	0	0	0	165
Hong Kong	0	0	0	0	0	0	0	0	0
Hungary	0	0	0	0	0	0	0	0	64
India	0	0	0	6	0	6	2,597	0	4,267
Indonesia	2,248	2,299	828	1,070	8,811	10,231	10,194	7,586	10,828
Ireland	0	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0	0
Malaysia	119	119	0	0	0	0	779	307	1,584
Mexico	0	0	0	0	0	0	0	0	24
Philippines	0	0	0	0	0	0	0	0	1,878
Poland	0	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0	3
Russia	0	0	0	0	0	0	0	0	3,228
Saudi Arabia	0	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	4,039	0	7,750
Slovenia	0	0	0	0	0	0	0	0	81
South Africa	0	0	0	0	0	0	0	0	0
South Korea	0	0	0	0	0	0	0	0	243
Spain	0	0	0	0	0	0	6,791	818	8,637
Taiwan	363	363	558	629	1,976	2,737	3,345	1,508	4,616
Thailand	138	138	530	730	2,211	2,665	3,935	1,676	5,511
Ukraine	1,500	1,512	181	181	975	881	4,421	3,029	5,295
Count of Predictable Currencies	5	5	6	8	6	8	11	8	28

This table reports the number of significant outperforming trading rules in 21,195 TTRs based on eight MHT control approaches. The performance appraisal is the SR of total return. The total return is equivalent to the daily excess return of each currency minus the transaction cost. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). Bonferroni and Holm are the results of the fundamental FWER control approaches of Bonferroni (1936) and Holm (1979), respectively. Step-RC refers to the stepwise reality check of Romano and Wolf (2005). Step-SPA represents the stepwise superiority predictive ability test documented by Hsu et al. (2010). Both stepwise procedures also control FWER. FDP-RC/SPA refers to the RC/SPA tests controlling FDP, where $\gamma = 5\%$. BH and BY are the FDR control approaches of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) where $\delta = 5\%$. Following Chordia et al. (2020), we set the level of significance, α, at 5% for each control approach. *Predictable Currencies* represents the number of currencies with at least one significant outperforming TTR

Table 8: The number of outperforming technical trading rules based on the Sharpe ratio of MT returns (*#risk-adjusted*)

Currencies	MHT control approaches								SHT
	FWER				FDP		FDR		
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Developed									
Australia	0	0	0	0	0	0	0	0	0
Belgium	10	10	469	567	5,268	5,178	7,733	3,805	8,380
Canada	0	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	6,004	0	5,167
Euro Area	0	0	0	0	0	0	0	0	4
France	0	0	10	33	10	368	5,106	1,057	4,884
Germany	0	0	0	0	0	0	0	0	0
Italy	86	87	0	0	0	0	9,668	3,249	10,850
Japan	0	0	42	96	2,298	4,370	6,308	1,268	8,332
Netherlands	0	0	0	0	0	0	702	0	5
New Zealand	0	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0	17
Sweden	0	0	0	0	0	0	2,769	0	2,823
Switzerland	0	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0	0
Others									
Austria	0	0	0	0	0	0	0	0	4,808
Brazil	0	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0	1,089
Egypt	0	0	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0	0	0	78
Greece	0	0	0	0	0	0	0	0	175
Hong Kong	0	0	0	0	0	0	0	0	0
Hungary	0	0	0	0	0	0	0	0	68
India	0	0	0	11	0	11	3,074	0	4,449
Indonesia	2,847	2,935	957	1,064	8,516	10,893	10,528	8,323	11,026
Ireland	0	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0	0
Malaysia	66	66	0	0	0	0	345	188	1,278
Mexico	0	0	0	0	0	0	0	0	24
Philippines	0	0	0	0	0	0	0	0	2,306
Poland	0	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0	10
Russia	0	0	0	0	0	0	0	0	3,253
Saudi Arabia	0	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	0	0	7,395
Slovenia	0	0	0	0	0	0	0	0	82
South Africa	0	0	0	0	0	0	0	0	0
South Korea	0	0	0	0	0	0	0	0	242
Spain	76	76	0	132	0	8,186	9,912	4,593	11,262
Taiwan	373	373	471	694	1,418	2,616	3,394	1,546	4,654
Thailand	146	146	577	732	2,067	2,821	4,013	1,752	5,583
Ukraine	823	826	128	142	141	155	3,159	1,722	4,607
Count of Predictable Currencies	8	8	7	9	7	9	14	10	28

This table reports the number of significant outperforming trading rules in 21,195 TTRs based on eight MHT control approaches. The performance appraisal is the SR of MT returns. The MT return is computed as the total return minus the tilt component. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). Bonferroni and Holm are the results of the fundamental FWER control approaches of Bonferroni (1936) and Holm (1979), respectively. Step-RC refers to the stepwise reality check of Romano and Wolf (2005). Step-SPA represents the stepwise superiority predictive ability test documented by Hsu et al. (2010). Both stepwise procedures also control FWER. FDP-RC/SPA refer to the RC/SPA tests controlling FDP, where $\gamma = 5\%$. BH and BY are the FDR control approaches of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) where $\delta = 5\%$. Following Chordia et al. (2020), we set the level of significance, α, at 5% for each control approach. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

MHT approaches. We posit that this divergence could potentially give rise to what we term 'MHT-hacking'. The concept is similar to the p-hacking bias, with Chen (2021) demonstrating p-hacking as a phenomenon whereby seemingly 'significant' financial anomalies come from collecting only 'notable' results from a giant factor pool, rather than based on fundamental academic theory. While we acknowledge that MHT-hacking may not be widespread as of yet, similar to SHT users, MHT users could choose to exclusively adopt an MHT approach that fits their demands (i.e., MHT-hacking). For example, when an academic work seeks to uncover if a given currency is highly predictable or if a group of trading rules outperform, the scholar may adopt MHT approaches to show robustness against p-hacking, the multiple comparisons problem or data-snooping issues. However, given the increasing prevalence of MHT innovations in recent years, there is now a potential for scholars to apply various MHT approaches to seemingly address these issues, however to subsequently exclusively report select MHT results that support their propositions, in order to increase their publication chances.

Asymmetric information and conflicts of interest may also provide motivation for practitioners to engage in MHT-hacking. Similar to the asymmetric information paradigm of Glosten and Milgrom (1985) and Kyle (1985), the over-the-counter (OTC) nature of the FX market, leads to information being widely dispersed across different groups of FX market participants (Liu and Wang, 2016, Babus and Kondor, 2018, Ranaldo and Somogyi, 2021). For instance, traders may possess information that asset holders do not. Traders are exclusively concerned with trading performance and portfolio size as these factors directly determine their income. However, asset holders (e.g., commercial banks) seek to maximise their preferences by striking a balance between trading performance (Grammatikos et al., 1986), risk hedging (Chamberlain et al., 1997, Allayannis and Ofek, 2001), inventory control (Westerhoff, 2003), and price competition with other market-markers (Dennert, 1993, Calcagno and Lovo, 2006).

As a result some potential stylised MHT-hacking examples that might arise, include a trader putting forward select MHT evidence in order to be assigned a greater trading budget, in the hope of ultimately securing a higher salary bonus. This could be in contrast with optimal risk management practice. Another potential conflict could be if a broker also provides advice to clients, they may be incentivised to show MHT evidence that highlights the benefits of strategies that trade currencies with a higher turnover or broker commission. A final potential example could be where an institution has concentrated exposure to a particular currency, meaning that an in-house trader might then be incentivised to bring forward select MHT evidence that would appear to reduce that exposure, in the hope that they receive institutional sign-off to trade.

Overall, our comparison first emphasises the need to adopt MHT instead of SHT approaches when assessing a large number of TTRs. Second, the diverging results across various MHT approaches and the sensitivities to performance criteria are two results that could leave such an analysis vulnerable to MHT-hacking in practice. For the two MHT approaches controlling the FDP, FDP-SPA is less sensitive to the specific performance criteria adopted and discovers more significant rejections than FDP-RC. For these reasons we view the SPA as a more desirable algorithm relative to the RC when controlling the FDP. Finally, our results indicate that controlling the FWER is too stringent. We now highlight the patterns that differ from those reported for equity markets by Chordia et al. (2020) and Harvey et al. (2020). First, when trading currencies, the traditional FWER approaches, Bonferroni and Holm, are not the most strict control approaches in terms of the lowest number of predictable currencies. For example, in Tables 5 and 8, the Step-RC/SPA tests find fewer predictable currencies than the Bonferroni or Holm. Second, for the FDP controls, Chordia et al. (2020) and Harvey et al. (2020) demonstrate that the test power of the FDP approaches lies between that of BH and BY. However, according to our *#profitable* and *#risk-adjusted* results, we see that the FDP-SPA indeed follows this conclusion, however the FDP-RC does not.

6. Out-of-Sample Evaluation

Our above results depend on the entire sample horizon and can be affected by the look-ahead bias. Also, the *#profitable* and *#risk-adjusted* results may not fully reflect the benefits of adopting a particular MHT approach in practice as the measures do not illustrate portfolio performance. More importantly, investing in the in-sample *#profitable* and *#risk-adjusted* trading rules can not guarantee significant outperformance over out-of-sample periods. To achieve more robust findings, we assess various MHT approaches in practice by examining whether the uncovered in-sample outperforming models indeed achieve superior out-of-sample (OOS) results (Lo and MacKinlay, 1990, Sullivan et al., 1999, Harvey et al., 2016, Hsu et al., 2016). We divide the entire sample horizon into two sub-periods: the out-of-sample period ranges from 3/5/2008 to 3/18/2019 (25% observations), with the remaining data referring to the in-sample period (75% observations), from 1/5/1976 to 3/4/2008.

6.1. The average out-of-sample survival rate

FX market participants apply MHT approaches to explore profitable trading strategies. They might expect outperforming trading rules uncovered using historical samples to also outperform in the future. To assess if this holds in practice, we evaluate various MHT approaches out-of-sample. One intuitive method is to count how many in-sample outperforming trading rules retain their outperforming status during the out-of-sample period. To this end, we introduce a criteria called the MHT survival rate. First, a trading rule is defined as a survivor if it shows significant outperformance in-sample and also achieves superior profitability out-of-sample. Then, we compute the survival rate for a given hypothesis testing approach, as:

$$Survival \quad rate = \frac{Num_{OOS}}{Num_{IS}}, \tag{22}$$

where Num_{OOS} represents the number of survivors, and Num_{IS} refers to the number of in-sample instances of significance. This measurement reflects the probability that a given testing approach achieves its in-sample performance in the OOS setting. The higher the survival rate, the more in-sample outperforming trading rules also outperform out-of-sample. An MHT approach with a higher survival rate is therefore most desirable. [12]

Furthermore, we illustrate the cross currency evaluation via the average survival rate, as:

$$Average \quad survival \quad rate = \frac{\sum_{c=1}^{NC} Num_{OOS}^{c}}{\sum_{c=1}^{NC} Num_{IS}^{c}}, \tag{23}$$

where NC represents the number of currencies with in-sample instances of significance based on the results of a specific MHT approach. Compared to directly averaging the survival rates of different currencies, this measurement alleviates the impact of variation in the number of significant rejections across currencies. Table

[12] We define a given trading rule as an OOS outperforming trading rule if its t-value of total return exceeds 1.96 over the OOS period. This constant threshold may cause the Num_{OOS} to be higher than the number of true OOS significant rejections.

Table 9: OOS survivors

Metric	Control approaches								SHT
	FWER				FDP		FDR		
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Panel A: average OOS survival rates									
Total return	29.9%	30.2%	12.3%	23.3%	34.1%	36.8%	30.5%	32.3%	22.3%
MT return	40.3%	40.7%	18.8%	27.9%	42.2%	37.3%	28.1%	33.9%	21.1%
Total SR	38.7%	39.1%	22.0%	25.7%	44.0%	44.5%	31.4%	32.7%	22.3%
MT SR	45.6%	45.8%	19.5%	23.6%	40.4%	44.5%	27.6%	33.3%	21.2%
Mean	38.6%	39.0%	18.2%	25.1%	40.2%	40.8%	29.4%	33.1%	21.7%
Panel B: Number of significant rules during the in-sample period (Num^c_{IS})									
Total return	1,337	1,345	545	1,141	807	13,502	44,451	13,435	95,538
MT return	1,640	1,658	1,386	1,804	15,081	19,605	44,847	14,395	95,876
Total SR	1,633	1,660	590	1,406	969	26,982	54,683	21,181	10,4892
MT return	2,199	2,233	1,615	2,195	18,148	30,410	56,504	22,083	10,5507
Mean	1,702.25	1,724	1,034	1,636.5	8,751.25	22,624.75	50,121.25	17,773.5	10,0453.3

This table reports the average survival rate computed as Equation 23, $\frac{\sum_{c=1}^{NC} Num^c_{OOS}}{\sum_{c=1}^{NC} Num^c_{IS}}$, where Num_{OOS} represents the number of survivors, Num_{IS} refers to the number of in-sample instances of significance, and NC represents the number of currencies with in-sample instances of significance based on the results of a specific MHT approach. The first column reports the performance metric used to identify the in-sample outperforming TTRS. The last row presents the mean of rates across the four metrics. MT return refers to 'market timing' profit derived from timing the trades being placed in the market, as opposed to the 'tilting' component, which is derived from the average FX position held over the full sample period. SR refers to the risk-adjusted Sharpe ratio measure.

9 summarises average survival rates across four performance metrics, where we include only currencies with at least one significant in-sample rejection. The survival rate for each currency calculated using Equation 22 is detailed in Appendix D.

First, the average survival rates emphasise that adopting MHT is essential to FX investors. Average survival rates of the SHT range from 21.1% to 22.3% across the four metrics. In other words, over 87% of the in-sample outperforming trading rules identified by the SHT fail to yield significant out-of-sample profits. As mentioned above, FX market practitioners apply MHT approaches to uncover profitable trading rules from a large number of candidates. Upon finding in-sample outperforming ones, investors expect the selected ones to yield significant profits out-of-sample. However, the survival rates highlight the inability of SHT to achieve this, which supports the need to apply MHT approaches. Second, the average survival rates highlight the importance of identifying a suitable MHT approach. Wrongly selecting an MHT approach could result in an average survival rate that is even lower than the SHT. For instance, despite the Step-RC stringently controlling Type I errors, only 12.3% to 22.0% of the uncovered in-sample instances of significance achieve outperformance in the OOS setting. Third, across the eight MHT approaches, Table 9 highlights the superiority of adopting FDP-SPA. It reports the highest mean rate at 40.8%, with average survival rates of 44.5% for both Total and MT Sharpe ratio. In comparison with FDP-RC, the best MHT approach identified by Chordia et al. (2020), the average survival rate of FDP-SPA exceeds that of FDP-RC across three of the four performance metrics, with MT return being the outlier. Furthermore, upon closer inspection we observe that the FDP-RC has the additional disadvantage of having its average MT return survival rate base being determined exclusively by the Indian Rupee. Please see Appendix D for further details.

6.2. OOS portfolios

Besides the survival rate analysis above, we next construct OOS portfolios based on MHT approaches to provide evidence from the portfolio aspect. In line with Hsu et al. (2016), we use in-sample results of MHT approaches to construct OOS portfolios and assess MHT approaches based on the performance of these OOS portfolio. We invest in every instance of significance uncovered by the MHT approach to construct our OOS portfolios. Total return is used as our standard performance metric to compare the various MHT approaches.[13]

Table 10 reports total return summary statistics of our OOS portfolios. We see that OOS portfolio performance reconciles with our findings using average survival rate. First we observe that the performance summary results also support the view that FDP-SPA is the preferable MHT algorithm to adopt, with the FDP-SPA based portfolio yielding the highest return (4.20%), the highest maximum monthly return (3.61%), and the best Sharpe ratio (1.270). Second, we see that the portfolio based on Step-RC fails to produce significant profits at 0.40% per annum with a t-value of only 0.298. Also, it reports the maximum drawdown at 20.32%.

The cumulative performance further supports the view the FX market practitioners should adopt FDP-SPA. This is further highlighted in Figure 1, which plots the OOS cumulative net profits for each MHT determined portfolio. The net profits of the FDP-SPA based portfolio (light-blue dotted line) reach a peak at 0.36 at the end of 2015, after which the cumulative net profits reduce but still represent the maximum profitability at the end of the OOS period. Interestingly, the BY based portfolio yields similar performance but is much more volatile than the FDP-SPA based portfolio, consistent with higher volatility reported in

[13] Hsu et al. (2016) exclusively invest in the best-performing trading rules uncovered by the different MHT approaches. However, we argue that such a method excludes crucial information uncovered by the MHT approaches, such as the number of instances of significance, and fails to differentiate appropriately between competing MHT approaches. Therefore, we build our OOS portfolios using all in-sample instances of significance.

Table 10: Performance summary of the OOS portfolios

	FWER				FDP		FDR	
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY
Mean	2.44%	2.48%	0.40%	2.11%	2.49%	4.20%	2.44%	2.48%
t-value	2.807	2.844	0.298	2.825	2.816	3.564	2.807	2.844
Vol	0.024	0.024	0.037	0.021	0.025	0.033	0.024	0.024
SR	1.000	1.013	0.106	1.006	1.003	1.270	1.000	1.013
Max	1.28%	1.29%	1.22%	0.99%	1.26%	3.61%	1.28%	1.29%
Min	-0.75%	-0.76%	-4.69%	-1.16%	-1.18%	-2.77%	-0.75%	-0.76%
Cumulative	0.209	0.213	0.031	0.166	0.196	0.331	0.209	0.213
Drawdown	3.94%	3.93%	20.32%	5.23%	5.07%	4.67%	3.94%	3.93%
Skewness	0.032	0.032	-0.317	-0.010	0.005	0.125	0.032	0.032
Kurtosis	0.027	0.027	0.360	0.038	0.035	0.268	0.027	0.027

This table reports the summary of OOS portfolio performance based on various MHT approaches. In-sample results of MHT approaches are used to construct OOS portfolios and the OOS peformance of these portfolios are presented. Each instance of significance uncovered by the MHT approaches is used to construct the OOS portfolios shown. Five statistics are presented at annual level, including the mean, volatility (Vol), Sharpe ratio (SR), skewness (Skew), and kurtosis (Kurt). Max/Min refers to the max/min daily return. Cum presents the cumulative net profits over the sample period, with Drawdown reporting the semi-annual maximum drawdown.

Table 10.

The superiority of the FDP-SPA based portfolio partly stems from better diversification across trading rules. Diversification across various currencies is desirable, however, a higher number of in-sample instances of significance might also be of benefit to an investor as it allows FX market participants to further diversify risks by investing across various trading rules. An example can be observed in the comparison between the Step-RC and FDP-RC portfolios, which both exclusively invest in Indonesian Rupiah. According to our in-sample results, the Step-RC uncovers 545 outperforming trading rules, whereas the FDP-RC finds 807, 48% more. The OOS portfolio of the Step-RC suffers from negative cumulative net profits before 2014 and yields the lowest cumulative net profits of all the MHT approaches at the end of the OOS period. The crash in 2008/2009 is the main driver of the poor performance of the Step-RC portfolio. In contrast, although the FDP-RC based OOS portfolio also produces negative returns at the beginning of the OOS period, it rebounds much sooner and is not subject to such extreme drawdowns. The performance difference between these two portfolios provides initial evidence highlighting the benefits of trading rule diversification brought about through a higher number of instances of significance uncovered by a given MHT approach.

Overall, both the survival rate analysis and the portfolio evaluation emphasise the importance of applying the appropriate MHT approach. We identify the FDP-SPA of Hsu et al. (2014) as the optimal algorithm for FX market practitioners. Although MHT approaches are regarded as efficient methods to control Type 1 errors, we highlight how their results may not always benefit an investor's out-of-sample wealth, for example, the Step-RC. Moreover, in line with our in-sample conclusions, the OOS analysis reconciles our findings that

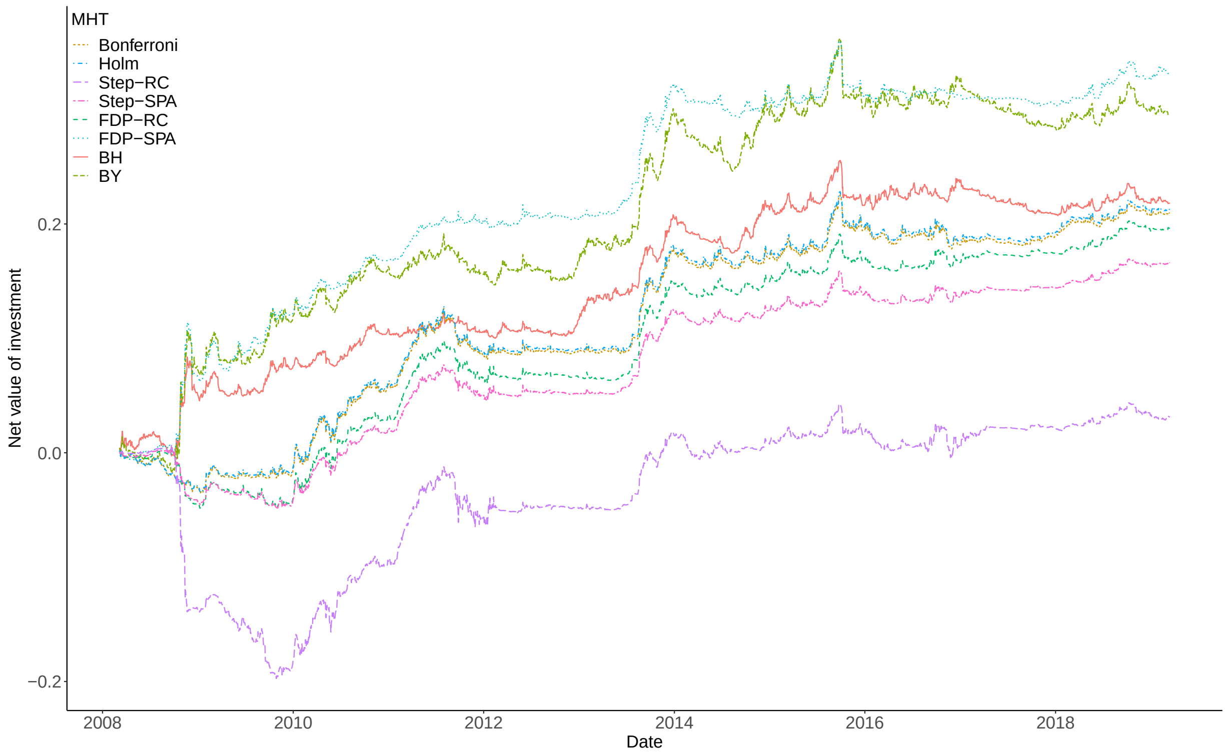

Figure 1: OOS cumulative net profits across eight MHT based portfolios

the FDP-SPA test is the optimal MHT algorithm in the FX market.

7. Conclusion

Our research contributes to the literature by examining the test powers of various MHT control approaches in the FX market. Our comprehensive comparisons allow us to draw up a guide for applying MHT when seeking truly profitable FX trading rules. Our work complements the findings of previous studies across three dimensions.

First, we uncover the potential for 'MHT-hacking' by showing diverging MHT results. The concept is similar to the p-hacking bias (see, for example Chen (2021)); p-hacking is a phenomenon whereby seemingly 'significant' financial anomalies come from presenting only 'notable' results from a giant factor pool. While we acknowledge that MHT-hacking may not be widespread as of yet, similar to SHT users, MHT users could choose to exclusively adopt an MHT approach that fits their demands. There is a potential for this to impact academics and practitioners, with researchers motivated to present only notable results to enhance their work's publication potential, and different groups of FX market participants experiencing preference conflicts due to information asymmetry (Liu and Wang, 2016, Babus and Kondor, 2018, Ranaldo and Somogyi, 2021). Second, we highlight the importance of adopting out-of-sample evaluation. Market participants can never know the realised FDP using an in-sample testing sample. By contrast, our out-of-sample assessment can illustrate how many in-sample outperforming rules achieve the same superiority over the out-of-sample period and which MHT results can produce the best portfolio performance. Finally, our empirical results find that the FDP-SPA of Hsu et al. (2014) has the advantage of uncovering more truly significant outperforming trading rules than the other MHT approaches.

This study suggests new interpretations of the findings of Chordia et al. (2020). We agree that controlling the FDP is the most appropriate controlling method for MHT. However, the FDP-RC test is not the optimal algorithm as it is sensitive to performance metrics, whereas SPA tests do not suffer from this downside. More importantly, the FDP-SPA approach reports higher survival rates and produces the best portfolio performance in the out-of-sample evaluation. Our work, combined with the equity market insights provided by Chordia et al. (2020), shows that a given MHT approach can perform differently across various asset classes. Therefore, we appeal to future researchers to assess the test power and practical implications of applying various MHT approaches in other asset classes.

References

Akram, Q. F., Rime, D. and Sarno, L. (2008), 'Arbitrage in the foreign exchange market: Turning on the microscope', *Journal of International Economics* **76**(2), 237–253.

Alexander, S. S. (1961), 'Price movements in speculative markets: Trends or random walks', *Industrial Management Review (pre-1986)* **2**(2), 7.

Allayannis, G. and Ofek, E. (2001), 'Exchange rate exposure, hedging, and the use of foreign currency derivatives', *Journal of International Money and Finance* **20**(2), 273–296.

Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013), 'Value and momentum everywhere', *The Journal of Finance* **68**(3), 929–985.

Babus, A. and Kondor, P. (2018), 'Trading and information diffusion in over-the-counter markets', *Econometrica* **86**(5), 1727–1769.

Bailey, D. H. and Lopez de Prado, M. (2012), 'The sharpe ratio efficient frontier', *Journal of Risk* **15**(2), 13.

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* pp. 1165–1188.

Bonferroni, C. (1936), 'Teoria statistica delle classi e calcolo delle probabilita', *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* **8**, 3–62.

Boons, M. and Prado, M. P. (2019), 'Basis-momentum', *The Journal of Finance* **74**(1), 239–279.

Burnside, C., Eichenbaum, M., Kleshchelski, I. and Rebelo, S. (2010), 'Do peso problems explain the returns to the carry trade?', *The Review of Financial Studies* **24**(3), 853–891.

Burnside, C., Eichenbaum, M. and Rebelo, S. (2007), 'The returns to currency speculation in emerging markets', *American Economic Review* **97**(2), 333–338.

Calcagno, R. and Lovo, S. (2006), 'Bid-ask price competition with asymmetric information between market-makers', *The Review of Economic Studies* **73**(2), 329–355.

Chamberlain, S., Howe, J. S. and Popper, H. (1997), 'The exchange rate exposure of us and japanese banking institutions', *Journal of Banking & Finance* **21**(6), 871–892.

Chen, A. Y. (2021), 'The limits of p-hacking: Some thought experiments', *The Journal of Finance* **76**(5), 2447–2480.

Chordia, T., Goyal, A. and Saretto, A. (2020), 'Anomalies and false rejections', *The Review of Financial Studies* **33**(5), 2134–2179.

Cornell, B. (1982), 'Money supply announcements, interest rates, and foreign exchange', *Journal of International Money and Finance* **1**, 201–208.

Cornell, W. B. and Dietrich, J. K. (1978), 'The efficiency of the market for foreign exchange under floating exchange rates', *The Review of Economics and Statistics* pp. 111–120.

Dahlquist, M. and Hasseltoft, H. (2020), 'Economic momentum and currency returns', *Journal of Financial Economics* **136**(1), 152–167.

Dennert, J. (1993), 'Price competition between market makers', *The Review of Economic Studies* **60**(3), 735–751.

Eatwell, J., Milgate, M. and Newman, P. (1992), *The New Palgrave Dictionary of Money and Finance*, Springer.

Fama, E. F. (1984), 'Forward and spot exchange rates', *Journal of Monetary Economics* **14**(3), 319–338.

Filippou, I., Gozluklu, A. E. and Taylor, M. P. (2018), 'Global political risk and currency momentum', *Journal of Financial and Quantitative Analysis* **53**(5), 2227–2259.

Froot, K. A. and Thaler, R. H. (1990), 'Anomalies: Foreign exchange', *Journal of Economic Perspectives* **4**(3), 179–192.

Glosten, L. R. and Milgrom, P. R. (1985), 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders', *Journal of Financial Economics* **14**(1), 71–100.

Grammatikos, T., Saunders, A. and Swary, I. (1986), 'Returns and risks of us bank foreign currency activities', *The Journal of Finance* **41**(3), 671–682.

Gu, J. and Shen, S. (2018), 'Oracle and adaptive false discovery rate controlling methods for one-sided testing: Theory and application in treatment effect evaluation', *The Econometrics Journal* **21**(1), 11–35.

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

Hansen, P. R., Lunde, A. and Nason, J. M. (2011), 'The model confidence set', *Econometrica* **79**(2), 453–497.

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

Harvey, C. R. and Liu, Y. (2020), 'False (and missed) discoveries in financial economics', *The Journal of Finance* **75**(5), 2503–2553.

Harvey, C. R., Liu, Y. and Saretto, A. (2020), 'An evaluation of alternative multiple testing methods for finance applications', *The Review of Asset Pricing Studies* **10**(2), 199–248.

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

Hochberg, Y. and Tamhane, A. C. (1987), *Multiple Comparison Procedures*, John Wiley & Sons, Inc.

Hodrick, R. J. and Srivastava, S. (1986), 'The covariation of risk premiums and expected future spot exchange rates', *Journal of International Money and Finance* **5**, S5–S21.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* pp. 65–70.

Hsu, P.-H., Han, Q., Wu, W. and Cao, Z. (2018), 'Asset allocation strategies, data snooping, and the 1/n rule', *Journal of Banking & Finance* **97**, 257–269.

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. and Kuan, C.-M. (2005), 'Reexamining the profitability of technical analysis with data snooping checks', *Journal of Financial Econometrics* **3**(4), 606–628.

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.

Hsu, Y.-C., Kuan, C.-M. and Yen, M.-F. (2014), 'A generalized stepwise procedure with improved power for multiple inequalities testing', *Journal of Financial Econometrics* **12**(4), 730–755.

Hudson, R. and Urquhart, A. (2021), 'Technical trading and cryptocurrencies', *Annals of Operations Research* **297**(1), 191–220.

Kearney, F., Cummins, M. and Murphy, F. (2014), 'Outperformance in exchange-traded fund pricing deviations: Generalized control of data snooping bias', *Journal of Financial Markets* **19**, 86–109.

Kim, A. Y., Tse, Y. and Wald, J. K. (2016), 'Time series momentum and volatility scaling', *Journal of Financial Markets* **30**, 103–124.

Kyle, A. S. (1985), 'Continuous auctions and insider trading', *Econometrica* pp. 1315–1335.

Lequeux, P. and Acar, E. (1998), 'A dynamic index for managed currencies funds using cme currency contracts', *The European Journal of Finance* **4**(4), 311–330.

Levy, R. A. (1967), 'Relative strength as a criterion for investment selection', *The Journal of Finance* **22**(4), 595–610.

Liu, H. and Wang, Y. (2016), 'Market making with asymmetric information and inventory risk', *Journal of Economic Theory* **163**, 73–109.

Lo, A. W. and MacKinlay, A. C. (1990), 'Data-snooping biases in tests of financial asset pricing models', *The Review of Financial Studies* **3**(3), 431–467.

Lustig, H., Roussanov, N. and Verdelhan, A. (2011), 'Common risk factors in currency markets', *The Review of Financial Studies* **24**(11), 3731–3777.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–91.

Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A. (2012*a*), 'Carry trades and global foreign exchange volatility', *The Journal of Finance* **67**(2), 681–718.

Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A. (2012*b*), 'Currency momentum strategies', *Journal of Financial Economics* **106**(3), 660–684.

Menkhoff, L. and Taylor, M. P. (2007), 'The obstinate passion of foreign exchange professionals: technical analysis', *Journal of Economic Literature* **45**(4), 936–972.

Moskowitz, T. J., Ooi, Y. H. and Pedersen, L. H. (2012), 'Time series momentum', *Journal of Financial Economics* **104**(2), 228–250.

Neely, C. J. and Weller, P. A. (2013), 'Lessons from the evolution of foreign exchange trading strategies', *Journal of Banking & Finance* **37**(10), 3783–3798.

Neuhierl, A. and Schlusche, B. (2011), 'Data snooping and market-timing rule performance', *Journal of Financial Econometrics* **9**(3), 550–587.

Papailias, F., Liu, J. and Thomakos, D. D. (2021), 'Return signal momentum', *Journal of Banking & Finance* **124**, 106063.

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

Psaradellis, I., Laws, J., Pantelous, A. A. and Sermpinis, G. (2019), 'Performance of technical trading rules: evidence from the crude oil market', *The European Journal of Finance* **25**(17), 1793–1815.

Qi, M. and Wu, Y. (2006), 'Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market', *Journal of Money, Credit and Banking* pp. 2135–2158.

Ramadorai, T. (2008), 'What determines transaction costs in foreign exchange markets?', *International Journal of Finance & Economics* **13**(1), 14–25.

Ranaldo, A. and Somogyi, F. (2021), 'Asymmetric information risk in FX markets', *Journal of Financial Economics* **140**(2), 391–411.

Romano, J. P., Shaikh, A. M. and Wolf, M. (2008), 'Formalized data snooping based on generalized error rates', *Econometric Theory* pp. 404–447.

Romano, J. P., Shaikh, A. M. and Wolf, M. (2010), 'Hypothesis testing in econometrics', *Annu. Rev. Econ.* **2**(1), 75–104.

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

Romano, J. P., Wolf, M. et al. (2007), 'Control of generalized error rates in multiple testing', *The Annals of Statistics* **35**(4), 1378–1408.

Sager, M. J. and Taylor, M. P. (2006), 'Under the microscope: The structure of the foreign exchange market', *International Journal of Finance & Economics* **11**(1), 81–95.

Sharpe, W. F. (1966), 'Mutual fund performance', *The Journal of Business* **39**(1), 119–138.

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

Sweeney, R. J. (1986), 'Beating the foreign exchange market', *The Journal of Finance* **41**(1), 163–182.

Taylor, M. P. (1989), 'Covered interest arbitrage and market turbulence', *The Economic Journal* **99**(396), 376–391.

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

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

Westerhoff, F. H. (2003), 'Market-maker, inventory control and foreign exchange dynamics', *Quantitative Finance* **3**(5), 363.

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

Xu, Q. and Wang, Y. (2021), 'Managing volatility in commodity momentum', *Journal of Futures Markets* **41**(5), 758–782.

Appendix A. Details of technical trading rules

Appendix A.1. Oscillator trading rules

Oscillator rules determine the trading positions in terms of the *Relative Strength Indicators* (RSI) introduced by Levy (1967). If RSI of a given currency over the previous h days exceeds $50 + v$ for d days and subsequently goes below this level, we short this currency. If RSI over the previous h days moves below $50+v$ and then exceeds this level, we go long this currency. We further separate this family into two sub-groups.

$O1$: *after a trading signal is generated, hold the trading position until the next trading signal is generated.*

$O2$: *after a trading signal is generated, hold the trading position for k days and then neutralise the positions until the next trading signal is generated.*[14]

By changing the parameters h, v, d, k, we are left with 120 $O1$ trading rules and 480 $O2$ trading rules.

Appendix A.2. Filter trading rules

Filter rules are simple technical trading rules firstly introduced by Alexander (1961). In practice, once the exchange rate rises to at least x percent greater than the most recent low and remains for $d(x)$ days, long the currency; if the exchange rate moves below at least x percent lower than the most recent high and remains for $d(x)$ days, short the currency. The most recent high (low) equals the most recent daily end spot rate that is larger (less) than the j previous daily end rates. Furthermore, we classify this family into three sub-categories:

$F1$: *after a trading signal is generated, hold the trading position until the next trading signal is generated.*

$F2$: *after a buying (shorting) trading signal is generated, hold the trading position until the exchange rate reduces (increases) to at least y percent lower (greater) than the highest (lowest) exchange rate over the previous j days, at which point we neutralise the position.*

$F3$: *after a trading signal is generated, hold the trading positions for k days, then neutralise the position until the next trading signal occurs.*

By changing the parameters $x, y, d, d(x), d(y), j, k$, we build 210 $F1$ trading rules, 1,575 $F2$ trading rules, and 1,050 $F3$ trading rules.

[14] For any constant holding period greater than one day, we ignore the trading signals generated during the holding period, k days, to avoid constructing overlapping portfolios.

Appendix A.3. Moving average trading rules

Moving average trading rule is the most extensively used technical analysis in the FX market. This study computes the basic moving average over the previous q days, $MA(q)$, as the mean of daily exchange rates. Once the short-term moving average, $MA(p)$,[15] of a currency moves to at least x percent greater than the long-term moving average, $MA(q)$, for d days, buy this currency; when the short-term moving average, $MA(p)$, moves to at least x percent lower than the long-term moving average, $MA(p)$, for d days, sell this currency. We further break this TTR down into the following sub-categories.

$MA1$: *the trading signals are determined by the comparison between the most recent daily exchange rate,* $MA(1)$, *and* $MA(q)$. *After one trading signal is generated, hold the trading positions until the next trading signal is determined.*

$MA2$: *the trading signals are determined by the comparison between the most recent daily exchange rate,* $MA(1)$, *and* $MA(q)$. *After one trading signal is generated, hold the trading positions for* k *days, then neutralise the position until the next trading signal occurs.*

$MA3$: *the trading signals are determined by the comparison between* $MA(p)$ *and* $MA(q)$. *After one trading signal is generated, hold the trading position until the next trading signal is determined.*

$MA4$: *the trading signals are determined by the comparison between* $MA(p)$ *and* $MA(q)$. *After one trading signal is generated, hold the trading positions for* k *days, then neutralise the position until the next trading signal occurs.*

Apart from these four sub-categories, we further cover another type of complicated moving average trading rules introduced by Lequeux and Acar (1998), and name it $MA5$. The construction is shown as:

$MA5$: *the trading signals are determined by the comparison among the most recent daily end exchange rate,* $MA(1)$, *and three moving averages across periods,* $MA(n)$, $MA(p)$, $MA(q)$, *where* $n < p < q$. *If* $MA(1)$ *is at least* x *percent over any two of these three moving averages for* d *days, take a one-third long position (buy this currency with 1/3 of the entire budget). If* $MA(1)$ *moves above all three moving averages, take a full long position (buy this currency with the entire budget). If* $MA(1)$ *moves at least* x *below any two of three moving averages for* d *days, take a one-third short position (sell this currency with 1/3 of the entire budget). If* $MA(1)$ *moves below all three moving averages, take a full short position (sell this currency with the entire budget). After a trading position is generated, hold it until the next*

[15] We restrict p to be less than or equal to q.

trading signal is determined.

Based on different sets of parameters, q, p, n, x, d, k, we construct 330 $MA1$ trading rules, 990 $MA2$ trading rules, 1,650 $MA3$ trading rules, 4,950 $MA4$ trading rules, and 4,950 $MA5$ trading rules.

Appendix A.4. Support-resistance trading rules

The *Support-resistance trading rule* is based on identifying a level where it is hard for the exchange rate to reduce (support) or increase (resistance) further. This trading rule assumes that once the spot rate breaks through the support or resistance levels, the spot rate moves further in the same direction. More specifically, a long position is taken once the spot rate is at least x percent greater than the highest exchange rate over the previous j days, and remains this way for d days; with a short position being taken if the spot rate is at least x percent lower than the lowest exchange rate over the previous j days, and remains this way for d days. This type of trading rule is similar to the *filter rules*, in which the trading signal is allocated when the daily rate moves above (below) a benchmark that is greater (less) than the resistance (support) level. However, the trading signal of *support-resistance rule* is generated when the exchange rate moves over (below) the resistance (support level). Based on different holding periods, we form two sub-categories in this family as.

$SR1$*: after a trading signal is generated, hold the trading position until the next trading signal is determined.*

$SR2$*: after a trading signal is generated, hold the trading position for k days, then neutralise the position until the next trading signal occurs.*

With different sets of parameters, x, d, j, k, we have 378 $SR1$ trading rules, and 1,512 $SR2$ trading rules.

Appendix A.5. Channel breakout trading rules

The *Channel break out trading rule* attempts to identify the channel of exchange rates that the highest level of the daily exchange rate over the previous j days is in a $c\%$ range of the lowest level over the same period. This trading rule family is similar to the *support-resistance trading rule* but with time dynamic support and resistance levels. The dynamic support and resistance levels under the channel break rule are the lower and upper bounds of the trading channels. The upper bound on a given day is set at a point where the exchange rate is $c\%$ greater than the lowest of previous j days; the lower bound on a given day is where the daily end exchange rate is $c\%$ lower than the highest of the previous j days. Under this trading rule, if a $c\%$ trading channel exists, go long the currency when the spot rate increases to x percent above the upper bound of the channel and remains in this state for d days; go short the currency when the spot rate declines to x percent below the lower bound of the channel and remains in this state for d days. Similar to the support-resistance family, we have two sub-groups in this family as.

$CB1$: after a trading signal is generated, hold the trading position until the next trading signal is determined.

$CB2$: after a trading signal is generated, hold the trading position for k days, then neutralise the position until the next trading signal occurs.

With different sets of parameters, x, d, j, c, k, we have 600 $CB1$ trading rules, and 2,400 $CB2$ trading rules. Overall, the entire universe of technical trading rules consists of 21,195 trading rules across the five families outlined above.

Appendix A.6. Parameters in the technical trading rules

This section presents the parameters used for the technical trading rules highlighted above. For the *oscillator trading rules*, we set h, v, d, k as:

$$h \in \{5, 10, 15, 20, 25, 50, 100, 150, 200, 250\}.$$

$$v \in \{10, 15, 20, 25\}.$$

$$d \in \{1, 2, 5\}.$$

$$k \in \{1, 5, 10, 25\}.$$

For the *filter rules*, we set the parameters $x, y, d(x), d(y), j, k$ as:

$$x \in \{0.05, 0.1, 0.5, 1.0, 5.0, 10.0, 20.0\}.$$

$$y \in \{0.05, 0.1, 0.5, 1.0, 5.0, 10.0, 20.0\}.$$

$$d \in \{0, 1, 2, 3, 4, 5\}.$$

$$d(x) \in \{0, 1, 2, 3, 4, 5\}.$$

$$d(y) \in \{0, 1, 2, 3, 4\}.$$

$$j \in \{1, 2, 5, 10, 20\}.$$

$$k \in \{5, 10, 15, 20, 25\}.$$

For the *moving average trading rule*, we set the parameters $q, y, d(x), d(y), j, k$ as:

$$q \in \{2, 5, 10, 15, 20, 25, 50, 100, 150, 200, 250\}.$$

$$p \in \{2, 5, 10, 15, 20, 25, 50, 100, 150, 200\}.$$

$$n \in \{2, 5, 10, 15, 20, 25, 50, 100, 150\}.$$

$$x \in \{0, 0.05, 0.1, 0.5, 1.0, 5.0\}.$$

$$d \in \{0, 2, 3, 4, 5\}.$$

$$k \in \{5, 10, 25\}.$$

For the *support-resistance trading rule*, we set the parameters $q, y, d(x), d(y), j, k$ as:

$$x \in \{0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10.0\}.$$

$$d \in \{0, 2, 3, 4, 5\}.$$

$$j \in \{2, 5, 10, 15, 20, 25, 50, 100, 250\}.$$

$$k \in \{1, 5, 10, 25\}.$$

For the *channel breakout trading rule*, we set the parameters $q, y, d(x), d(y), j, k$ as:

$$x \in \{0.05, 0.1, 0.5, 1.0, 5.0\}.$$

$$d \in \{0, 1, 2\}.$$

$$j \in \{5, 10, 15, 20, 25, 50, 100, 200\}.$$

$$c \in \{0.1, 0.5, 1.0, 5.0, 10.0\}.$$

$$k \in \{1, 5, 10, 25\}.$$

Appendix B. The generalised Stepwise RC and SPA tests with various K

This section presents the test power of the generalised Step-RC/SPA(K) tests. Tables B.1 to B.6 show the number of significant outperforming TTRs for trading each currency, where $K = 2, 3, 10, 20, 100, 200$, respectively. As K increases, both Stepwise algorithms uncover an increased number of significant rejections, with a greater number of currencies reporting outperforming strategies. For instance, based on the SR of total returns, the Step-SPA(200) test identifies 20 more predictive currencies than the standard Step-SPA, $K = 1$, as shown in Table 7. Meanwhile, the Step-SPA(200) test finds 8,866 significant rejections when trading the Indonesian Rupiah, but only 935 when K=1.

In general, the test power of Step-SPA(K) is superior to that of Step-RC(K). When the performance metric is the SR of either total or MT returns, the $\#$*risk-adjusted* of both tests are similar to each other, but the $\#$*risk-adjusted* of Step-SPA(K) is higher than that of Step-RC(K) in general. In line with our findings in Section 5, when we shift our attention to the mean of total or MT returns, the Step-RC(K) tests exhibit extremely poor performance. This test only uncovers significant rejections across 5–8 currencies even when K increases to 200, whereas the Step-SPA(K) test identifies significantly outperforming strategies in at least in 13 currencies. When both tests find instances of significance in a given currency, the $\#profitable$ of Step-RC(K) is much lower than that of the Step-SPA(K) test. For instance, when measuring performance through the mean of total returns, the Step-SPA(200) test uncovers 2,657 significant rejections when trading the Thai Baht, with only 21 significance discovered by the Step-RC(200) test. These results further support the view that the SPA of Hansen (2005) is a more powerful algorithm than the RC of White (2000).

Table B.1: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=2).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	374	533	0	58	772	806	0	171
Canada	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	0	0
Euro Area	0	0	0	0	0	0	0	0
France	0	30	0	12	15	65	0	47
Germany	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	0	0
Japan	27	111	0	2	103	260	0	94
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	0	0	0	0	0	0	0	0
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	0	0	0	0	0	0	0	0
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0
Egypt	0	0	0	0	0	0	0	0
Finland	1	0	0	0	1	0	0	0
Greece	0	0	0	0	0	0	0	0
Hong Kong	0	0	0	0	0	0	0	0
Hungary	0	1	0	0	0	1	0	0
India	0	33	0	6	0	39	0	6
Indonesia	1,070	1,439	408	1,218	1,412	1,677	727	1,519
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	0	3	0	6	0	2	0	2
Mexico	0	0	0	0	0	0	0	0
Philippines	0	0	0	0	0	0	0	0
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	0	1	0	0	0	1	0	2
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	1	1	0	0	0	1	0	0
Slovenia	0	0	0	0	0	0	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	0	0	0	0	0	0	0
Spain	0	1	0	0	0	132	0	28
Taiwan	625	833	19	446	600	801	31	336
Thailand	767	956	13	500	831	914	16	635
Ukraine	181	185	188	796	128	145	141	249
Predictable Currencies	8	13	4	9	8	13	4	11

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 2$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table B.2: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=3).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	550	630	0	192	1,085	853	0	255
Canada	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	0	0	0
Euro Area	0	0	0	0	0	0	0	0
France	7	48	0	17	25	102	0	66
Germany	0	0	0	0	0	0	0	0
Italy	0	0	0	0	0	0	0	0
Japan	80	151	0	11	196	357	0	148
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	0	0	0	0	0	0	0	0
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	0	0	0	0	0	0	0	0
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0
Egypt	0	0	0	0	0	0	0	0
Finland	1	0	0	0	1	1	0	0
Greece	0	0	0	0	0	0	0	0
Hong Kong	0	0	0	0	0	0	0	0
Hungary	0	1	0	0	0	1	0	0
India	0	69	0	10	0	95	0	12
Indonesia	1,274	1,778	465	1,507	1,598	2,030	727	1,620
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	0	3	0	8	0	2	0	4
Mexico	0	0	0	0	0	0	0	0
Philippines	0	0	0	0	0	0	0	0
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	1	4	0	1	1	2	0	3
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	1	1	0	1	0	1	0	0
Slovenia	0	1	0	0	0	0	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	1	0	0	0	1	0	0
Spain	0	50	0	0	0	181	0	32
Taiwan	706	957	33	509	686	920	38	421
Thailand	925	1,185	15	650	958	1,031	16	885
Ukraine	181	227	188	821	128	146	141	299
Predictable Currencies	10	15	4	11	9	15	4	11

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 3$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table B.3: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=10).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	1,361	1,323	0	1,014	2,100	1,625	0	1,374
Canada	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	0	5	0	0
Euro Area	0	0	0	0	0	0	0	0
France	55	123	0	94	73	216	0	130
Germany	0	0	0	0	0	0	0	0
Italy	0	33	0	0	0	393	0	55
Japan	479	602	0	200	784	853	0	699
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	0	0	0	0	0	5	0	3
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	0	12	0	11	4	23	0	37
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0
Egypt	0	0	0	0	0	0	0	0
Finland	1	1	0	1	1	1	0	1
Greece	0	0	0	0	0	0	0	0
Hong Kong	0	0	0	0	0	0	0	0
Hungary	0	1	0	1	0	1	0	1
India	10	236	0	116	0	356	0	163
Indonesia	2,741	2,784	518	3,269	3,102	3,504	875	3,026
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	0	24	0	16	0	8	0	23
Mexico	0	0	0	0	0	0	0	0
Philippines	0	0	0	0	0	12	0	0
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	18	39	0	10	23	23	0	12
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	1	1	0	1	1	1	0	1
Slovenia	1	1	0	0	1	1	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	1	0	0	0	1	0	0
Spain	0	962	0	3	0	1,263	0	1,222
Taiwan	1,134	1,449	44	897	1,085	1,305	47	869
Thailand	1,290	1,633	33	1,163	1,234	1,592	16	1,335
Ukraine	310	380	189	1,006	142	163	147	615
Predictable Currencies	12	18	4	15	12	21	4	17

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 10$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table B.4: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=20).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	1,936	1,711	0	1,895	2,599	2,324	0	2,244
Canada	0	0	0	0	0	0	0	0
Denmark	0	0	0	0	9	33	0	15
Euro Area	0	0	0	0	0	0	0	0
France	122	204	0	179	154	373	0	186
Germany	0	0	0	0	0	0	0	0
Italy	0	815	0	0	0	1,698	0	561
Japan	1,156	1,061	0	543	1,187	1,410	0	1,332
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	0	3	0	0	0	13	0	19
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	43	75	0	66	37	51	0	126
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0	0	0
Egypt	0	0	0	0	0	0	0	0
Finland	1	1	0	1	1	1	0	1
Greece	0	2	0	0	0	2	0	0
Hong Kong	0	0	0	0	0	0	0	0
Hungary	0	1	0	1	0	1	0	1
India	19	404	0	332	0	650	0	363
Indonesia	3,923	4,292	518	4,390	4,316	4,805	875	4,357
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	0	62	0	25	2	21	0	35
Mexico	0	0	0	0	0	0	0	0
Philippines	0	9	0	0	0	29	0	22
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	83	120	0	86	91	139	0	73
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	1	18	0	12	1	10	0	1
Slovenia	1	2	0	0	1	2	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	1	0	0	0	1	0	0
Spain	0	1529	0	321	0	3,511	0	2,728
Taiwan	1,292	1,788	46	1,180	1,288	1,569	49	1,094
Thailand	1,609	1,878	42	1,517	1,561	1,942	16	1,587
Ukraine	767	740	245	1257	154	239	150	722
Predictable Currencies	13	21	4	15	14	22	4	19

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 20$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table B.5: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=100).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	3,657	3,532	0	4,210	4,311	4,075	0	4,094
Canada	0	0	0	0	0	0	0	0
Denmark	17	166	0	104	265	484	0	344
Euro Area	0	0	0	0	0	0	0	0
France	466	608	0	465	667	984	0	569
Germany	0	0	0	0	0	0	0	0
Italy	0	2,826	0	1,843	5	4,373	0	3,802
Japan	2,440	2,687	0	2,286	2,239	3,373	0	3,389
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	9	64	0	14	61	288	0	291
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	304	465	0	445	294	405	0	620
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	3	0	4	0	7	0	4
Egypt	0	0	0	0	0	0	0	0
Finland	1	1	0	1	1	1	0	1
Greece	2	2	0	0	2	2	0	0
Hong Kong	0	0	0	0	0	8	0	7
Hungary	0	2	0	1	0	2	0	2
India	1,069	1,156	0	1,146	1,230	1,603	0	1,045
Indonesia	6,787	6,840	518	6,749	7,301	7,763	887	6,814
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	85	166	0	169	55	171	0	129
Mexico	0	0	0	0	0	0	0	0
Philippines	0	89	0	70	2	161	0	233
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	427	471	0	557	571	541	0	615
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	92	197	0	413	58	180	0	166
Slovenia	2	2	0	0	2	2	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	3	0	2	0	2	0	1
Spain	0	4,041	0	2,121	0	5,573	0	4,562
Taiwan	1,976	2,560	65	1,966	1,481	2,466	53	2,106
Thailand	2,145	2,532	67	2,483	2,050	2,688	16	2,224
Ukraine	1,238	1,200	386	1,568	594	701	154	937
Predictable Currencies	17	23	4	21	19	24	4	22

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 100$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Table B.6: The number of outperforming technical trading rules based on generalised Step-RC/SPA tests (K=200).

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA	Step-RC	Step-SPA
Developed								
Australia	0	0	0	0	0	0	0	0
Belgium	4,517	4,226	0	4,795	5,059	4,939	0	4,792
Canada	0	0	0	0	0	0	0	0
Denmark	97	412	0	307	648	1,095	0	804
Euro Area	0	0	0	0	0	0	0	0
France	723	864	0	689	1,104	1,576	0	745
Germany	0	0	0	0	0	0	0	0
Italy	0	3,572	0	3,221	5	5,758	0	5,153
Japan	2,932	3,411	0	3,298	2,541	4,266	0	4,263
Netherlands	0	0	0	0	0	0	0	0
New Zealand	0	0	0	0	0	0	0	0
Norway	0	0	0	0	0	0	0	0
Sweden	82	186	0	80	263	472	0	486
Switzerland	0	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0	0
Others								
Austria	699	809	0	705	555	710	0	953
Brazil	0	0	0	0	0	0	0	0
Bulgaria	0	0	0	0	0	0	0	0
Croatia	0	0	0	0	0	0	0	0
Cyprus	0	0	0	0	0	0	0	0
Czech Republic	0	8	0	8	0	17	0	8
Egypt	0	0	0	1	0	0	0	0
Finland	5	1	0	1	5	1	0	1
Greece	3	3	0	0	3	3	0	0
Hong Kong	0	10	0	10	0	8	0	8
Hungary	0	2	0	2	0	2	0	2
India	1,488	1,649	0	1,556	1,704	2,077	0	1,376
Indonesia	7,920	8,193	518	7,606	7,970	8,897	1,015	7,896
Ireland	0	0	0	0	0	0	0	0
Israel	0	0	0	0	0	0	0	0
Iceland	0	0	0	0	0	0	0	0
Kuwait	0	0	0	0	0	0	0	0
Malaysia	190	237	0	278	115	236	0	226
Mexico	0	1	0	0	1	0	0	1
Philippines	0	197	0	141	10	305	0	600
Poland	0	0	0	0	0	0	0	0
Portugal	0	0	0	0	0	0	0	0
Russia	636	664	0	791	950	797	0	973
Saudi Arabia	0	0	0	0	0	0	0	0
Singapore	0	0	0	0	0	0	0	0
Slovakia	118	725	0	1,140	76	605	0	481
Slovenia	2	2	0	0	2	3	0	0
South Africa	0	0	0	0	0	0	0	0
South Korea	0	5	0	4	0	4	0	2
Spain	0	5,186	0	3,298	0	6,464	0	5,547
Taiwan	2,164	3,006	65	2,381	1,702	2,803	53	2,456
Thailand	2,566	2,986	67	2,902	2,355	3,079	16	2,837
Ukraine	1,615	1,518	832	1,670	730	869	166	1,268
Predictable Currencies	17	25	4	23	20	24	4	23

This table reports the number of significant outperforming trading rules uncovered by the generalised Step-RC and Step-SPA tests of Hsu et al. (2014) in terms of different performance metrics and return types, when $K = 200$. The *'developed'* and *'others'* samples follow the classification of Lustig et al. (2011). The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return) and the MT return (i.e., the profit generated by the time trend). Step-RC refers to the stepwise reality check of Romano et al. (2007). Step-SPA represents the generalised stepwise superiority ability test documented by Hsu et al. (2014). Following Hsu et al. (2016), we set the level of significance, α, at 5% for each data snooping bias test. *'Predictable Currencies'* represents the number of currencies with at least one significant outperforming TTR.

Appendix C. Critical values and K of the FDP-RC and SPA controls

We further report the critical values of the K-th FWER for the FDP-RC/SPA tests in Table C.1 and Table C.2, respectively. According to Table C.1, the K required by the FDP-RC to ensure the realised FDP is below 0.05 is normally lower than that of the FDP-SPA. This is consistent with what Hansen (2005) proposed, that for a given level of FDP, the FDP-SPA test has more tolerance in comparison with the FDP-RC test.[16]

Table C.2 reports the critical values of the FDP-RC/SPA in terms of various performance metrics. Not surprisingly, the critical values of the FDP-RC are always higher than that of the FDP-SPA. After considering transaction costs, the t-statistics of trading strategies are all negative for these currencies. In other words, the TTRs are no longer profitable after controlling for transaction costs. Besides the zero critical values highlighted above, the critical values of FPD-RC/SPA are normally much higher than the SHT threshold of 1.96. Consistent with Chordia et al. (2020) and Harvey et al. (2020), the critical values are around three on average. We also observe some extreme thresholds, such as 21.314 when trading Brazilian Real and 18.455 for Portuguese Escudo, when the performance metric is the SR of total excess returns.

[16] As the FDP guarantees that the realised FDR is below the pre defined γ, a higher K for FDP-SPA represents that the number of errors is easier to control than the FDP-RC.

Table C.1: Step-RC/SPA(K) tests the K required to ensure <5% FDP.

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA
Developed								
Australia	1	1	1	1	1	1	1	1
Belgium	239	216	1	251	264	259	1	250
Canada	1	1	1	1	1	1	1	1
Denmark	1	1	1	1	1	1	1	1
Euro Area	1	1	1	1	1	1	1	1
France	1	1	1	1	1	19	1	4
Germany	1	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1	1
Japan	1	158	1	1	115	219	1	1
Netherlands	1	1	1	1	1	1	1	1
New Zealand	1	1	1	1	1	1	1	1
Norway	1	1	1	1	1	1	1	1
Sweden	1	1	1	1	1	1	1	1
Switzerland	1	1	1	1	1	1	1	1
United Kingdom	1	1	1	1	1	1	1	1
Others								
Austria	1	1	1	1	1	1	1	1
Brazil	1	1	1	1	1	1	1	1
Bulgaria	1	1	1	1	1	1	1	1
Croatia	1	1	1	1	1	1	1	1
Cyprus	1	1	1	1	1	1	1	1
Czech Republic	1	1	1	1	1	1	1	1
Egypt	1	1	1	1	1	1	1	1
Finland	1	1	1	1	1	1	1	1
Greece	1	1	1	1	1	1	1	1
Hong Kong	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1
India	1	1	1	1	1	1	1	1
Indonesia	441	512	26	443	426	545	44	474
Ireland	1	1	1	1	1	1	1	1
Israel	1	1	1	1	1	1	1	1
Iceland	1	1	1	1	1	1	1	1
Kuwait	1	1	1	1	1	1	1	1
Malaysia	1	1	1	1	1	1	1	1
Mexico	1	1	1	1	1	1	1	1
Philippines	1	1	1	1	1	1	1	1
Poland	1	1	1	1	1	1	1	1
Portugal	1	1	1	1	1	1	1	1
Russia	1	1	1	1	1	1	1	1
Saudi Arabia	1	1	1	1	1	1	1	1
Singapore	1	1	1	1	1	1	1	1
Slovakia	1	1	1	1	1	1	1	1
Slovenia	1	1	1	1	1	1	1	1
South Africa	1	1	1	1	1	1	1	1
South Korea	1	1	1	1	1	1	1	1
Spain	1	1	1	1	1	410	1	1
Taiwan	99	137	1	99	71	131	1	107
Thailand	111	134	1	137	104	142	1	117
Ukraine	49	45	10	78	8	8	8	42

This table reports the K for Step-RC/SPA(K) tests to ensure the realised FDP of FDP-RC/SPA tests is below 0.05. The performance metrics contain the SR and mean of returns for each technical trading strategy. Return types contain the total return equivalent to the daily excess return of each currency and the MT return is the profit generated using the time trend of a particular currency. FDP-RC refers to the stepwise reality check controlling FDP of Romano et al. (2007). FDP-SPA represents the stepwise superiority ability test controlling FDP documented by Hsu et al. (2010). Following Chordia et al. (2020), we set the level of significance, α, at 5% for each data snooping bias test.

Table C.2: Critical values for FDP-RC/SPA(0.1) tests.

Return type	Total returns				MT returns			
Performance metric	Sharpe ratio		Mean return		Sharpe ratio		Mean return	
	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA	FDP-RC	FDP-SPA
Developed								
Australia	3.968	3.334	4.621	3.282	3.888	3.155	4.663	3.270
Belgium	2.622	2.737	5.393	2.559	2.588	2.614	5.869	2.669
Canada	4.288	3.040	5.362	3.158	4.218	3.080	5.526	3.008
Denmark	4.220	3.971	4.943	3.908	3.974	3.813	5.221	3.972
Euro Area	4.771	3.828	5.539	3.833	4.341	3.772	5.504	3.869
France	4.147	3.813	5.216	4.015	4.034	3.209	5.169	3.698
Germany	4.117	3.707	4.748	3.751	4.248	3.720	4.651	3.764
Italy	5.711	4.957	7.027	5.194	5.999	5.022	6.076	5.093
Japan	4.257	2.803	6.187	4.432	3.137	2.605	6.242	4.323
Netherlands	4.252	3.807	4.300	3.753	3.992	3.746	4.102	3.783
New Zealand	4.444	2.941	5.760	2.926	4.233	2.855	5.873	2.898
Norway	4.000	3.884	5.376	3.856	4.026	3.838	5.260	3.880
Sweden	4.048	3.712	6.331	3.766	4.047	3.781	5.708	3.757
Switzerland	4.393	2.696	5.699	2.862	4.462	3.057	5.831	2.824
United Kingdom	4.508	0.000	5.904	0.000	4.645	0.000	5.632	0.000
Others								
Austria	4.035	3.833	5.512	3.887	3.929	3.840	5.462	3.868
Brazil	21.314	3.577	6.580	3.558	19.747	3.607	6.562	3.557
Bulgaria	10.021	3.377	6.810	3.408	12.128	3.351	7.459	3.504
Croatia	4.034	3.759	6.289	3.801	4.143	3.788	6.288	3.687
Cyprus	5.921	0.000	4.695	0.000	6.008	0.000	4.823	0.000
Czech Republic	4.015	3.896	5.471	4.217	3.968	3.926	5.778	4.344
Egypt	8.616	5.816	7.853	2.863	9.691	5.578	7.400	2.973
Finland	3.627	3.692	4.276	3.976	3.648	3.717	4.683	3.962
Greece	4.536	4.225	4.236	4.264	4.419	4.196	4.400	4.273
Hong Kong	4.161	3.244	4.742	3.802	4.014	3.082	4.747	3.789
Hungary	4.910	3.784	6.194	4.121	4.645	3.780	5.511	3.965
India	4.636	4.219	6.465	4.345	6.857	4.176	6.243	4.352
Indonesia	2.523	1.960	5.856	2.475	2.834	1.725	5.422	2.427
Ireland	4.908	0.000	5.327	0.000	5.056	0.000	5.178	0.000
Israel	3.887	3.672	5.850	3.745	3.968	3.780	5.745	3.801
Iceland	4.616	3.587	6.652	3.704	4.426	3.546	6.709	3.686
Kuwait	6.150	0.000	6.229	0.000	6.731	0.000	5.739	0.000
Malaysia	15.280	5.626	9.502	4.965	16.363	5.751	9.200	4.810
Mexico	3.746	3.761	5.939	3.956	3.784	3.806	5.483	3.916
Philippines	3.995	3.757	5.798	3.946	4.091	3.736	6.373	3.802
Poland	5.010	3.739	6.398	3.950	5.271	3.761	5.651	3.917
Portugal	18.455	3.458	6.971	3.639	20.543	3.568	6.654	3.930
Russia	4.869	4.611	6.181	4.243	5.044	4.634	5.989	4.064
Saudi Arabia	4.269	0.890	5.631	0.000	4.451	0.862	6.346	0.000
Singapore	4.213	1.506	5.793	1.436	4.099	1.464	5.626	1.601
Slovakia	4.139	3.954	5.758	4.286	4.211	4.092	5.439	4.301
Slovenia	4.421	3.833	4.297	3.889	4.419	4.005	4.526	3.985
South Africa	3.912	3.623	4.905	3.571	3.875	3.750	5.328	3.732
South Korea	6.593	4.445	6.236	4.025	6.629	4.276	6.682	3.983
Spain	8.617	5.544	7.221	5.073	8.699	2.329	6.773	4.467
Taiwan	3.147	2.737	5.784	3.076	3.477	2.819	5.833	3.011
Thailand	3.115	2.921	5.633	2.963	3.205	2.883	5.537	3.170
Ukraine	5.187	5.329	6.630	4.187	6.425	5.772	6.670	4.311

This table reports the critical values of FDP-RC/SPA tests. The performance metrics contain the SR and mean of returns for each technical trading strategy. Return types contain the total return equivalent to the daily excess return of each currency, and the MT return that is the profit generated using the time trend of a particular currency. FDP-RC refers to the stepwise reality check controlling FDP of Romano et al. (2007). FDP-SPA represents the stepwise superiority ability test controlling FDP documented by Hsu et al. (2010). Following Chordia et al. (2020), we set the level of significance, α, at 5% for each data snooping bias test.

Appendix D. The out-of-sample survival rate for each currency

Table D.1 reports the Out-Of-Sample (OOS) survival rates for *'Predictable Currencies'* based on either MHT or SHT. First, according to the SHT results, the survival rates of 12 currencies are equal to zero in each panel, which further illustrates the poor test power of the SHT. We also find that some in-sample outperforming trading rules identified by the MHT approaches fail to achieve the same level of superiority over the out-of-sample periods. For instance, in Panel A, the survival rates of trading the Kuwaiti Dinar and the Mexican Peso are all equal to zero when controlling the FWER. This also verifies that the MHT approach can manage Type I errors but can not guarantee out-of-sample performance.

Table D.1: The OOS survival rates

Currencies	Control approaches								
	FWER				FDP		FDR		SHT
	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	
Panel A: Mean of total returns									
Brazil	-	-	-	-	-	-	-	-	0.0%
Bulgaria	-	-	-	-	-	-	-	-	0.0%
Croatia	-	-	-	-	-	-	-	-	0.0%
Czech Republic	-	-	-	-	-	-	-	-	0.0%
Denmark	-	-	-	-	-	-	-	-	0.0%
Euro Area	-	-	-	-	-	-	-	-	0.0%
Hungary	-	-	-	-	-	-	-	-	0.0%
Iceland	-	-	-	-	-	-	-	-	29.6%
India	40.9%	41.3%	12.3%	32.4%	34.1%	58.9%	47.5%	52.9%	45.1%
Ireland	-	-	-	-	-	-	-	-	0.0%
Japan	0.0%	0.0%	-	0.0%	-	0.1%	0.2%	0.0%	0.2%
Malaysia	0.0%	0.0%	-	0.0%	-	0.0%	0.0%	0.0%	0.0%
Mexico	-	-	-	-	-	-	-	-	0.0%
Norway	-	-	-	-	-	-	-	-	0.0%
Philippines	-	-	-	-	-	-	-	-	5.6%
Poland	-	-	-	-	-	-	-	-	0.0%
Russia	-	-	-	-	-	-	-	-	19.9%
Slovakia	-	-	-	-	-	-	18.5%	-	30.5%
South Korea	-	-	-	-	-	-	-	-	0.0%
Sweden	-	-	-	-	-	-	-	-	0.0%
Taiwan	-	-	-	-	-	-	73.1%	-	60.1%
Thailand	66.7%	66.7%	-	80.0%	-	80.0%	76.3%	-	54.1%
Mean	29.9%	30.2%	12.3%	23.3%	34.1%	36.8%	30.5%	32.3%	22.3%
Panel B: Mean of MT returns									
Brazil	-	-	-	-	-	-	-	-	0.0%
Bulgaria	-	-	-	-	-	-	-	-	0.0%
Croatia	-	-	-	-	-	-	-	-	0.0%
Czech Republic	-	-	-	-	-	-	-	-	0.0%
Denmark	-	-	-	0.0%	-	0.0%	0.0%	-	0.0%
Euro Area	-	-	-	-	-	-	-	-	0.0%
Hungary	-	-	-	-	-	-	-	-	0.0%
Iceland	-	-	-	-	-	-	-	-	27.9%
India	48.7%	49.0%	18.8%	38.3%	42.2%	58.7%	45.9%	51.7%	44.4%
Ireland	-	-	-	-	-	-	-	-	0.0%
Japan	0.0%	0.0%	-	0.0%	-	0.1%	0.2%	0.0%	0.4%
Malaysia	0.0%	0.0%	-	0.0%	-	0.0%	0.0%	0.0%	0.0%
Mexico	-	-	-	-	-	-	-	-	0.0%
Norway	-	-	-	-	-	-	-	-	0.0%
Philippines	-	-	-	-	-	-	-	-	5.1%
Poland	-	-	-	-	-	-	-	-	0.0%
Russia	-	-	-	-	-	-	-	-	20.1%
Slovakia	-	-	-	-	-	-	13.9%	-	28.9%

Continued on next page

Table D.1 – continued from previous page

	Bonferroni	Holm	Step-RC	Step-SPA	FDP-RC	FDP-SPA	BH	BY	SHT
South Korea	-	-	-	-	-	-	-	-	0.0%
Sweden	-	-	-	-	-	-	-	-	0.0%
Taiwan	-	-	-	-	-	-	73.5%	-	60.0%
Thailand	66.7%	66.7%	-	81.8%	-	81.8%	75.0%	-	53.1%
Mean	40.3%	40.7%	18.8%	27.9%	42.2%	37.3%	28.1%	33.9%	21.1%
Panel C: Sharpe Ratio of Total Returns									
Brazil	-	-	-	-	-	-	-	-	0.0%
Bulgaria	-	-	-	-	-	-	-	-	0.0%
Croatia	-	-	-	-	-	-	-	-	0.0%
Czech Republic	-	-	-	-	-	-	-	-	0.0%
Denmark	-	-	-	-	-	-	-	-	0.0%
Euro Area	-	-	-	-	-	-	-	-	0.0%
Hungary	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	0.0%	-	0.0%
Iceland	-	-	50.0%	0.0%	50.0%	0.0%	-	-	29.5%
India	51.4%	52.0%	24.0%	32.3%	59.1%	58.1%	47.0%	52.5%	44.8%
Ireland	-	-	-	-	-	-	-	-	0.0%
Japan	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.0%	0.2%
Malaysia	0.0%	0.0%	-	-	-	-	0.0%	0.0%	0.0%
Mexico	-	-	-	-	-	-	-	-	0.0%
Norway	-	-	-	-	-	-	-	-	0.0%
Philippines	-	-	-	-	-	-	-	-	5.7%
Poland	-	-	-	-	-	-	-	-	0.0%
Russia	-	-	-	-	-	-	-	-	19.6%
Slovakia	-	-	-	-	-	-	18.9%	-	30.5%
South Korea	-	-	-	-	-	-	-	-	0.0%
Sweden	-	-	-	-	-	-	-	-	0.0%
Taiwan	100.0%	100.0%	100.0%	94.0%	100.0%	88.6%	77.2%	-	59.1%
Thailand	90.9%	90.9%	94.2%	93.8%	92.0%	89.3%	76.6%	93.8%	54.4%
Mean	38.7%	39.1%	22.0%	25.7%	44.0%	44.5%	31.4%	32.7%	22.3%
Panel D: Sharpe Ratio of MT returns									
Brazil	-	-	-	-	-	-	-	-	0.0%
Bulgaria	-	-	-	-	-	-	-	-	0.0%
Croatia	-	-	-	-	-	-	-	-	0.0%
Czech Republic	-	-	-	-	-	-	-	-	0.0%
Denmark	-	-	-	0.0%	-	0.0%	0.0%	-	0.0%
Euro Area	-	-	-	-	-	-	-	-	0.0%
Hungary	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	0.0%	-	0.0%
Iceland	-	-	-	0.0%	-	0.0%	-	-	27.8%
India	56.4%	56.5%	30.0%	33.3%	57.9%	57.9%	45.5%	51.4%	44.2%
Ireland	-	-	-	-	-	-	-	-	0.0%
Japan	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.0%	0.4%
Malaysia	0.0%	0.0%	-	-	-	-	0.0%	0.0%	0.0%
Mexico	-	-	-	-	-	-	-	-	0.0%
Norway	-	-	-	-	-	-	-	-	0.0%
Philippines	-	-	-	-	-	-	-	-	5.4%
Poland	-	-	-	-	-	-	-	-	0.0%
Russia	-	-	-	-	-	-	-	-	20.1%
Slovakia	-	-	-	-	-	-	14.2%	-	28.9%
South Korea	-	-	-	-	-	-	-	-	0.0%
Sweden	-	-	-	0.0%	-	0.0%	0.0%	-	0.0%
Taiwan	100.0%	100.0%	100.0%	95.5%	100.0%	90.5%	76.8%	-	58.9%
Thailand	92.3%	92.3%	94.0%	93.5%	92.7%	89.5%	75.7%	93.8%	53.5%
Mean	45.6%	45.8%	19.5%	23.6%	40.4%	44.5%	27.6%	33.3%	21.2%

This table reports the survival rates ($\frac{Num_{OOS}}{Num_{IS}}$) measured by Equation 22, where N_{OOS} represents the number of survivors, and Num_{IS} refers to the number of in-sample instances of significance. We only present the currencies with at least one instance of significance in terms of either MHT or SHT, with 22 currencies reported in each panel. The performance metrics considered are the SR and mean of returns. Return types considered include total return (i.e., the daily excess return), and the MT return (i.e., the profit generated by the time trend). Each panel presents the rates based on a given performance appraisal, including the total return, MT return, total SR, and MT SR. Mean refers to the average survival rate computed using Equation 23, $\frac{\sum_{c=1}^{NC} Num_{OOS}^{c}}{\sum_{c=1}^{NC} Num_{IS}^{c}}$, where NC represents the number of currencies with in-sample instances of significance based on the results of a specific MHT approach.