Controlling for Luck in Picking Outperforming Trading Strategies

Tren Ma

Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, UK

DinhTren.Ma@glasgow.ac.uk

This version: July 8, 2023

Abstract

This paper develops a new framework to estimate the false discovery rate (FDR) using informative covariates, namely, the multivariate functional false discovery rate ($mfFDR$). The method gains a considerably higher power than the one that utilizes only one covariate and the ones that use only $p-$values. This advantage is robust under various settings of dependencies and estimation errors that are typical in economic and financial data. In further empirical analyses, the method is applied to detecting truly profitable trading rules from a large and widely used universe of technical trading rules in the foreign exchange market. Particularly, four informative covariates are introduced and analyzed. The $mfFDR$ using the four covariates detects many more profitable rules than existing methods regardless of the group of currencies and sample periods. The portfolios of trading rules selected by $mfFDR$ produce significantly positive net Sharpe ratios which are higher than those of portfolios selected by the existing methods. More importantly, the more covariates being used the higher Sharpe ratio the portfolios gain. Finally, via analyzing profitability of the technical rules conditionally on their category, useful insights on the use of the technical rules in practice are uncovered.

Keywords: Multiple testing, Multivariate functional false discovery rate, Informative covariates, Technical analysis, Sharpe ratio, Foreign exchange market.

---

1. Introduction

Data-snooping is a statistical bias that appears when a dataset is used more than once, for inference and model selection. It can lead to results that seem statistically significant but are due to luck, i.e. false discoveries. In financial research, data-snooping is prevalent and in some cases unavoidable (White, 2000). To rectify this bias, several multiple testing corrections approaches have been presented, such as the contributions of White (2000), Hansen (2005), Barras *et al.* (2010), Bajgrowicz and Scaillet (2012) and Hsu *et al.* (2010) among others. However, those approaches show a lack of power and this has recently gained attention. A common feature of those existing approaches is that, they utilize only p-value of hypothesis testings and ignoring side information which might be useful.

In this study, we introduce a new framework to control the false discovery rate (FDR) of Benjamini and Hochberg (1995), BH henceforth, which is the expectation of false discovery proportion among the ones called significant. Our new method tests the nulls conditionally on realized informative covariates and estimates the FDR as a function of them. The method shows a superior power in comparison to the non-covariate FDR approach of Storey (2002), which is introduced into finance by Barras *et al.* (2010), and the functional FDR ($fFDR$) in Chen *et al.* (2021), CRS henceforth, and Hsu *et al.* (2021), HKMS henceforth, which estimates the FDR as a function of only one covariate. Via simulations, we show that our method performs well in controlling for the FDR under various settings. Its performance in terms of power is impressive and beats that of the aforementioned methods with gaps of up to about 67% and 44%, respectively. The advantages of the proposed procedure remain under: i) weak signal-to-noise data; ii) weak dependent among tests; iii) covariates estimated with noise; and iv) correlated covariates.

As our method estimates multivariate functions, it performs best when having a sufficient number of inputs, i.e. the number of hypothesis tests. Our method therefore perfectly fits in applying to control the lucky rate in selecting outperforming technical trading rules where the size of universe of rules can reach thousands. In this particular study we focus on the rules which are applied in the foreign exchange (FX) trading. We have some good reasons for this choice. First, in the FX trading, technical analysis is widely used in both the academia and in practice (Goodman, 1979, e.g.,). The studies of Allen and Taylor (1990) and Taylor and Allen (1992) documented that technical analysis is, indeed, an essential tool in decision making in the FX market. Analyses of Menkhoff and Taylor (2007) additionally show that almost all foreign exchange professionals use technical analysis as a tool at least to some degree. Second, the academics have treated this popularity of the technical analysis, which has no economic fundamental, with skepticism and a debate on its profitability is still ongoing. Meese and Rogoff (1983) and Chinn and Meese (1995) document that major exchange rates follow a random walk, and Cialenco and Protopapadakis (2011) claim that simple technical trading rules are not significantly profitable. In contrast, Neely *et al.* (2009) argue that genuine technical trading rules exist in 1970s and 1980s and observe a decline in the profitability of the rules by early 1990s. Hsu *et al.* (2010, 2016) also find substantial profitability of the rules both in developed and emerging currencies.

We implement our new method to detect genuinely profitable trading rules in the set of more than 21,000 rules used in Hsu *et al.* (2016). Using daily data over a maximum of 50 years for 30 U.S. dollar exchange rates, which spans from 1971 to 2020, we construct monthly rolling portfolios of significantly profitable trading rules, selected ex-ante based on past 12-month window, and invested in the following month. We use as inputs of the $mfFDR$ four covariates including the auto-correlation of a trading rule's excess return and, the R-square, the beta and alpha coefficients of a regression of the trading rule's excess return on the excess return of passive buy and hold strategy. The informativeness of those covariates are confirmed via two facts: i) the $mfFDR$ with use of those covariates gains higher power compared to the FDR and; ii) portfolio of rules selected by $mfFDR$ out-performs those selected by its opponents. More specifically, on average the $mfFDR$ detects at least twice number of out-performing strategies compared to FDR when implementing on whole sample period and control FDR at target of 20%. This advantage is unchanged for sub-samples of 10 years. In terms of profitability, we show that the $mfFDR$ based portfolio with the use of multiple covariates performs better than the $fFDR$ based portfolios with the use of only one covariate, which in turn outperforms the FDR based portfolio of Bajgrowicz and Scaillet (2012). In particular, the $mfFDR$ based portfolio gains positive profit in 28 over 30 considered currencies for an extended period from 1973 to 2020 after transaction cost, and 16 among them are statistically significant. A similar $mfFDR$ based portfolio, when constructed based on a pool of more than 630,000 trading rules applied on all 30 currencies, gains a Sharpe ratio of 1.06 and 0.95 before and after transaction cost, respectively. The results are retained regardless the choice of FDR targets and performance metrics.[1] Finally, we investigate the profitability of portfolios constructed conditionally on each of the five popular categories of technical rules including Channel Breakout, Filter, Moving Average, Relative Strength Index (RSI) and Support-resistant. We show that the portfolio corresponding to Filter category is the best but only when transaction cost is relatively low. In contrast, the ones corresponding to Moving Average and Channel Breakout categories are the best after transaction cost as they have lower trading frequency. The results also suggest traders to exclude the RSI category from the pool when constructing a portfolio.

Our paper contributes to econometrics and finance literature in several aspects. First, our development of the multivariate functional false discovery rate ($mfFDR$) offers a novel approach to correct for data-snooping bias and shows a strong validation to financial data which are typically dependent and noisy. It not only provides a method that is more powerful than existing ones in detecting out-performers, but also allows us adapting sufficiently information in decision making. Second, our results from the implementation of the largest set of rules up to date, which consists of more than 630,000 rules traded over maximum 50 years and 30 currencies, allow us to conclude the benefit of using technical analysis in FX market. Third, the results offer valuable guidance to traders regarding the specific categories of technical rules they should include when constructing a portfolio, taking into consideration transaction costs.

The rest of the paper is organized as follows. The next section develops the $mfFDR$ framework and designs simulations to show its performance. Section 3 presents descriptions on the data and trading rule universe and Section 4 is devoted for trading rule's performance measure. In Section 5, we present the empirical results where $mfFDR$ based portfolios are constructed both on individual currency and a basket of 30 currencies. Finally, Section 6 concludes the paper.

2. The use of covariates in FDR framework

In this section, we introduce the $mfFDR$ that estimates the false discovery rate as a function of more than one informative covariates. Our approach develops the frameworks of CRS and HKMS on the $fFDR$ with a single covariate. First, we present the setting of our method and its implementation in the context of data snooping and FX trading. Second, we conduct a set of simulations that demonstrate the value of multiple informative

[1] We investigate with FDR targets from 10% to 40% and use Sharpe ratio and mean return as performance metrics of the rules.

covariates in controlling the false discovery rate and the superior power of $mfFDR$ compared to the related existing approaches. Third, we validate the performance of our method under certain dependent types of data which are typical in finance. We confirm that our approach retains its power and control of the false discovery rate when the statistics are weakly dependent and when informative covariates are correlated and contain estimation errors - all these features are common in most financial data and topics.

2.1. The multivariate functional false discovery rate

Suppose we are having n trading rules and each of them produces an excess return. Assume that there are covariates that convey information about the performance of each trading rule, which is measured by a metric ϕ. As will be detailed in following sections, in this study we use Sharpe ratio (SR) as our main performance metric. To assess the performance of each strategy, conditional on realization of the covariates, we conduct a hypothesis test

$$H_0 : \phi = 0, \qquad H_1 : \phi \neq 0. \tag{1}$$

The n trading rules produce n conditional tests as in (1). Our aim is to detect a maximal number of strategies having significant none-zero ϕ while controlling for FDR, the expected proportion of false discoveries among the hypothesis tests called significant as introduced in BH.

For convenience of notation, let us consider the single test (1). To formulate the assumptions, we assume there are d covariates, represented by random variables $Z^1, \ldots, Z^d$, conveying information about the probability of a hypothesis being true null as well as the distribution of the $p-$value of the false null hypothesis. Let us denote by P the random variable representing the $p-$value of the test and $\boldsymbol{Z} = (Z^1, \ldots, Z^d)$. Also, let h be the true status of the hypothesis, that is, $h = 0$ if the null hypothesis is true and $h = 1$ if the null hypothesis is false. We assume that, conditional on $\boldsymbol{Z} = \boldsymbol{z}$, the hypothesis is a priori true null with probability $\pi_0(\boldsymbol{z})$, i.e. $(h|\boldsymbol{Z} = \boldsymbol{z}) \sim Bernoulli(1 - \pi_0(\boldsymbol{z}))$. Its estimation procedure will be discussed in a latter stage. For each j, we transform the observed value of the covariate Z_i^j to a form so that Z_i^j uniformly distributed on $[0, 1]$, via using $z_i^j = r_i^j/n$ where r_i^j is the rank of the observed value of Z_i^j in the set of observed values of $Z_1^j, \ldots, Z_n^j$, $j = 1, \ldots, d$. In the followings, Z^js are the ones in the transformed forms. To formulate the theoretical framework, we require $\boldsymbol{Z}$ and P satisfying: i) Z^js are mutually independent, $j = 1, \ldots, d$; ii) Conditional on $\boldsymbol{Z} = \boldsymbol{z}$, when the null hypothesis is true, P is uniformly distributed on the interval $[0, 1]$ and, when the null hypothesis is false, P has a distribution determined by some density function $f_{alt}(p|\boldsymbol{z})$. To develop the theoretical framework, we assume that the aforementioned n tests are independent replications of the test (1), i.e., the triple h, $p-$value and covariates of the tests are independent and each of them has the same distribution as the triple $(h, P, \boldsymbol{Z})$. Note that, in the next Sections, we show that our method can be also applied to dependent cases, when there is a dependence among the tests and among the covariates.

The gist of our method is as following. Given a target τ of FDR, we do not decide to reject the null hypotheses based on their $p-$values solely. We instead reject a null hypothesis by a rule based on both the $p-$value and the covariates. Thereby, we define for each hypothesis with an observed $(p, \boldsymbol{z})$, a posterior probability of being null denoted by $r(p, \boldsymbol{z})$. If there are any significant hypotheses, the hypothesis with smallest $r(p, \boldsymbol{z})$ will be selected first, then the second smallest one and so on. Each time a hypothesis is added to the significant set, the FDR is raising. We stop the procedure when the FDR target is reached. More specifically, the posterior probability of being true null is

$$r(p, \boldsymbol{z}) = \mathbb{P}(h = 0|(P, \boldsymbol{Z}) = (p, \boldsymbol{z})) \tag{2}$$

which can be developed further as

$$r(p, \boldsymbol{z}) = \frac{\mathbb{P}(h = 0|\boldsymbol{Z} = \boldsymbol{z})}{\mathbb{P}((P, \boldsymbol{Z}) = (p, \boldsymbol{z}))} = \frac{\pi_0(\boldsymbol{z})}{f(p, \boldsymbol{z})} \tag{3}$$

where $f(p, \boldsymbol{z})$ is the joint density function of the $p-$value and covariates.[2]

Empirically, $r(p, \boldsymbol{z})$ is estimated by $\hat{r}(p, \boldsymbol{z}) = \hat{\pi}_0(\boldsymbol{z})/\hat{f}(p, \boldsymbol{z})$ where $\hat{\pi}_0(\boldsymbol{z})$ and $\hat{f}(p, \boldsymbol{z})$ are estimators of $\pi_0(\boldsymbol{z})$ and $f(p, \boldsymbol{z})$, respectively. For the sake of space, we refer the details of the estimation procedures to the Appendix A.

Consequently, the rejection region has a form $\Gamma(\theta) = \{(p, \boldsymbol{z})|\hat{r}(p, \boldsymbol{z}) \leq \theta\}$ where $\theta \in (0, 1)$ is satisfying

$$\int_{\Gamma(\theta)} \hat{r}(p, \boldsymbol{z}) dp d\boldsymbol{z} \leq \tau. \tag{4}$$

The left side of the equation (4) is the estimate of the FDR corresponding with the rejection region $\Gamma(\theta)$. Hence, we are choosing a maximal threshold $\theta = \theta^*$ such as the condition (4) holds. Thus, a hypothesis is significant if and only if its observed $(p, \boldsymbol{z})$

[2]The equation (3) is obtained by using the fact that $\mathbb{P}(h = 0, P = p, \boldsymbol{Z} = \boldsymbol{z}) = \mathbb{P}(P = p|(h = 0, \boldsymbol{Z} = \boldsymbol{z})).\mathbb{P}(h = 0|\boldsymbol{Z} = \boldsymbol{z}).\mathbb{P}(\boldsymbol{Z} = \boldsymbol{z}) = \mathbb{P}(h = 0|\boldsymbol{Z} = \boldsymbol{z})$ where the first and last factors equal 1 as they are density functions of uniform distributions.

belongs to the set $\Gamma(\theta^*)$. This rejection rule is implemented via the "functional $q-$value" introduced in CRS. For each hypothesis i with observed $(p, \boldsymbol{z}) = (p_i, \boldsymbol{z}_i)$, we determine its $q-$value as the estimate of the FDR when we reject the null of all hypotheses j having $\hat{r}(p_j, \boldsymbol{z_j}) \leq \hat{r}(p_i, \boldsymbol{z}_i)$. Thus, at the given target τ of FDR, a null hypothesis is rejected if and only if its corresponding $q-$value $\leq \tau$.

We name the proposed procedure, where the $\boldsymbol{z}$ is a vector of more than one covariate, as $mfFDR$. When $d = 1$, the $mfFDR$ is the $fFDR$ of HKMS. In the followings, we present an intuitive illustration on the value of informative covariates in the $mFDR$ framework and we compare the performance of the proposed method to others in terms of FDR control and power.

2.2. Simulation studies

We consider the simplest case where we have two informative covariates $\boldsymbol{Z} = (U, V)$. We simulate $n = 10,000$ hypotheses where the proportion of the null hypotheses is approximately 0.8. Suppose that the two covariates convey information of the hypotheses. We will demonstrate that by using $mfFDR$, which utilizes both covariates as inputs, we obtain a higher power in detecting false null hypotheses than the $fFDR$ which uses only one of the two covariates.

The data generating process in our $mfFDR$ simulations is as following. In each iteration we draw independently the elements of the covariates $U = (u_1, \ldots, u_n)$ and $V = (v_1, \ldots, v_n)$ from the uniform distribution $U(0, 1)$. For each hypothesis i we draw its null status h_i from a Bernoulli distribution with a probability of being null $\mathbb{P}[h_i = 0|(u, v) = (u_i, v_i)] = \pi_0(u_i, v_i)$, i.e. $h_i \sim Bernoulli(1 - \pi_0(u_i, v_i))$, where $\pi_0(u, v)$ has one of the two following forms

- $\pi_0(u, v) = \sin[\pi(u + v)/2]$, i.e. a sine function;
- $\pi_0(u, v) = 1-(u^4+v^4)/2$, i.e. a monotonic function (with respect to each covariate).

We obtain two bundles of (U, V, H) that correspond to the two forms of the $\pi_0(u, v)$, where $H = (h_1, \ldots, h_n)$.

Next, for each triple (U, V, H), we generate $p-$values for tests such that for each true null hypothesis, i.e. the one with $h_i = 0$, its $p-$value is drawn from the uniform distribution $U[0, 1]$, whereas the $p-$value of a false null hypothesis the distribution with density function $f_{alt}(p|(u, v))$. We specify $f_{alt}(p|(u, v))$ by using a Beta distribution $Beta(\alpha, \beta)$, i.e. $f_{alt}(p|(u, v)) \propto p^{\alpha-1}(1 - p)^{\beta-1}$ where the α and β are positive real parameters determining the shape of the distribution. Here, similar to CRS, we set the β as a function of the covariates, specifically as $\beta = 3 + 1.5(u + v)$. Aiming to study the performance of the method in various circumstances, we consider three cases of $\alpha \in \{0.5, 1, 1.5\}$ which we name as strong, weak and very weak signal, respectively (thus, we have three specifications of a Beta distribution). In the strong signal case, the false null hypotheses are easier to be distinguished from the true null ones than those in the weak and very weak signal cases.

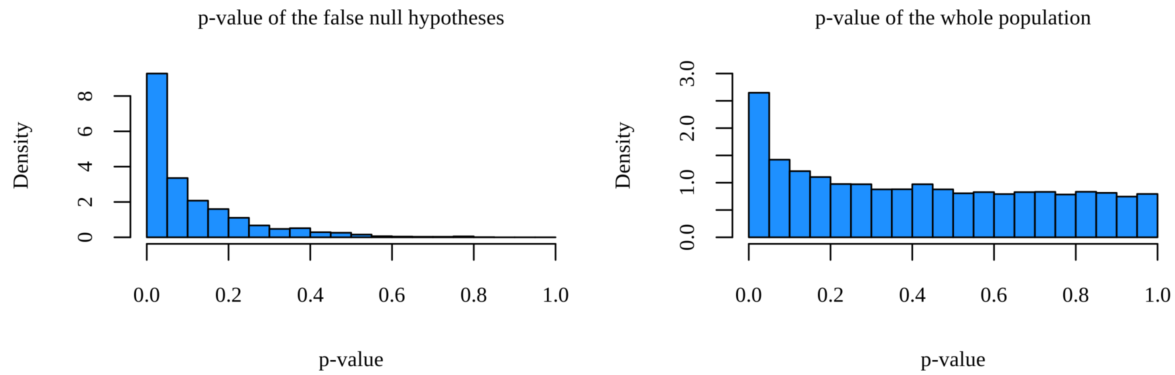
To illustrate clearer how different the three cases are, Figure 1 depicts in Panel A (Panel B and C) the distribution of $p-$values of the false null hypotheses drawn from the strong (weak and very weak) signal case on the left, and of the whole population on the right of the panel. In these panels, the probability of being null each hypothesis is generated from the same sine form of $\pi_0(u, v)$ and the $p-$values of the true null hypotheses are drawn from the uniform distribution $[0, 1]$. In the strong signal case ($\alpha = 0.5$), the $p-$values of the false null ones are mostly concentrated near the zero point. In contrast, those $p-$values under the weak signal setting are less condensed at the zero point and dispersed remarkably up to 0.6. In the very weak case the peak departs from the zero point. The false null hypotheses in the weak and the very weak cases are more difficult to be detected. For example, if we reject a null hypothesis whenever its $p-$value is less than 0.05, then we detect half of the false null hypotheses in the strong signal case while in the very weak signal case the detected portion is much smaller.[3] In the context of technical trading rule, a weak signal means that the truly out/under-performing rules are having small absolute Sharpe ratios and thus they are having large $p-$values which make them more difficult to be detected from a random walk.

The task is to detect the false null hypotheses from the simulated sample with control of the FDR at given targets by using only the $p-$value and the covariates. In our first experiment, we illustrate the of role of informative covariates in detecting false null hypotheses. Then we benchmark our procedure, the $mfFDR$ with both U and V as covariates, against the $fFDR$ with only U as covariate, the FDR of Storey (2002) which we notate as standard FDR ($StdFDR$), and the FDR of BH. By comparing the obtained results with the null status of the hypotheses (i.e., the H) we calculate a false discovery proportion, which is the ratio of the number of true null hypotheses falsely rejected over the number of the discoveries, and a correct detection proportion, which is the ratio of the number of false null hypotheses detected over the number of false null ones in the

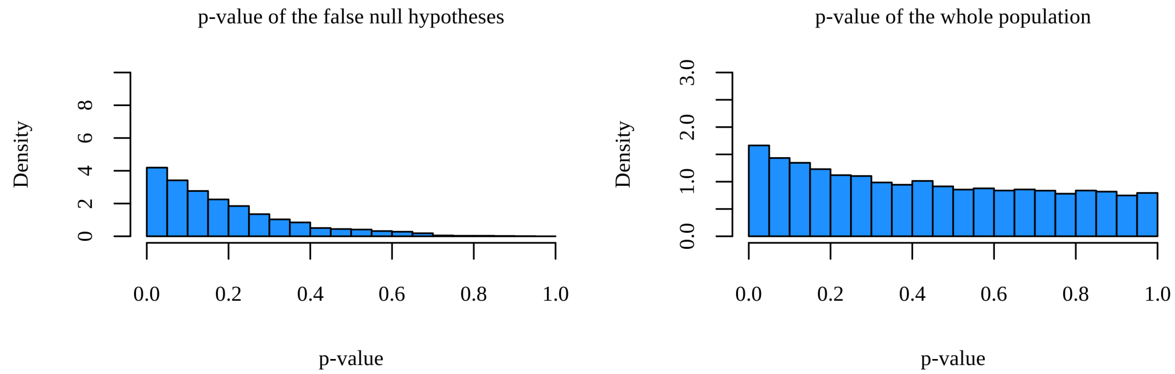
[3] We also note that, the same number of true null hypotheses will be wrongly rejected at this threshold for all cases, consequently the false discovery proportion is much higher in the very weak signal case.

Figure 1: Distribution of $p-$values of false null hypotheses component and of the whole population in three scenarios: the $p-$values of the false null hypotheses are drawn from a strong signal case (Panel A), a weak signal case (Panel B) and a very weak signal case (Panel C).

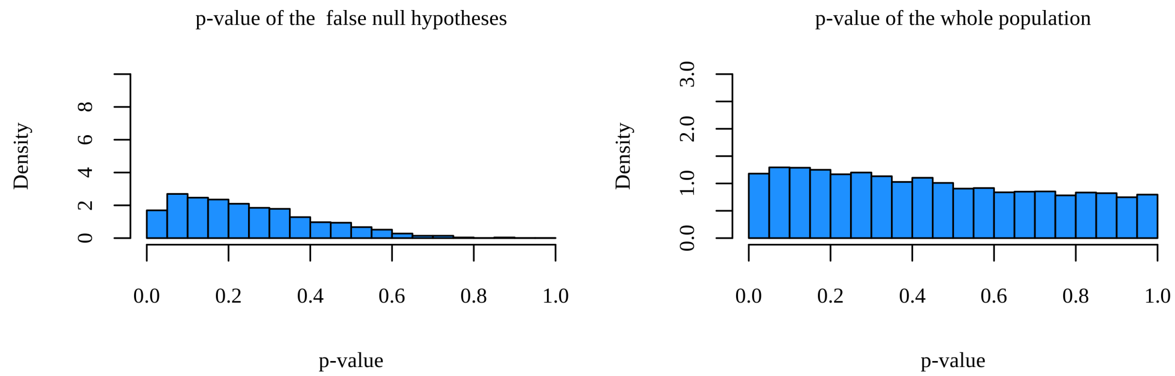
Panel A: Distribution of $p-$value under the strong signal case ($\alpha = 0.5$).



Panel B: Distribution of $p-$value under the weak signal case ($\alpha = 1$).



Panel C: Distribution of $p-$value under the very weak signal case ($\alpha = 1.5$).

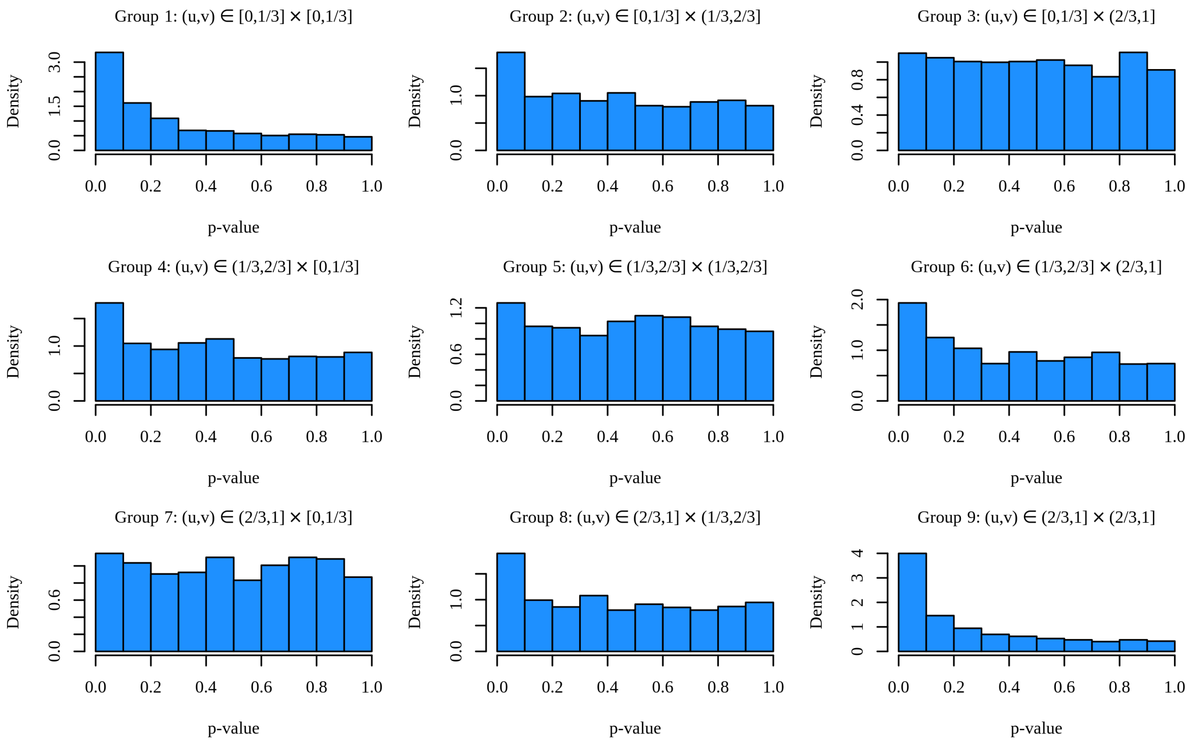


population.

2.2.1. *The role of informative covariates in detecting false nulls*

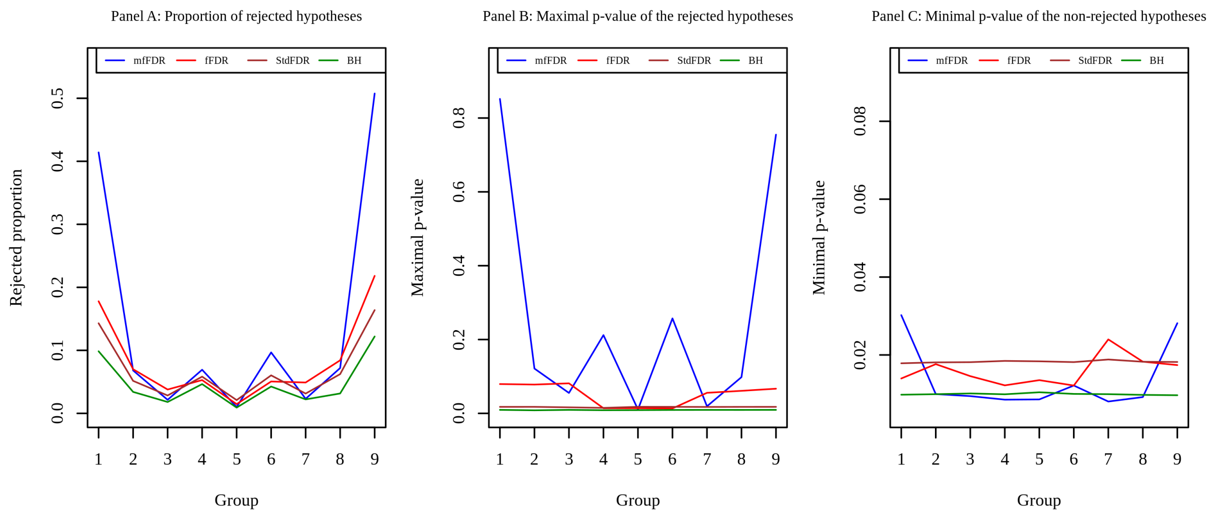
In order to show the usefulness of informative covariates and how they work in the FDR framework, we analyze a simulated sample. Particularly, we select the one depicted in the Panel A of the Figure 1. This sample is generated under the strong signal setting with the sine form of the $\pi_0(u, v)$. To see how the covariates convey the information of the hypotheses, we partition the sample into nine groups based on dividing each of the u and v into three equal segments. For convenience, the groups are named from 1 to 9, corresponding to the nine areas of (u, v). In doing so, each group contains roughly the same number of hypotheses. The $p-$value distributions of the groups are depicted in Figure 2.

Figure 2: Distribution of $p-$values of nine groups which are partitioned from the sample in Panel A of the Figure 1 based on the value of the covariates u and v. Each sub-figure represents for a group of hypotheses corresponding to the values of (u, v) shown in its title.



To control the FDR at a target τ, which is $\tau = 0.2$ in this particular example, the BH and $StdFDR$ reject null hypotheses based on the $p-$values by seeking a threshold at which all null hypotheses with smaller $p-$value are rejected. Thus, this threshold is fixed for all groups. However, as shown in the Figure 2, groups 1, 2, 4, 6 and 9 contain much more false null hypotheses than others.[4] contains more false null hypotheses. More specifically, as the $p-$values of the true null hypotheses are uniformly distributed, a $p-$value histogram such as the one of the group 3 indicates that most of hypotheses in this group are true null. In other words, the true null proportion in this group is very high. In contrast, the null proportion in group 1, for instance, is much lower. Hence, when the purpose is to maximize the number of discoveries (while controlling for the FDR at the given target), instead of rejecting all null hypotheses up to a single threshold as in a traditional approach, say 0.05, we could simply use a threshold of 0.2 in group 1 and 0.01 for group 3. This is reflected clearer in Figure 3.

Figure 3: Comparison of the procedures across groups. Panel A presents the proportion of rejections of each procedure whereas Panel B the corresponding maximal $p-$value of the those null hypotheses rejected in each group. Panel C, in contrast, shows the minimal $p-$value of those hypotheses whose the nulls are not rejected. The partition of hypotheses into groups is described in the Figure 2.



In Panel A of Figure 3, each line represents the proportion of null hypotheses rejected by the four procedures across groups 1 to 9. Here, the rejected proportion is the ratio of the number of null hypotheses rejected in a group over the number of null hypotheses of the group. It is clear that all procedures reject more null hypotheses in groups 1 and 9 than they do in groups 3, 5 and 7. Especially, we witness that the rejected proportions of the $mfFDR$ and $fFDR$ are much higher than those of the BH and $StdFDR$ in groups

[4] Since the $p-$values of the tests are independent, the ones of true null hypotheses are uniformly distributed and therefore producing a flat histogram. Consequently, a group with a skewed histogram contains more false nulls.

1 and 9 while the figures of the former are slightly less than that of the $StdFDR$ in group 5. In Panel B, we show the maximal $p-$value of those null hypotheses rejected by each procedure. As discussed, a null hypothesis is rejected by the BH and $StdFDR$ if its $p-$value is less than some threshold (these thresholds are 0.01 for the BH and 0.018 for the $StdFDR$, which virtually coincide with the green and brown lines in Panel B, respectively). Hence, the maximal $p-$values of the null hypotheses rejected in the nine groups for BH and $StdFDR$ are roughly the same. In contrast, $mfFDR$ rejects some null hypotheses having $p-$value up to more than 0.8 in group 1 while in group 5 it does not reject any null hypotheses having $p-$value more than 0.011, which is less than the significant threshold of the $StdFDR$. Thus, there are a few null hypotheses rejected by the $StdFDR$ but not by the $mfFDR$. However, while by partitioning the hypotheses into groups we intuitively see how $mfFDR$ works, it is not based on grouping hypotheses and establishing a particular rejection threshold for each group. As presented in Section 2.1, the rejection rule in $mfFDR$ is based on the $q-$value which calculated from the posterior probability that a null hypothesis is true, given its observed $p-$value and the covariates. The $mfFDR$ rejects more null hypotheses in groups where the false null ones are rich, and it is not using a $p-$value threshold for rejecting null hypotheses in a group. As noted from Panels B and C, it is not uncommon in a group, that a null hypothesis with high $p-$value is rejected while an another with a lower $p-$value is not. For instance, in group 1, there is a null hypothesis with a $p-$value of 0.03 that is not rejected by the $mfFDR$ while on the same group there are null hypotheses with $p-$values of more than 0.8 that are rejected.

2.2.2. Performance of $mfFDR$: FDR control and power comparison

In this section, we assess the two most important criteria of an FDR procedure: the control of the FDR and the power. In order to do so, we implement the $mfFDR$ and benchmark procedures at FDR targets $\tau \in \{0.05, 0.1, \ldots, 0.95\}$ over 1000 iterations and average the false discovery proportions and the correct detection proportions to have estimates of the actual FDR and the power, i.e. the expectation of the mentioned correct detection proportion, respectively. In total, we are studying six cases corresponding to the combinations of the two forms of the function $\pi_0(u, v)$ and the three aforementioned specifications of the Beta distribution. We also demonstrate that our method retains its power when the covariates are correlated with each other and are subject to estimation errors. Finally, we illustrate the FDR control and the power of $mfFDR$ when the $p-$values are weakly dependent (as for example, there will be when each hypothesis test represent a technical trading rule). To have an assessment for the first criterion, we compare the estimated actual FDR of the $mfFDR$ to the given FDR targets while for the second one we compare the power of $mfFDR$ against that of the $fFDR$, $StdFDR$ and BH approaches.

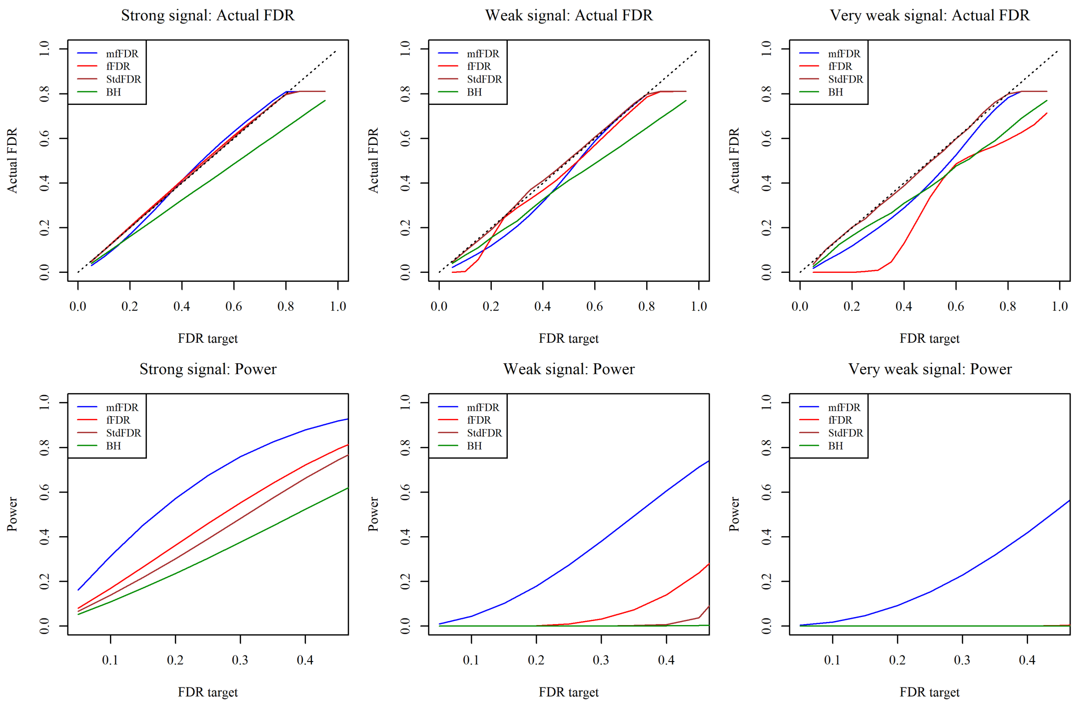
In Figure 4, we show the estimate of the actual FDR and the power of all procedures under the sine (Panel A) and monotonic (Panel B) form of $\pi_0(\boldsymbol{z})$. In each panel, the top three sub-figures exhibit the estimated actual FDR corresponding to the three cases of the signal (strong, weak and very weak). In each of those sub-figures, each line presents for the estimated actual FDR of a procedure at the given FDR targets. Ideally, a procedure perfectly (strictly) controls the FDR at a target if its estimated actual FDR at that target lies on (below) the 45° dotted line. For instance, all procedures either perfectly or strictly, control for FDR at target 20%. As controlling of FDR is a sample property, it is acceptable to observe a point positioning slightly above the dotted line since we estimate the actual FDR over only 1000 iterations. In general, from the sub-figures we see that all procedures control well for the FDR at any given targets and in all considered cases.

In terms of power, it makes sense to focus on only the FDR targets less than 0.5 (i.e. up to 0.45 in our simulation). Hence, the three bottom sub-figures in each panel present the power of the four methods for only those targets. In all cases, the lines representing for the power of the $mfFDR$ are always at the top regardless the form of $\pi_0(u, v)$ as well as the strength of signal. In other words, the $mfFDR$ beats its benchmarks in terms of power. Apparently, all procedures have higher power when the signal is strong. In this case, the $mfFDR$ has the gaps up to about 21%, 29% and 37% in comparison with the $fFDR$, $StdFDR$ and BH procedures, respectively. Those figures are 44%, 67% and 76% (57%, 62% and 62%) for the weak (very weak) signal case. At the FDR target of 20%, which will be used later in our main analysis, the gap of $mfFDR$ over $fFDR$ is varying from 10% to 20%. Finally, the weak and the very weak signal cases highlight the benefit of using the $mfFDR$ when the data has a low signal to noise ratio. While the $fFDR$, $StdFDR$ and BH procedures can hardly detect a single false null hypothesis even at the FDR target of 20% (see the very weak signal case), the $mfFDR$ quickly gains a significant power of more than 10%.

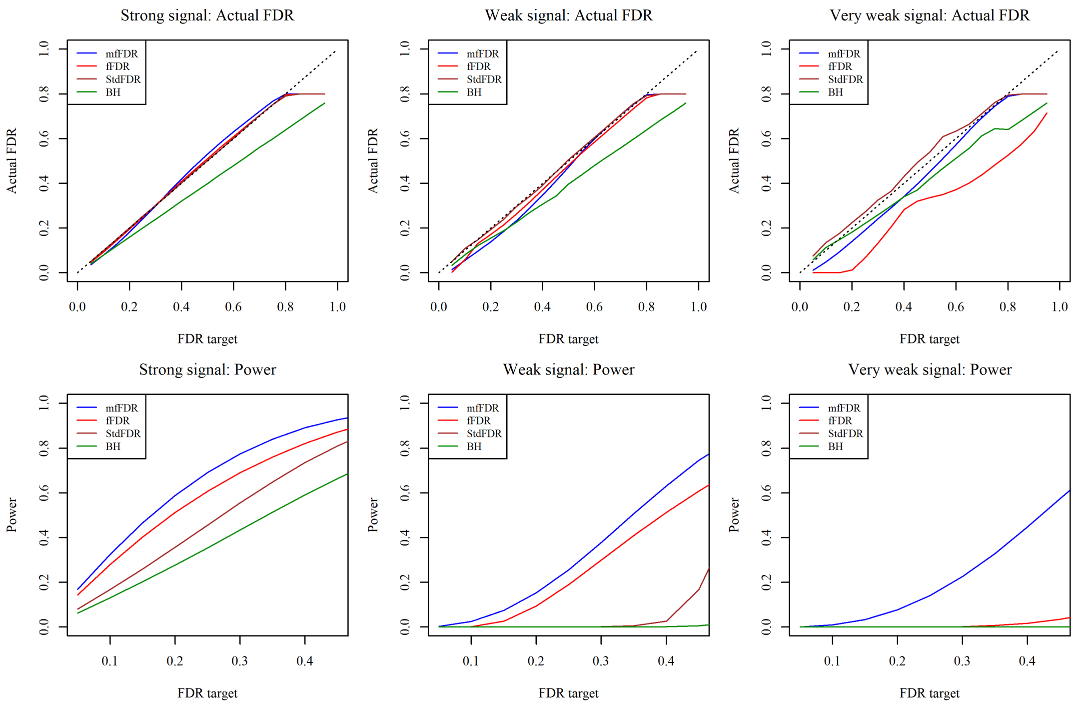
To sum up, we have developed $mfFDR$ which enables the use of multiple informative covariates in detecting false null hypotheses and compare our method to the FDR procedures of Storey (2002) and BH which do not use covariates. We have shown that the power of the former largely surpasses that of the latter ones while controlling well for the FDR at any given targets. In other words, we are able to detect more outperforming

Figure 4: Performance comparison of the $mfFDR$, the $fFDR$, the Standard FDR of Storey ($SdtFDR$) and the FDR procedure in BH. Panel A shows the performance when the $\pi_0(u, v)$ has a sine form whereas in Panel B is the monotonic one.

Panel A: $\pi_0(u, v)$ is a sine function.



Panel B: $\pi_0(u, v)$ is monotonic with respect to each covariate.



technical rules under a testing framework that is conditional on multiple information (i.e., $mfFDR$) than the ones that are unconditional or using less information. On the other hand, we also show that when more than one informative covariates that are mutually independent exists, the $mfFDR$ gains remarkably higher power than the $fFDR$, which is the $mfFDR$ with $d = 1$, especially when the signal of the false null hypothesis is weak.

2.3. Correlation and estimation errors of covariates

Hitherto, we have studied the performance of $mfFDR$ where the covariates are independent drawn from a uniform distribution. In financial data, this is unlikely the case. In order to take this issue into account, in this section we design a simple model where the covariates are positively correlated.[5] Given a correlation coefficient r, the two covariates studied in previous section are transformed (with use of Cholesky factorization) to two new covariates having a correlation coefficient approximately r. The simulated data then are generated similarly to the previous section. For interest of space, we present the results for only the sine form of the null proportion function.[6] Aiming to study the impact of the correlation in covariates on the performance of the $mfFDR$, we consider varying values of r from 0.1 to 0.8.

Figure 5 depicts the performance of the $mfFDR$ in terms of FDR control (Panel A) and its power compared to others (Panels B and C). For the former aspect, the FDR is well controlled when $r < 0.7$ regardless the level of the signal. This range of the coefficient covers most cases in our real data in which, as being shown in Section 5, more than 95% of the empirical coefficients are having absolute value less than 0.7. When $r \geq 0.7$ the FDR is controlled well under the weak and very weak signals and asymptotically controlled in the strong case.

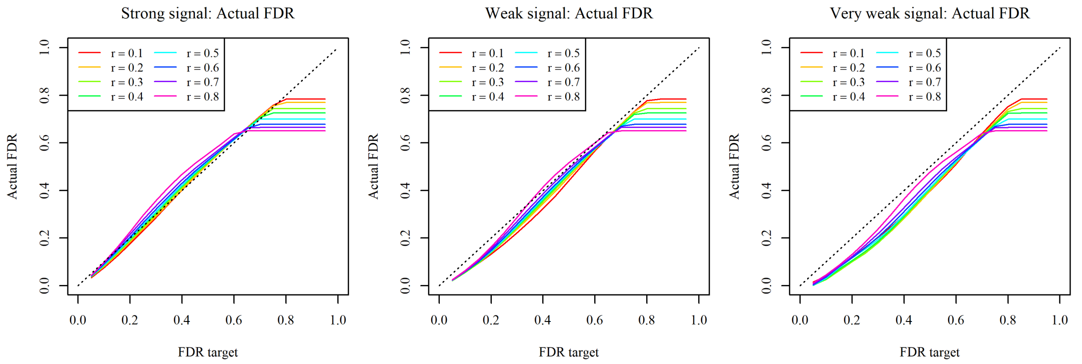
It is noted that, the powers of $mfFDR$ under different correlation coefficients are incomparable due the differences in the level of signal. That is, the signal in higher correlation settings might be stronger due to the transformed covariates as the $p-$values of tests are generated based on them. Consequently, to assess the performance in terms of power, we calculate the gaps in power of the $mfFDR$ over the $fFDR$ and $StdFDR$ under each case of r and depict them in panels B and C, respectively. In sub-figures of these panels, a line above zero indicates that a higher power of $mfFDR$ in comparison to its benchmarks. This is the case for all of considered settings with the peak varying

[5]The performances of the procedures under negatively correlated covariates is similar.

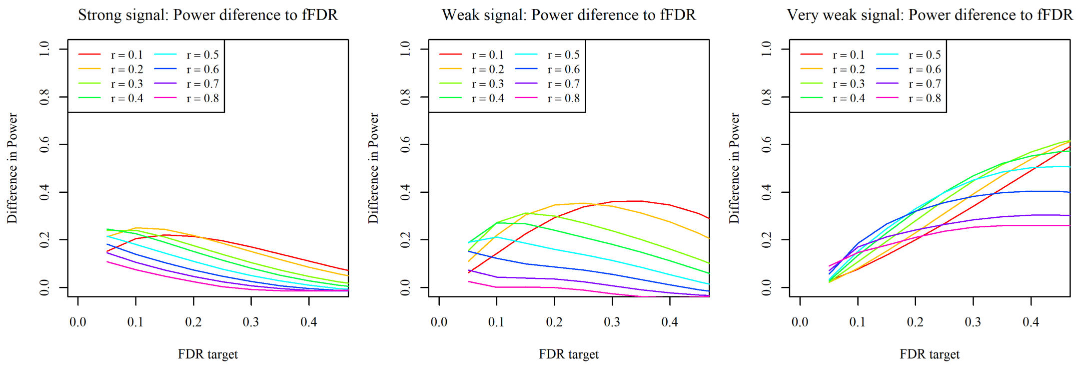
[6]These two new covariates are not uniformly distributed on $[0, 1]$, thus, the null proportion function is slightly modified, $\pi_0(u, v) = \min\{1, \max\{0, \sin(\pi(u + v)/2)\}\}$, so that its values are in $[0, 1]$.

Figure 5: Performance of $mfFDR$ under correlated covariates.

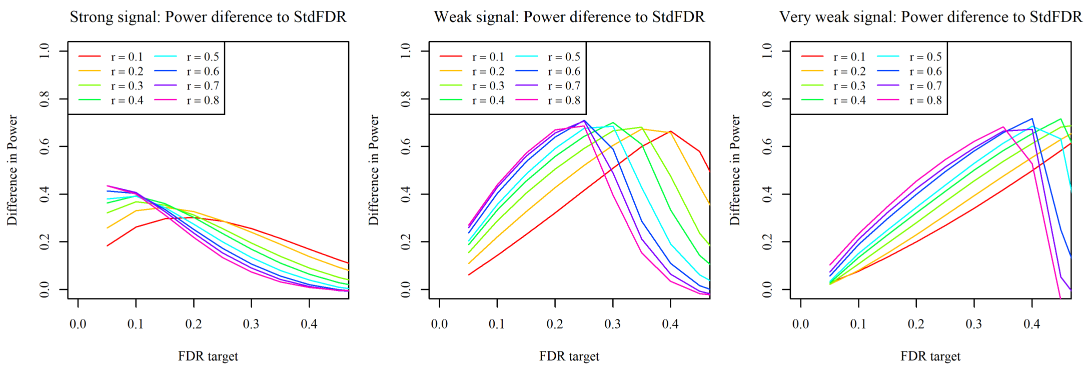
Panel A: FDR control of $mfFDR$ under correlated covariates



Panel B: Gap in power of $mfFDR$ over $fFDR$ under correlated covariates



Panel C: Gap in power of $mfFDR$ over $StdFDR$ under correlated covariates



across cases and could reach to about 70% or 30%, under FDR target of 20%, when the benchmark is the $StdFDR$ or $fFDR$, respectively.

Alongside with the concern in the correlation of covariates, the estimation errors of covariates (i.e., the noise in the estimating covariates) also potentially affect the performance of the $mfFDR$. In empirical finance, the covariates will be estimated quantities and are thus subject to estimation errors. To address this concern, we additionally conduct simulations where the covariates used as input contain noise. For interest of space, we defer the results of this experiment to Appendix B. Generally, we find that $mfFDR$ still controls FDR well. Perhaps not surprising, its power is lower than the prior case with uncorrelated covariates as presented in previous section. Nevertheless, the power of $mfFDR$ is still remarkably higher than that of other methods. As a result, the proposed $mfFDR$ method is reasonably robust to correlation and estimation biases.

2.4. Weak dependence in p-values

In developing theoretical $mfFDR$ framework, we assume that the tests are independent replications of the test (1). In financial applications, this scenario is unlikely to be the case. In this particular study, we will consider hypothesis tests comparing the Sharpe ratio of technical trading rules against zero. In a particular category of trading rule, the rules with close parameters tend to have highly correlated returns. This leads to a weak dependency among the testing statistics or $p-$values of the corresponding hypotheses. In this section, we show that our method is robust under this type of dependence. Specifically, we are generating data such that the hypotheses are partitioned into groups with the same size k. In each group, the $p-$values of testing hypotheses are mutually dependent at the same level which is characterized by a covariance matrix $\mathbf{\Sigma}$. The $p-$values corresponding to the hypotheses from different groups are independent.

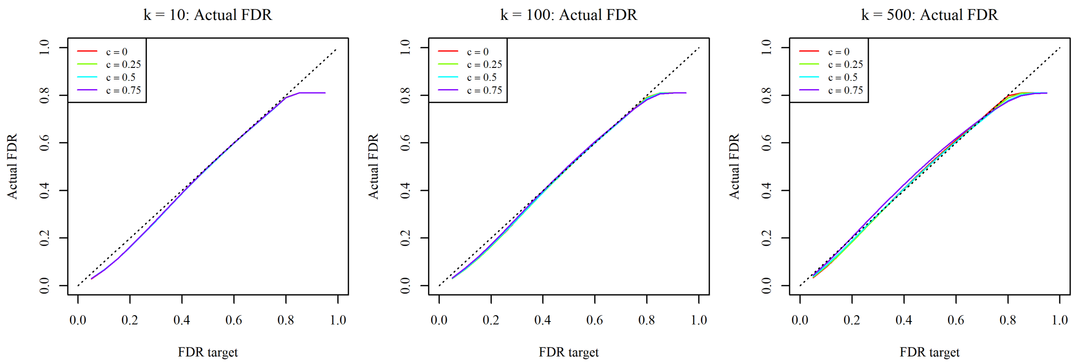
The data generating process is as following. The simulated covariates, $\pi_0(u, v)$ and the true status of null hypotheses H are as described in Section 2.2. For the sake of space, we present only the results for the sine form of $\pi_0(u, v)$. In order to account for weak dependence in hypotheses, we fist partition the true null hypotheses into groups of size k. Secondly, we generate the $z-$scores for the null hypotheses of each group from the multivariate normal distribution $\mathcal{N}(0, \mathbf{\Sigma})$. We process similarly for the false null ones, but the $z-$scores of each groups are drawn from $\mathcal{N}(2, \mathbf{\Sigma})$. To simplify, the matrix $\mathbf{\Sigma} = (\Sigma_{ij})_{k\times k}$ is set as $\Sigma_{ii} = 1$ and $\Sigma_{ij} = c$ for $i \neq j = 1, \ldots, k$ for some c. By considering various values of the parameters k and c, we reveal the impact of the weak dependence in different levels on the performance of the $mfFDR$. Here, we choose $c \in \{0, 0.25, 0.5, 0.75\}$ where the case $c = 0$ indicates the absence of the dependence among $p-$values and will be used as a benchmark for a comparison purpose, and $k \in \{10, 100, 500\}$. The parameter k represents the dependence scale. A larger k indicates the presence of more hypotheses that are mutually dependent.[7] Finally, the $p-$value of each (two-sided) test is calculated from its $z-$score by using the cumulative distribution function of the standard normal

[7] This type of weak dependent setting is also studied in Storey (2003) with $k = 10$. Here we extend the cases of k aiming to study its impact on the performance of the method.

distribution. Note that, in this setting, the covariates convey only the information on the probability of being true null of the hypotheses and not on the $p-$values as in the simulations conducted in the previous sections.

Figure 6: Performance of the $mfFDR$ under weakly dependent $p-$values.

Panel A: FDR control of the $mfFDR$ under weakly dependent $p-$values



Panel B: Power of the $mfFDR$, $fFDR$ and $StdFDR$ under weakly dependent $p-$values

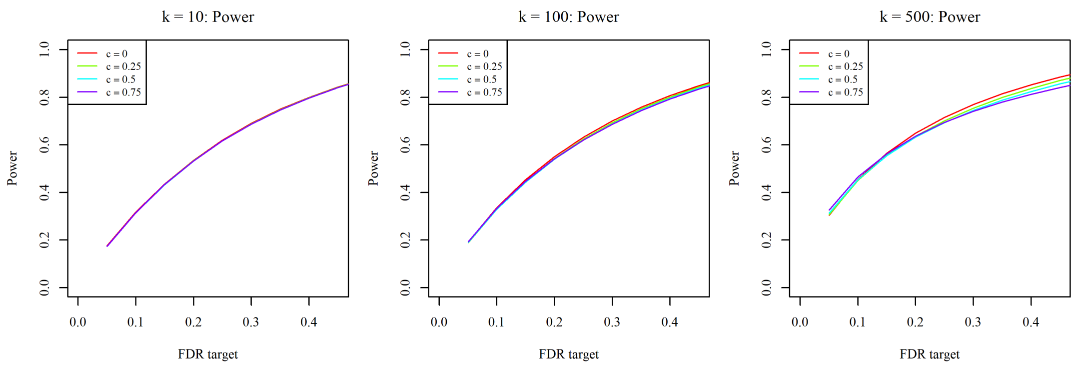


Figure 6 presents the performance of the $mfFDR$ under the settings of c and k. Panel A of the figure shows that the dependence does not affect the FDR control of the method. In terms of power in Panel B, when k is small, the first two sub-figures show an insignificant difference between the power of the $mfFDR$ with different levels of dependence among $z-$scores in each group. Therefore, the lines corresponding to different values of c are virtually identical and covered by a purple line. Their powers are only distinguishable in the third sub-figure, where k is large and for high targets of FDR. In this case, the power is decreasing concerning c, i.e. the level of dependence level among the $z-$scores.

In this section, our simulations cover most concerns about applying our $mfFDR$ method to financial data. We find that the $mfFDR$ performs well in terms of FDR control and power with correlated covariates, under weakly dependent $p-$values, and with estimation errors in covariates.

3. Data and strategy universe

3.1. Data

We collect daily spot and 1-month forward exchange rates data against the U.S. dollar for 30 countries, including nine developed markets (Australian dollar, Canadian dollar, German mark/euro, Japanese yen, New Zealand dollar, Norwegian krone, Swedish krona, Swiss franc, and U.K. pound) and 21 emerging markets (Argentine peso, Brazilian real, Chilean peso, Colombian peso, Czech koruna, Hungarian forint, Indian rupee, Indonesian rupiah, Israeli shekel, Korean won, Mexican peso, Philippine peso, Polish zloty, Romanian new leu, Russian ruble, Singaporean dollar, Slovak koruna, South African rand, Taiwanese dollar, Thai baht, and Turkish lira). The sample periods for developed market currencies start from January 4, 1971 and end on December 31, 2020, while the sample periods for emerging market countries start from various dates due to data availability on both exchange rates and short-term interest rates. Israel has the earliest starting date among emerging market currencies (January 1978) and is followed by South Africa (January 1981), Singapore (January 1982) and Taiwan (October 1983); all emerging market data end in December 2020. Our data on exchange rates and short-term interest rates were kindly supplied by the London branch of the asset manager BlackRock and are based on midday quotations in the London market.

Table 1 presents summary statistics of gross currency returns and short-term interest rates of developed and emerging economies. We define the gross currency return that is the return from buying a foreign currency unit and holding it for one day, $r_t = ln(s_t/s_{t-1})$, where s_t represents the spot exchange rate on day t. We define the spot exchange rate as units of U.S. dollars per foreign currency so that an increase of s is associated with an appreciation of the foreign currency. We report the mean, the standard deviation, the minimum and the maximum and the corresponding sample period. We find that gross returns tend to be more volatile for emerging economies offering higher minimum and maximum values.

Another important aspect of our analysis is the short-term interest rate as it affects the overall return of the trading strategies even though technical analysis focuses more exchange rate fluctuations. We convert our annual short-term interest rate (i^a) into daily data $i_t = ln(1 + i^a)/360$. In Table 1, we report the mean, the standard deviation, the minimum and the maximum values of short-term interest rates for both developed and emerging countries. We find that short-term interest rates for developed countries range

Table 1: Summary Statistics. The table presents summary statistics of gross returns on foreign currencies and short-term interest rates. We report the mean, the volatility, the minimum and the maximum. We also report the sample period for each currency in our sample. We show results for developed and emerging economies.

Countries	Gross returns on foreign currencies				Short-term interest rates				Sample Period
	mean(%)	vol	min	max	mean(%)	vol(%)	min(%)	max(%)	
Developed									
Australia	-0.0028	0.0068	-0.1925	0.1073	0.0191	0.0119	0.0001	0.1709	1/4/1971-12/31/2020
Canada	-0.0018	0.0041	-0.0434	0.0505	0.0172	0.0125	0.0005	0.0596	1/4/1971-12/31/2020
Germany/E.U.	-0.0017	0.0058	-0.0421	0.0462	0.0092	0.0078	-0.0020	0.0344	1/4/1971-12/31/2020
Japan	0.0095	0.0063	-0.0626	0.0950	0.0108	0.0124	-0.0006	0.0406	1/4/1971-12/31/2020
New Zealand	-0.0033	0.0073	-0.2050	0.0995	0.0194	0.0111	-0.0007	0.1286	1/4/1971-12/31/2020
Norway	-0.0014	0.0066	-0.0814	0.0646	0.0163	0.0106	0.0023	0.1372	1/4/1971-12/31/2020
Sweden	-0.0036	0.0065	-0.1507	0.0555	0.0171	0.0139	-0.0011	0.2090	1/4/1971-12/31/2020
Switzerland	0.0122	0.0073	-0.0892	0.1267	0.0045	0.0098	-0.0086	0.2276	1/4/1971-12/31/2020
U.K.	-0.0044	0.0058	-0.0848	0.0467	0.0160	0.0112	0.0004	0.0460	1/4/1971-12/31/2020
U.S.	-	-	-	-	0.0128	0.0097	-0.0000	0.0466	1/4/1971-12/31/2020
Emerging									
Argentina	-0.0579	0.0098	-0.3418	0.1712	0.0343	0.0368	0.0035	0.4144	4/1/1991-12/31/2020
Brazil	-0.0238	0.0097	-0.1080	0.1178	0.0428	0.0283	0.0051	0.2792	1/6/1992-12/31/2020
Chile	-0.0070	0.0064	-0.1160	0.1114	0.0014	0.0016	0.0000	0.0214	1/1/1991-12/31/2020
Colombia	-0.0224	0.0062	-0.0593	0.0562	0.0315	0.0240	0.0052	0.0908	1/3/1986-12/31/2020
Czech	0.0040	0.0073	-0.0707	0.0522	0.0104	0.0133	0.0002	0.2628	1/4/1978-12/31/2020
Hungary	-0.0177	0.0080	-0.0842	0.0520	0.0297	0.0227	0.0021	0.0834	1/2/1987-12/31/2020
India	-0.0178	0.0046	-0.1281	0.0376	0.0254	0.0149	0.0002	0.1944	9/1/1994-12/31/2020
Indonesia	-0.0276	0.0130	-0.3576	0.2361	0.0296	0.0242	0.0000	0.2054	1/2/1997-12/31/2020
Israel	-0.0683	0.0059	-0.1725	0.0645	0.0555	0.0962	0.0002	0.6309	4/27/1993-12/31/2020
South Korea	-0.0047	0.0078	-0.1809	0.2012	0.0154	0.0138	0.0012	0.0676	7/4/1994-12/31/2020
Mexico	-0.0349	0.0114	-0.2231	0.2231	0.0403	0.0416	0.0077	0.3387	1/3/1994-12/31/2020
Philippines	-0.0096	0.0044	-0.0860	0.1015	0.0198	0.0165	0.0014	0.1962	4/22/1992-12/31/2020
Poland	-0.0117	0.0079	-0.0715	0.0670	0.0212	0.0198	-0.0004	0.0792	6/4/1991-12/31/2020
Romania	-0.0368	0.0091	-0.3887	0.0953	0.0407	0.0518	0.0006	0.3667	1/3/1992-12/31/2020
Russia	-0.0515	0.0132	-0.3863	0.2779	0.0288	0.0391	0.0022	0.3208	1/1/1987-12/31/2020
South Africa	-0.0286	0.0097	-0.1030	0.1440	0.0289	0.0115	0.0000	0.0588	6/4/1993-12/31/2020
Singapore	0.0043	0.0033	-0.0276	0.0414	0.0053	0.0050	0.0000	0.0181	1/2/1981-12/31/2020
Slovakia	0.0021	0.0063	-0.1097	0.0462	0.0130	0.0164	0.0001	0.2068	1/4/1982-12/31/2020
Taiwan	0.0035	0.0029	-0.0420	0.0430	0.0082	0.0067	0.0004	0.0483	10/4/1983-12/31/2020
Thailand	-0.0022	0.0055	-0.2077	0.0635	0.0102	0.0107	0.0004	0.0664	1/2/1991-12/31/2020
Turkey	-0.1002	0.0109	-0.3348	0.1256	0.0833	0.0643	0.0130	1.0328	2/1/1990-12/31/2020

between 0.45 to 1.94 basis points on average. We also find that short-term interest rates are higher for most of the currencies of the emerging markets category and they exhibit a particular variation. The highest interest rate in this group is for Turkey which is 8.33 basis points and the lowest is for Chile which offers 0.14 basis points.

3.2. Trading rule universe

Based on past daily spot exchange rates, a trading rule determines positions (long, short or neutral) that traders should take in the next day. In this study, we assess the universe of trading rules used in Hsu *et al.* (2016), which consists of the following five family trading rules widely used by traders.

The filter trading rules: The rules generate a long (short) position whenever the closing exchange spot rate has risen (fallen) by a given percentage above (below) its most recently high (low). This family rules is generally based on momentum of the exchange rate where traders believe the rising (falling) rate continue rises (fall). The threshold percentages are set so that the traders are not misled by the small fluctuations.

Moving average trading rules: These rules generate long or short positions based on comparing the closing spot exchange rate to one or three simple moving averages of given different lengths or comparing the moving averages of two different lengths. For example, the simplest moving average rule generates a long (short) position when the spot exchange rate moves up (down) at least a certain percent above (below) the moving average of a specific length.

Relative strength indicators: The relative strength indicator is a popular form of oscillators, which aims to identify imminent market corrections after rapid exchange rate movements. Generally, the indicator of a given length has values from 0 to 100 and generate an oversold or overbought signal when it crosses a predetermined lower or upper extreme, respectively.

Support-resistance rules: These rules rely on determining of a support or resistance level, for which the exchange rate appears to have difficulty in penetrating in a previous given number days, and a premise that when the closing exchange rate breaches the level, it will trigger further movement in the same direction.

Channel breakouts: The rules establish time-varying support and resistance levels, forming a trading channel with upper and lower bounds. Once a bound is breached, a long or short position is initiated in the similar way as the support-resistance rules.

Given the described family rules, a number of their variants are generated by varying plausible parameters, the delay time and the position's holding time. Ultimately, we obtain 2,835 filter rules, 12,870 moving average rules, 600 relative strength indicators, 1,890 support-resistance rules and 3,000 channel breakout rules, making up 21,195 trading rules in total. Readers are referred to Appendix A in Hsu *et al.* (2016) for the detail specifications of our technical rules.

4. Performance measure

In this study, we distinguish the excess return and net excess return gained by a trading strategy before and after counting for transaction cost, respectively. The excess return from buying one unit of foreign currency (against U.S. dollar) and holding it for one day is calculated as the summation of returns due to appreciation/depreciation of the foreign currency and the return obtained from lending the money in foreign currency, minus the benchmark return which is the return would gain if the money is deposited in the U.S., that is

$$r_t = \ln(s_t/s_{t-1}) + \ln(1 + i^*_{t-1}) - \ln(1 + i_{t-1}) \tag{5}$$

where s_t and s_{t-1} are spot rates at the midday of the days t and $t-1$, respectively, and i_{t-1} and i^*_{t-1} are daily interest rates on U.S. dollar deposits and foreign currency deposits on day $t-1$, respectively.

The daily excess return of the trading rule j, earned from day $t-1$ to day t, $R_{j,t}$, is determined as

$$R_{j,t} = S_{j,t-1}.r_t \tag{6}$$

where $S_{j,t-1}$ is position guided by the trading rule based on past data up to day $t-1$ and taking value in the interval $[-1, 1]$. For most of the trading rules in our universe, $S_{j,t-1}$ takes value $+1$ for the long, (-1) for the short and 0 for the neutral position on the foreign currency. For some moving average trading rules, where the signal guides for positioning one-third of funds on a long while the rest is kept in U.S. dollar or on a short position while the rest is kept in foreign currency, $S_{j,t-1}$ takes value $+1/3$ or $-1/3$, respectively.

In this study, we assess the performance of a trading rule j based on its Sharpe ratio.[8] From day T_1 to day T_2, it is defined as

$$SR_j = \frac{\overline{R}_j}{\sigma_j}, \tag{7}$$

where $\overline{R}_j$ and σ_j are the mean and standard deviation of excess returns of the trading rule over the mentioned period, that are $\overline{R}_j = \sum_{t=T_1}^{T_2} R_t/N$, $\sigma_j = \sqrt{\sum_{t=T_1}^{T_2} (R_{j,t} - \overline{R}_j)^2/(N-1)}$ where $N = T_2 - T_1 + 1$, respectively.

When a trading rule changes its guiding signal, a new or existing position is triggered or closed and transaction costs occurs consequently. The net excess return in a daily basis, therefore, is the daily excess return less the cost caused by the transactions and the net excess return and Sharpe ratio after transaction cost of a trading rule are calculated accordingly. As in Neely and Weller (2013), we use one third of the quoted one-month forward rate bid-ask spread in each currency as an estimate of the one-way transaction

[8]We repeat empirical experiments with use of the mean excess return as an alternative performance metric. Our following conclusions remain persistent as exhibited in Section IC of the Internet Appendix.

costs on any particular day. For periods before the forward data is available, we set fixed transaction costs for each period in the same way as in Neely and Weller (2013): for developed country currencies we set the transaction cost at a flat 5 basis points in the 1970s, 4 basis points in the 1980s and 3 basis points in the 1990s, and for emerging market currencies we set the daily cost at one third of the average of the first 500 bid-ask observations.

Table 2: Summary performance of trading rules in whole sample period. The table shows the number of significant rules, which are those having bootstrap p-value < 0.05 and positive estimated performance, the highest performing rule in each currency with its performance. We report two performance metrics - the Sharpe ratio and the mean of net excess return (both are annualized) which is the excess return after transaction cost."*", "**" and "***" respectively indicate statistical significance at levels of 10%, 5% and 1%.

	Mean net return			Sharpe ratio		
	Number of significant rules	Highest return (%)	Best rule	Number of significant rules	Highest ratio	Best rule
Australia	900	5.73***	F3	1155	0.54***	F3
Canada	196	2.67***	MA4	217	0.41***	MA4
Germany/E.U.	8585	5.33***	F3	8536	0.66***	MA5
Japan	7862	6.13***	MA4	7954	0.61***	MA4
New Zealand	6204	6.14***	MA1	6747	0.56***	MA1
Norway	443	4.24***	SR1	449	0.41***	SR1
Sweden	6827	5.99***	MA4	7038	0.59***	MA5
Switzerland	601	5.05***	F3	599	0.48***	CB1
U.K.	2161	4.58***	MA2	2217	0.54***	CB1
Argentina	9254	11.22***	SR1	10973	0.77***	MA5
Columbia	3915	11.04***	MA1	4025	1.17***	MA5
India	1844	4.44***	F3	2287	0.67***	F3
Indonesia	1744	10.71***	MA4	2401	0.56***	MA5
Israel	15271	9.26***	MA1	15452	1.22***	MA5
Philippines	6401	4.74***	SR1	6444	0.69***	MA5
Romania	67	5.53***	F3	118	0.52***	MA5
Russia	3897	15.63**	F1	6246	0.83***	MA5
Slovak	1331	5.69***	MA4	1348	0.58***	MA4
Brazil	4938	10.97***	SR1	6623	0.77***	MA5
Chile	1545	7.16***	SR1	1431	0.85***	MA5
Czech	107	5.87***	MA4	114	0.52***	MA5
Hungary	34	5.39***	SR1	80	0.45***	RSI
Korea	200	8.71***	F3	456	0.7***	F3
Mexico	8	4.66**	SR1	190	0.38***	CB1
Poland	39	6.3***	F3	74	0.51***	F3
Singapore	739	2.45***	CB1	673	0.56***	MA5
South Africa	914	7.82***	CB1	1013	0.53***	CB1
Taiwan	9782	4.3***	MA3	9827	1***	MA5
Thailand	3850	6.85***	MA1	3842	0.87***	MA5
Turkey	13516	16.07***	F3	14568	1.1***	MA5

Table 2 captures important features of trading rules in terms of performance corresponding to each exchange rate, after transaction costs. It is noted that the number significant rules in this particular table is merely based on conventional $p-$value of 5%. If all rules are independent random walks, one should expect there are roughly 530 significantly out-performing rules for each exchange rate.[9] With both performance measures under study, there are 22 over 30 exchange rates having more than 530 significant rules. This fact implicitly indicates that in those exchange rates, there are a portion of technical trading rules that are truly profitable. All the best rules of the 30 exchange rates are significant with five of them gaining more than 10% per annum and four with Sharpe ratio at least one. However, we do not know if these rules are truly profitable or just lucky before we implement tests to correct data snooping bias.

5. Empirical results

In this section, we discuss our covariates in technical trading rules and the information they potentially convey. Based on these covariates, we construct rolling portfolios and show the benefit of using the $mfFDR$ in selecting outperforming rules on each currency separately and to all currencies combined. As documented in Neely and Weller (2013), the performance of technical trading rules in the foreign currency market is persistent. Consequently, by investing on truly profitable strategies in a in-sample period, an investor should be able to generate an out-of-sample (OOS) profit. In this section, we are demonstrating the performance of the $mfFDR$ in detecting out-performing trading rules. For this purpose, we introduce a procedure similar to the functional false discovery "plus" ($fFDR^+$) documented in HKMS. Thereby, to control FDR in a group of significantly outperforming trading rules at a given target, we implement the $mfFDR$ on the group of positive estimated SR trading rules to control for FDR at the given target. We name this procedure the multivariate functional false discovery rate "plus" ($mfFDR^+$). For interest of space, in the followings, we present the results with control of FDR at target 20%. In Section IB of the Internet Appendix, we show that our conclusions are robust for other FDR targets.

5.1. Covariates

We apply within our $mfFDR$ framework a set of informative continuous variables that are generated by the return of the trading rules. These covariates reflect the performance

[9]That is, as the $p-$values of true null hypotheses has a uniform distribution, there are approximate $21195 \times 0.05/2 \approx 530$ significantly out-performing rules.

persistence, financial risk, activeness and out-performance compared to a passive strategy of the trading rules under study. As a first covariate we consider the auto-correlation of trading rule's return (ρ) or else, its performance persistence. For the remaining covariates and for each individual currency, we regress the excess return of a trading rule with the excess return from the corresponding buy-and-hold strategy on a particular currency

$$r_{i,t} = \alpha_{i,bh} + \beta_{i,bh} r_{bh,t} + \varepsilon_{bh,t} \tag{8}$$

where the $r_{i,t}$ is the excess return of the trading rule i, $r_{bh,t}$ is the excess return of the buy-and-hold strategy and $\epsilon_{bh,t}$ is the noise at day t. The $\alpha_{i,bh}$, $\beta_{i,bh}$ and R–square of regression (8), respectively, represent the alpha and the riskiness level the activeness of the trading rule i compared to the buy-and-hold strategy. We denote the three covariates as α_{bh}, β_{bh} and R^2_{bh}.

When the target is to select outperforming trading rules from the universe of all trading rules of all currencies (or else, when we examine all currencies together), which consists of $21195 \times 30 = 635850$ rules, we compare a particular trading rule to the average currency excess return factor of Lustig *et al.* (2011), denoted by RX. That is, the return of strategy that invests equally weighted on all 30 currencies in our sample (henceforth, currency market factor). In these cases, instead of using regression (8), the latter three covariates are obtained from the regression model:

$$r_{i,t} = \alpha_{i,mk} + \beta_{i,mk} RX_t + \varepsilon_{mk,t} \tag{9}$$

where the RX_t is the currency market factor on the day t. Analogously, we denote the three new covariates as α_{mk}, β_{mk} and R^2_{mk}, respectively.

In the following analyses, the informativeness of the covariates will be revealed and the benefit of using multiple covariates under $mfFDR$ framework is highlighted. We first restrict constructing a portfolio on each currency where the covariates ρ, α_{bh}, β_{bh} and R^2_{bh} are used. Then, we examine all currencies together and we apply as covariates ρ, α_{mk}, β_{mk} and R^2_{mk}. Finally, we investigate the mobility of profitable rules across the currencies.

5.2. Individual Currencies

We construct monthly rolling portfolios which use one year daily data as in-sample. More specifically, at the end of each month, we utilize the past 12 months data up to that point of time as the in-sample data to calculate the $p-$value and covariates of the trading rules. We then implement the $mfFDR^{+}$ to detect outperforming strategies in the in-sample period at the FDR target of 20%. We combine the signals of these outperforming rules to determine the position of that day by neutralizing the opposite ones. For instance, suppose there are 100 trading rules are selected and 20 of them indicate a buy signal with a weight +1, 30 of them indicate sell signals with weight $-1/3$ and the remains 50 are neutral signals. After combining, we have $10(= 20.(+1) + 30.(-1/3))$ long signals. Thus, the trader takes a long position with $1/10(= 10/100)$ of her/his funds.[10] We follow the signals of these portfolios determine the position of each trading day in the following month (i.e., OOS period). We benchmark our results by those generated by the FDR^{+} and $fFDR^{+}$ with one covariate. The input data of all procedures are based on the excess return (before transaction cost) and we assess the OOS performance by the portfolio's net excess return (after transaction cost).

As a first exercise, we examine each currency individually. Therefore, the covariates of $mfFDR^{+}$ are $r^2_{bh}, \alpha_{bh}, \beta_{bh}$ and ρ as described in Section 5.1.

Table 3 provides the quantiles of the correlated coefficients of pairs of input covariates and those after combining all coefficients of the pairs as a pool. Those numbers capture an overview on the dependencies among the input covariates. The final row indicates that a majority of 98% coefficients are in $[-0.69, 0.59]$ for which the $mfFDR$ control

Table 3: The pairwise correlation of covariates. The table presents the quantiles of the correlation coefficients (calculated monthly with use of one-year in-sample data) of six covariate pairs which are combinations of the four covariates: $r^2_{bh}, \alpha_{bh}, \beta_{bh}$ and ρ. The final row shows the numbers for the set of all correlated pairs.

Covariate pairs	min	1%	5%	25%	50%	75%	95%	99%	max
r^2_{bh}, α_{bh}	-0.73	-0.44	-0.32	-0.12	0.05	0.21	0.40	0.54	0.70
r^2_{bh}, β_{bh}	-0.82	-0.75	-0.68	-0.51	-0.25	0.18	0.59	0.70	0.81
r^2_{bh}, ρ	-0.92	-0.68	-0.40	-0.09	0.02	0.12	0.28	0.47	0.91
α_{bh}, β_{bh}	-0.81	-0.54	-0.39	-0.20	-0.06	0.09	0.30	0.45	0.72
α_{bh}, ρ	-0.86	-0.45	-0.22	-0.10	-0.04	0.02	0.12	0.21	0.72
β_{bh}, ρ	-0.52	-0.32	-0.19	-0.06	0.01	0.08	0.22	0.34	0.58
All pairs	-0.92	-0.69	-0.51	-0.15	-0.02	0.10	0.36	0.59	0.91

[10]This approach is based on the idea of the $1/N$ portfolio strategy, where we invest equally funds into each of the selected rules. We choose this method for three reasons; First, DeMiguel *et al.* (2007) show that such approach is hard to be beaten by more sophisticated approaches (e.g. approaches weight funds differently on selected rules). Second, it reflects directly the performance of its components (e.g., the selected rules). Lastly, the funds allocated to opposite signals should be neutralized to avoid transaction costs. In a different context, Burnside *et al.* (2011) follow a similar approach to construct carry trade and momentum portfolios and Filippou *et al.* (2018) build an equally-weighted portfolio of a trading rule that is based on political risk.

very well for any FDR targets as shown in simulations in Section 2.3. We emphasize that there are only 1% of the coefficients having absolute value from or above 0.7.

In Table 4, we present the average number of trading rules selected by FDR^{+} and $mfFDR^{+}$, controlling for FDR at 20%, at the end of each month in different sample periods. Here, the superior in power of the $mfFDR^{+}$ is shown empirically, that is, controlling at the same level of FDR, the new approach detects more outperforming rules, compared to the FDR^{+}. This finding strongly holds for all currencies both in the whole period and in the sub-periods.

In Table 5, we present the performance of the portfolios including the FDR^{+}, $fFDR^{+}$ and $mfFDR^{+}$ in terms of Sharpe ratio based on the portfolios' excess return before transaction cost. We also study the performance of $fFDR^{+}$ having as covariate the first principal component (PC1) of the four aforementioned covariates. We see that most of portfolios are able to produce positive Sharpe ratio. The $mfFDR^{+}$ surpasses all procedures in terms of Sharpe Ratio. The performance of the $fFDR^{+}$ portfolio with PC1 as a covariate does not surpass $fFDR^{+}$ based portfolios with the original individual covariates. A finding that implies that our procedure can exploit the non-linearities between the four covariates and extract useful new information in detecting outperforming rules.

Table 6 presents a similar information but based on portfolios' excess return after the transaction cost. On average, the Sharpe ratio of the FDR^{+} portfolio now is indistinguishable from zero. This implies that, the portfolios selected by this procedure having a high turnover ratio. The final row of the table present the $t-$statistic comparing the mean of Sharpe ratios of the $fFDR^{+}$ and $mfFDR^{+}$ portfolios across currencies to that of the FDR^{+}. All covariate augmented portfolios statistically significantly outperform the FDR^{+}. The $mfFDR^{+}$ with use of the four covariates performs best. This portfolio gains positive Sharpe ratio for all currencies except the Hungarian forint and Polish zloty. Eighteen of those positive Sharpe ratios are statistically significant. The implications from the two tables 5 and 6 are threefold. First, the covariates are informative in a way that they help us to discover more out-performing rules and after combining, the obtained one efficiently reduces the trading frequency. Second, the more covariates we use, the more the efficiency we achieve. Third, our method can exploit the non-linearities between the four covariates and extract new information that is valuable in detecting out-performing technical rules.

Table 4: Empirical power comparison. The table presents the average number of trading rules in portfolios FDR^{+} and $mfFDR^{+}$ at beginning of each month, control for FDR at 20%. The first five columns report the numbers in five sub-periods while the last column shows those numbers across the months from the first time forming portfolios till December 2020.

	Period											
	1973-1980		1981-1990		1991-2000		2001-2010		2011-2020		1973-2020	
Currency	FDR^{+}	$mfFDR^{+}$	FDR^{+}	$mfFDR^{+}$	FDR^{+}	$mfFDR^{+}$	FDR^{+}	$mfFDR^{+}$	FDR^{+}	$mfFDR^{+}$	FDR^{+}	$mfFDR^{+}$
Australia	4138	8710	2451	8804	359	6615	1104	8095	114	6191	1530	7636
Canada	4603	8089	1349	6810	335	5826	370	6497	1099	5751	1433	6533
Germany/E.U.	2835	8244	3865	10832	496	7225	655	7707	864	6012	1697	7984
Japan	6121	11179	3453	9309	1173	8791	25	6211	724	6499	2145	8282
New Zealand	3227	10921	4483	10066	1141	8163	2232	8298	137	5047	2195	8388
Norway	2218	7348	3914	10194	127	6278	1243	6592	990	5615	1675	7192
Sweden	2485	9306	3050	10704	2142	8036	688	8146	1288	6079	1906	8410
Switzerland	3255	9296	1531	8960	1053	6513	259	6299	353	4684	1210	7056
U.K.	6303	10826	2889	9646	321	4828	22	6389	349	5413	1803	7280
Argentina					1674	2132	1565	3390	8213	10230	4011	5524
Columbia					3711	9102	3723	9918	1821	6589	3010	8463
India					1305	9112	3399	9262	387	5249	1716	7774
Indonesia			20	3936	988	5291	2801	9641	3084	8593	2090	7493
Israel	15551	3570	9365	5464	1882	8918	1542	7876	620	6223	3658	7021
Philippines			10529	13022	4671	9466	3644	8230	357	6962	3348	8509
Romania					11420	14408	1915	5954	896	6175	2312	6823
Russia					6787	11236	3806	9383	2597	8139	3836	9198
Slovak					345	5616	1643	8526	890	5813	1061	6818
Brazil					4185	5536	1350	10093	650	8451	1583	8580
Chile					5425	10570	3694	8572	210	6692	2635	8214
Czech					16	5919	2646	8221	1670	5788	1619	6726
Hungary					331	6843	473	6934	133	4711	310	6096
Korea					3098	9697	1899	8850	124	5346	1547	7761
Mexico			63	1388	628	5454	905	5326	147	6579	526	5518
Poland					2427	7506	1617	6950	359	5365	1298	6447
Singapore			1060	6313	2422	6740	1014	6655	1068	5400	1415	6268
South Africa			2045	9723	2817	8579	1661	8192	187	6635	1649	8199
Taiwan			8678	9708	4078	9220	2860	8490	116	6551	3278	8317
Thailand					1247	7120	2917	9792	2483	7851	2284	8330
Turkey					13393	11654	1477	8028	1935	9302	5334	9596

Table 5: The table presents the annualized Sharpe ratios before transaction costs of seven portfolios including the FDR^{+}, the $fFDR^{+}$ and $mfFDR^{+}$ controlling FDR at 20%. For the $fFDR^{+}$, we first consider four covariates: α_{bh}, β_{bh}, R^2_{bh}, ρ and the first principal component of the four mentioned covariates. For the $mfFDR^{+}$ we study $d = 4$ with the four covariates. The last row is the average of Sharpe ratio across currencies. The numbers in parentheses are the corresponding $p-$values. "*", "**" and "***" respectively indicate statistical significance at levels of $10\%, 5\%$ and 1%.

Currency	FDR^{+}	$fFDR^{+}$					$mfFDR^{+}$
		α_{bh}	β_{bh}	R^2_{bh}	ρ	PC1	
Australia	0.16 (0.15)	0.07 (0.60)	0.14 (0.30)	0.14 (0.30)	0.14 (0.30)	0.12 (0.34)	0.18 (0.17)
Canada	0.09 (0.53)	0.00 (0.94)	0.11 (0.44)	0.10 (0.52)	0.09 (0.59)	0.11 (0.46)	0.17 (0.23)
Germany/E.U.	0.24 (0.06)*	0.24 (0.11)	0.35 (0.01)***	0.34 (0.01)***	0.23 (0.13)	0.31 (0.03)**	0.49 (0.00)***
Japan	0.49 (0.00)***	0.37 (0.01)***	0.27 (0.07)*	0.31 (0.03)**	0.15 (0.29)	0.30 (0.03)**	0.45 (0.00)***
New Zealand	0.27 (0.00)***	0.22 (0.08)*	0.21 (0.07)*	0.29 (0.01)***	0.27 (0.02)**	0.34 (0.00)***	0.34 (0.00)***
Norway	0.11 (0.44)	0.17 (0.22)	0.07 (0.61)	0.25 (0.08)*	0.02 (0.87)	0.02 (0.90)	0.18 (0.17)
Sweden	0.34 (0.00)***	0.46 (0.00)***	0.32 (0.02)**	0.30 (0.02)**	0.28 (0.03)**	0.32 (0.02)**	0.40 (0.00)***
Switzerland	0.02 (0.89)	0.22 (0.13)	0.13 (0.35)	0.15 (0.27)	0.05 (0.70)	0.16 (0.28)	0.26 (0.06)*
U.K.	0.24 (0.09)*	0.09 (0.52)	0.22 (0.13)	0.19 (0.20)	0.20 (0.17)	0.12 (0.37)	0.29 (0.05)**
Argentina	0.39 (0.12)	0.42 (0.01)***	0.42 (0.01)***	0.41 (0.02)**	0.47 (0.01)***	0.41 (0.02)**	0.35 (0.06)*
Columbia	1.02 (0.00)***	0.59 (0.00)***	0.51 (0.02)**	0.64 (0.00)***	0.75 (0.00)***	0.52 (0.02)**	0.60 (0.00)***
India	0.37 (0.01)***	0.25 (0.15)	0.17 (0.32)	0.29 (0.08)*	0.28 (0.09)*	0.35 (0.04)**	0.34 (0.03)**
Indonesia	0.40 (0.04)**	0.34 (0.02)**	0.21 (0.17)	0.25 (0.08)*	0.22 (0.16)	0.30 (0.03)**	0.34 (0.02)**
Israel	1.10 (0.00)***	0.94 (0.00)***	0.78 (0.00)***	0.75 (0.00)***	0.81 (0.00)***	0.81 (0.00)***	0.54 (0.00)***
Philippines	0.51 (0.00)***	0.61 (0.00)***	0.59 (0.00)***	0.56 (0.00)***	0.60 (0.00)***	0.63 (0.00)***	0.62 (0.00)***
Romania	0.07 (0.72)	0.08 (0.65)	0.09 (0.62)	0.10 (0.58)	-0.03 (0.94)	0.03 (0.88)	0.23 (0.19)
Russia	0.44 (0.08)*	0.41 (0.06)*	0.39 (0.06)*	0.42 (0.05)**	0.42 (0.06)*	0.43 (0.05)**	0.51 (0.01)***
Slovak	0.04 (0.77)	-0.01 (0.98)	0.20 (0.31)	0.12 (0.51)	0.08 (0.67)	0.14 (0.46)	0.18 (0.35)
Brazil	0.08 (0.65)	0.24 (0.20)	0.02 (0.91)	0.10 (0.63)	0.06 (0.72)	0.11 (0.60)	0.36 (0.07)*
Chile	0.56 (0.01)***	0.65 (0.00)***	0.44 (0.02)**	0.46 (0.01)***	0.61 (0.01)***	0.56 (0.01)***	0.46 (0.02)**
Czech	0.11 (0.51)	-0.01 (0.98)	0.05 (0.72)	0.27 (0.11)	-0.08 (0.68)	0.00 (0.93)	0.12 (0.46)
Hungary	0.01 (0.89)	-0.01 (1.00)	-0.03 (0.89)	0.03 (0.85)	-0.04 (0.90)	-0.06 (0.76)	-0.11 (0.60)
Korea	0.41 (0.07)*	0.33 (0.09)*	0.23 (0.22)	0.22 (0.23)	0.18 (0.34)	0.27 (0.16)	0.29 (0.16)
Mexico	0.17 (0.20)	0.17 (0.25)	-0.05 (0.76)	0.04 (0.75)	0.06 (0.65)	0.10 (0.52)	0.10 (0.49)
Poland	-0.07 (0.70)	0.07 (0.67)	0.05 (0.76)	0.00 (0.98)	0.10 (0.60)	0.15 (0.45)	0.02 (0.92)
Singapore	0.14 (0.38)	0.14 (0.40)	0.12 (0.53)	0.21 (0.20)	0.13 (0.45)	0.03 (0.88)	0.32 (0.02)**
South Africa	0.16 (0.33)	0.21 (0.18)	0.09 (0.53)	0.12 (0.41)	0.22 (0.14)	0.28 (0.07)*	0.23 (0.14)
Taiwan	0.77 (0.00)***	0.70 (0.00)***	0.55 (0.00)***	0.69 (0.00)***	0.52 (0.00)***	0.70 (0.00)***	0.73 (0.00)***
Thailand	0.34 (0.07)*	0.51 (0.01)***	0.27 (0.13)	0.29 (0.12)	0.37 (0.04)**	0.23 (0.22)	0.47 (0.02)**
Turkey	0.60 (0.00)***	0.62 (0.00)***	0.61 (0.00)***	0.60 (0.00)***	0.64 (0.00)***	0.62 (0.00)***	0.51 (0.00)***
Average	0.32	0.30	0.25	0.29	0.26	0.28	0.33

Table 6: Net Sharpe ratios of portfolios. The table presents the annualized Sharpe ratio after transaction cost of seven portfolios including the FDR^+, the $fFDR^+$ and $mfFDR^+$ controlling FDR at 20%. For the $fFDR^+$, we first consider four covariates: α_{bh}, β_{bh}, R^2_{bh}, ρ and the first principal component of the four mentioned covariates. For the $mfFDR^+$ we study $d = 4$ with the four covariates. The second last row is the average of Sharpe ratio across currencies. The last row is $t-$statistic of the test comparing the (paired) means of the portfolios $fFDR^+/mfFDR^+$ to the portfolios FDR^+. The numbers in parentheses are the corresponding $p-$values. "*", "**" and "***" respectively indicate statistical significance at levels of $10\%, 5\%$ and 1%.

Currency	FDR^+	$fFDR^+$					$mfFDR^+$
		α_{bh}	β_{bh}	R^2_{bh}	ρ	PC1	
Australia	-0.02 (0.42)	-0.01 (0.97)	0.11 (0.42)	0.11 (0.42)	0.11 (0.42)	0.06 (0.63)	0.14 (0.29)
Canada	-0.25 (0.80)	-0.10 (0.50)	0.06 (0.69)	0.05 (0.80)	0.03 (0.89)	0.00 (0.98)	0.11 (0.46)
Germany/E.U.	0.13 (0.02)**	0.20 (0.17)	0.32 (0.02)**	0.31 (0.02)**	0.19 (0.19)	0.27 (0.07)*	0.46 (0.00)***
Japan	0.38 (0.06)*	0.31 (0.02)**	0.24 (0.11)	0.28 (0.06)*	0.11 (0.43)	0.25 (0.09)*	0.41 (0.00)***
New Zealand	0.00 (0.05)**	0.10 (0.43)	0.15 (0.24)	0.24 (0.05)**	0.19 (0.11)	0.24 (0.05)**	0.28 (0.02)**
Norway	-0.08 (0.14)	0.07 (0.64)	0.03 (0.88)	0.21 (0.14)	-0.04 (0.78)	-0.06 (0.64)	0.13 (0.36)
Sweden	0.17 (0.04)**	0.39 (0.01)***	0.28 (0.05)**	0.25 (0.04)**	0.24 (0.07)*	0.26 (0.05)**	0.36 (0.01)***
Switzerland	-0.12 (0.41)	0.16 (0.26)	0.10 (0.44)	0.11 (0.41)	0.01 (0.91)	0.10 (0.47)	0.22 (0.12)
U.K.	0.09 (0.30)	0.02 (0.87)	0.18 (0.22)	0.15 (0.30)	0.15 (0.29)	0.06 (0.64)	0.25 (0.08)*
Argentina	0.27 (0.03)**	0.36 (0.04)**	0.38 (0.02)**	0.37 (0.03)**	0.43 (0.01)***	0.35 (0.04)**	0.31 (0.11)
Columbia	0.71 (0.01)***	0.44 (0.04)**	0.40 (0.07)*	0.51 (0.01)***	0.59 (0.00)***	0.35 (0.10)*	0.53 (0.01)***
India	0.25 (0.12)	0.18 (0.29)	0.13 (0.45)	0.25 (0.12)	0.24 (0.15)	0.30 (0.07)*	0.30 (0.07)*
Indonesia	-0.09 (0.29)	0.23 (0.12)	0.12 (0.40)	0.15 (0.29)	0.11 (0.47)	0.19 (0.18)	0.24 (0.09)*
Israel	0.73 (0.00)***	0.75 (0.00)***	0.68 (0.00)***	0.63 (0.00)***	0.68 (0.00)***	0.65 (0.00)***	0.39 (0.01)***
Philippines	-0.42 (0.05)**	0.36 (0.05)**	0.40 (0.03)**	0.35 (0.05)**	0.39 (0.03)**	0.35 (0.07)*	0.46 (0.01)***
Romania	-0.31 (0.81)	-0.13 (0.53)	0.03 (0.86)	0.04 (0.81)	-0.13 (0.58)	-0.11 (0.60)	0.12 (0.51)
Russia	0.39 (0.07)*	0.38 (0.07)*	0.37 (0.07)*	0.40 (0.07)*	0.40 (0.07)*	0.40 (0.05)**	0.48 (0.01)***
Slovak	-0.12 (0.62)	-0.09 (0.68)	0.17 (0.38)	0.10 (0.62)	0.05 (0.76)	0.08 (0.63)	0.13 (0.48)
Brazil	-0.05 (0.71)	0.20 (0.29)	0.00 (0.98)	0.07 (0.71)	0.02 (0.94)	0.06 (0.78)	0.33 (0.10)*
Chile	0.22 (0.07)*	0.52 (0.01)***	0.35 (0.07)*	0.36 (0.07)*	0.47 (0.03)**	0.43 (0.05)**	0.38 (0.06)*
Czech	-0.11 (0.17)	-0.12 (0.55)	0.01 (0.89)	0.23 (0.17)	-0.14 (0.42)	-0.08 (0.70)	0.06 (0.65)
Hungary	-0.23 (0.97)	-0.09 (0.67)	-0.06 (0.75)	0.00 (0.97)	-0.08 (0.67)	-0.11 (0.55)	-0.17 (0.41)
Korea	0.15 (0.41)	0.18 (0.33)	0.15 (0.39)	0.14 (0.41)	0.09 (0.62)	0.14 (0.42)	0.23 (0.25)
Mexico	0.04 (0.97)	0.10 (0.49)	-0.09 (0.58)	0.00 (0.97)	0.02 (0.86)	0.03 (0.85)	0.04 (0.73)
Poland	-0.23 (0.89)	-0.01 (0.98)	0.03 (0.87)	-0.03 (0.89)	0.06 (0.77)	0.09 (0.64)	-0.03 (0.89)
Singapore	-0.38 (0.67)	-0.19 (0.25)	-0.02 (0.93)	0.08 (0.67)	-0.03 (0.87)	-0.20 (0.21)	0.15 (0.32)
South Africa	-0.27 (0.99)	0.01 (0.93)	-0.02 (0.93)	0.00 (0.99)	0.08 (0.62)	0.10 (0.54)	0.10 (0.49)
Taiwan	0.44 (0.00)***	0.56 (0.00)***	0.45 (0.00)***	0.59 (0.00)***	0.42 (0.01)***	0.57 (0.00)***	0.66 (0.00)***
Thailand	0.11 (0.30)	0.37 (0.05)**	0.18 (0.35)	0.19 (0.30)	0.27 (0.13)	0.07 (0.75)	0.38 (0.04)**
Turkey	0.36 (0.00)***	0.54 (0.00)***	0.53 (0.00)***	0.52 (0.00)***	0.56 (0.00)***	0.54 (0.00)***	0.45 (0.00)***
Average	0.06	0.19	0.19	0.22	0.18	0.18	0.26
t-stats		3.9	3.5	4.8	3.6	3.4	5.1

5.3. Basket of Currencies

Hitherto, we select out-performing trading rules separately by currency. The profitability of the trading rules are shown to be depending on the currency and the period of time. It, therefore, might be beneficial for a trader to assess the performance of trading rules across currencies simultaneously. In doing so, the trader will be able to diversify and/or switch her/his funds from a rule implemented on a particular currency to other profitable rules applied on other currencies. Thus, we assume that the trader has access to trade on all considering currencies and select, with control of luck, out-performing rules in a pool of all trading rules across currencies. Depending on the availability of the data, the pool consists of 190,755 ($= 9 \times 21195$) to 635,850 ($= 30 \times 21195$) trading rules. We construct a monthly rolling portfolio as in the previous section with use of the new trading rule set. The four input covariates of the $mfFDR^+$ now are ρ, α_{mk}, β_{mk} and r^2_{mk} described in Section 5.1. The quantiles of correlated coefficients of covariate pairs are shown in Table 7. Here, again a majority of more than 98% of the coefficients are in $[-0.73, 0.52]$.

Table 7: Summary of correlation coefficients of covariate pairs: the case of all currencies. The table presents the quantiles of the correlated coefficient of six covariate pairs which are combinations of the four covariates: $r^2_{mk}, \alpha_{mk}, \beta_{mk}$ and ρ. The final row shows the numbers for the set of all correlated pairs.

Covariate pairs	min	1%	5%	25%	50%	75%	95%	99%	max
r^2_{mk}, α_{mk}	-0.44	-0.38	-0.27	-0.08	0.02	0.10	0.37	0.67	0.71
r^2_{mk}, β_{mk}	-0.81	-0.76	-0.70	-0.52	-0.23	0.12	0.50	0.54	0.55
r^2_{mk}, ρ	-0.26	-0.24	-0.18	-0.08	-0.02	0.07	0.25	0.30	0.36
α_{mk}, β_{mk}	-0.93	-0.90	-0.68	-0.31	-0.16	-0.03	0.31	0.53	0.63
α_{mk}, ρ	-0.20	-0.15	-0.11	-0.03	0.04	0.12	0.26	0.33	0.36
β_{mk}, ρ	-0.40	-0.36	-0.27	-0.11	-0.03	0.05	0.15	0.20	0.21
All pairs	-0.93	-0.73	-0.55	-0.16	-0.03	0.07	0.32	0.52	0.71

At beginning of each month, the trading rules selected by the $mfFDR^+$ procedure (control for FDR at 20%) are combined by currency. For instance, suppose there are k rules are selected and that all of them are ones applied on two currencies namely A and B with sizes (i.e. numbers of rules selected) of k_A and $k_B = k - k_A$, respectively. The wealth is split into k portions and k_A (k_B) of them are invested on the k_A (k_B) rules traded on currency A (B). The selected rules on each currency are combined as in Section 5.2.

The OOS performance of the $mfFDR^+$ based portfolios in terms of annualized Sharpe ratios and mean returns before and after transaction cost are exhibited in Table 8. The first row of the table reveals the performances over the whole sample period, from 1973 to the end of 2020. The Sharpe ratios show an impressive $mfFDR^+$ performance with values of about 1.06 and 0.95 before and after transaction cost, respectively. The break-even point, which is the fixed transaction cost at which the Sharpe ratio of portfolio is set to zero, is very high at 60 basis points. Those numbers are evidently show that in whole sample period, the $mfFDR$ portfolio that is consisted by all currencies portfolio performs much better than any considered $FDR^+, fFDR^+$ and $mfFDR^+$ based portfolios that are traded on a single currency. The next rows of table break down the performance of the portfolio into sub-periods of roughly ten years. Though the numbers show a deterioration in Sharpe ratio overtime, the $mfFDR^+$ portfolio is still profitable even in the most recent decade.

Table 8: Performance of $mfFDR^+$ portfolio. The table shows the annualized Sharpe ratios (SR) and mean returns (before and after transaction cost) of implementing the $mfFDR^+$ on all strategies in all currencies to control the FDR at 20%. The last column is the related break-even point.

Period	Excess SR	Net SR	Excess Return (%)	Net Return (%)	Break-even (bps)
Whole Period	1.06	0.95	3.80	3.40	60
1973-1980	1.45	1.35	4.47	4.18	69
1981-1990	2.08	1.97	7.30	6.93	128
1991-2000	0.92	0.81	4.15	3.64	72
2001-2010	0.69	0.53	2.37	1.82	34
2011-2020	0.29	0.21	0.88	0.63	14

5.3.1. Contribution of Categories

In this section, we investigate further the contribution of each category into the $mfFDR$ portfolio. First, we uncover the profit of the selected strategies by each category. For this purpose, we split the selected rules of the $mfFDR$ portfolio into five groups based on trading rule category and track the OOS performance of each group. Table 9 shows the OOS performance metrics of those groups in whole period.

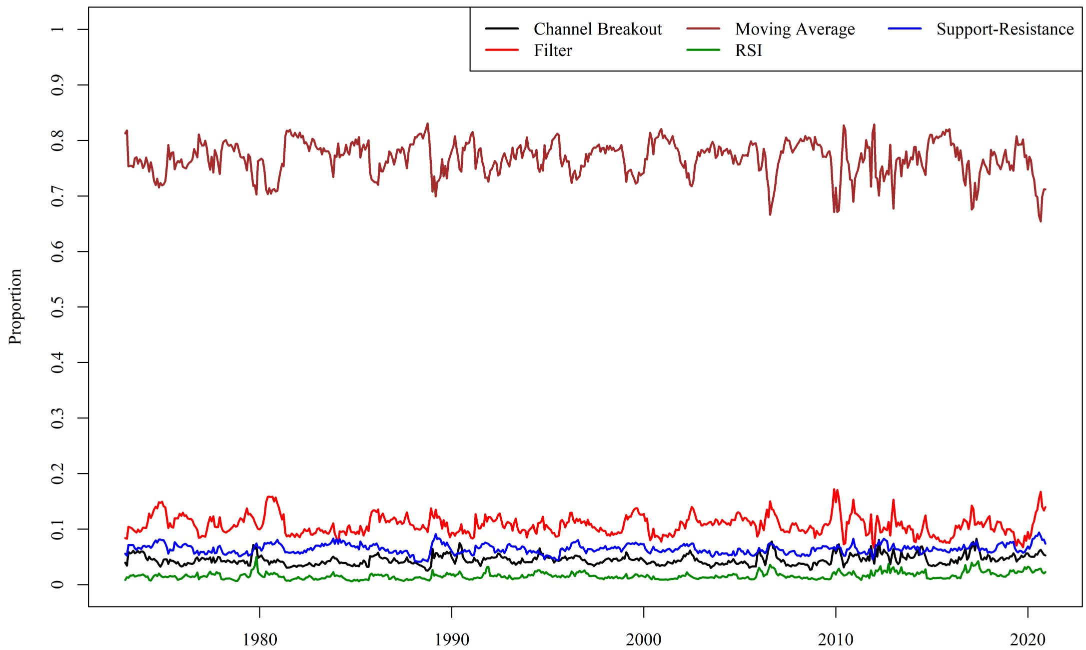
Table 9: Profit contributions of each group. The table shows the annualized Sharpe ratios (SR) and mean returns (before and after transaction cost) of the five groups which are split from the $mfFDR$ portfolio based on the trading rule category.

Category	Excess SR	Net SR	Excess Return (%)	Net Return (%)
Channel Breakout	0.84	0.77	3.08	2.82
Filter Rule	1.25	0.91	4.72	3.45
Moving Average	1.01	0.92	3.80	3.47
RSI	-0.48	-0.98	-0.55	-1.13
Support-Resistance	0.98	0.78	3.29	2.61

In terms of profit magnitude, the filter rule groups appear to be the best before transaction cost, followed by moving average, support-resistance and channel breakout in that order. The RSI negatively contribute to the whole portfolio with negative profit, however. By comparing the two tables 8 and 9 we see that, although the filter rule group gain higher Sharpe ratio before cost than that of the combination of all groups, it has a lower Sharpe ratio after transaction cost than that of the whole portfolio. This confirms the fact that combination of all rules reduces the transaction cost, thus the combined portfolio might gain higher Sharpe ratio at the end.

Yet, the profit of a category might contribute little to the whole portfolio if the portion of that category in the portfolio is small. We thus discover further the size aspect, that is, the proportion of each category in the portfolio of rules. Figure 7 depicts the proportions overtime, from 1973 to 2020. We see that about 80% of the selected rules belong to the moving average category, followed by filter rules with roughly more than 10%. The RSI contributes least to the portfolio. The performance of the portfolio, therefore, is influenced most by those selected moving average rules and though the selected RSI rules are perform worst, this group contributes very little on the performance of the whole portfolio.

Figure 7: Contribution of categories in terms of size overtime. The figure shows the proportion of each category in the selected profitable rules. More specifically, each of the selected strategy in the $mfFDR$ portfolio belongs to one of the five categories. To have the figure, each month we calculate the proportion of each category in the portfolio and depict it by a point.



The analyses above tell us the structure of the portfolio constructed by implementing $mfFDR$ on all rules in all currencies. The results somewhat reflect the answer for an important question: which is the most profitable category? However, they do not lead to a comprehensive conclusion since the estimation of FDR is based on all categories as a pool.

5.3.2. Portfolios Conditional on Category

As documented, performance of rules are different across category. Put differently, the category is itself an informative covariate. It is therefore interesting to see how $mfFDR$ performs conditionally on the category. However, this covariate is not continuous and thus cannot be an input of $mfFDR$. To study its informativeness, we therefore repeat the experiment with basket of currencies but on each of the category. In doing so, we are able to answer for two questions: first, which category is most profitable; and second, which categories traders should consider or avoid when they construct a portfolio. More specifically, we pool all trading rules of a particular category currencies to form a new set of rules. Thus, we have five new sets: the largest set is that of the Moving Average category with number of rules varying from 9×2,870 =25,830 to 30×2,870 = 86,100 rules while the smallest set is that of RSI category containing from 9 × 600 = 54,000 to 30 × 600 =18,000 rules. We construct five $mfFDR$-based portfolios which control FDR at 20%, each for one of the new sets.

We first investigate how prevalent the profitable rules is. Thereby, for each portfolio corresponding to each category, we calculate its profitable rate which is merely the ratio of the number of trading rules selected as profitable by $mfFDR^+$ divided by the overall number of input rules. This measure shows how rich each technical rule category is in terms of containing outperforming rules (in in-sample period). At the end of each month when the rules are selected, we have a profitable portion for each category and this portion

Table 10: Ratios (in %) of selected technical rules in each category selected by $mfFDR^+$ under controlling for FDR at 20%. The table exhibits the average of the ratios of trading rules (in %) in each category (Channel breakout (CB), Filter rule (FR), Moving average (MA), Relative strength relative (RSI) and Support-resistance (SR)) have been selected to invest each month over whole period (first row) and sub-periods (remaining rows).

	Selected strategy rate				
Period	CB	FR	MA	RSI	SR
Whole Period	7	17	24	11	15
1973-1980	10	24	31	13	20
1981-1990	9	21	34	10	20
1991-2000	7	15	22	10	14
2001-2010	6	16	20	9	13
2011-2020	5	12	16	11	11

is obviously varying overtime. In Table 10 we present the average of the portions over the whole sample period and sub-periods. We observe that, in whole sample period, the moving average category contains highest proportion of profitable strategies with 24% of the set to be outperforming, followed by filter rule, support-resistance, RSI and channel

Table 11: Performance of $mfFDR^{+}$ portfolios implemented on each category. The table shows the annualized Sharpe ratios (SR) and mean returns (before and after transaction cost) of portfolios generated by implementing the $mfFDR^{+}$ on each category of trading rule (on all currencies) to control the FDR at 20%. The last column is the related break-even point. In each panel, we present the mentioned metrics for the portfolio implemented on a category in whole sample period (first row) and sub-samples (next five rows).

Period	Excess SR	Net SR	Excess Return (%)	Net Return (%)	Break-even (bps)
Panel A: Channel Breakout Rule					
Whole Period	0.88	0.80	2.64	2.40	72
1973-1980	0.82	0.76	2.05	1.88	62
1981-1990	2.03	1.95	5.81	5.57	184
1991-2000	0.98	0.89	3.45	3.14	99
2001-2010	0.47	0.36	1.43	1.11	37
2011-2020	0.12	0.06	0.33	0.18	7
Panel B: Filter Rule					
Whole Period	1.26	0.91	4.41	3.19	23
1973-1980	1.77	1.53	5.62	4.86	36
1981-1990	2.17	1.81	7.14	5.97	41
1991-2000	0.98	0.66	4.90	3.27	27
2001-2010	1.02	0.44	2.95	1.28	14
2011-2020	0.67	0.37	1.71	0.95	7
Panel C: Moving Average					
Whole Period	1.01	0.92	3.80	3.47	71
1973-1980	1.39	1.32	4.49	4.24	89
1981-1990	1.99	1.91	7.42	7.12	156
1991-2000	0.92	0.83	4.23	3.82	89
2001-2010	0.65	0.53	2.43	1.97	41
2011-2020	0.18	0.11	0.57	0.37	9
Panel D: RSI					
Whole Period	-0.46	-0.96	-0.50	-1.04	-
1973-1980	-0.73	-1.16	-0.63	-0.99	-
1981-1990	-0.20	-0.70	-0.25	-0.84	-
1991-2000	-1.04	-1.67	-1.05	-1.70	-
2001-2010	-0.15	-0.73	-0.19	-0.90	-
2011-2020	-0.40	-0.74	-0.40	-0.74	-
Panel E: Support-resistance					
Whole Period	1.00	0.80	3.13	2.48	32
1973-1980	1.37	1.20	3.64	3.18	41
1981-1990	2.06	1.83	6.02	5.34	64
1991-2000	0.87	0.68	3.68	2.87	41
2001-2010	0.55	0.24	1.54	0.67	15
2011-2020	0.34	0.19	0.87	0.49	8

breakout categories with 17%, 15%, 11% and only 7%, respectively. When breaking down the rates into sub-periods, more useful facts are revealed. First, although the portions varied, the ranking in profitability rate of categories is consistent over time. Second, the profitable portion tend to be deteriorated overtime regardless categories. Nevertheless, although the percentage of genuine profitable rules has been generally decreased, this pattern seems to be stopped and stable in last three decades with a considerable portion of them identified as significant profitable.

The prevalence of in-sample profitable rules does not always imply a high profit gain in OOS, however. In Table 11, we present the OOS performance of those selected rules. We learn several facts from the results. First, generally the portfolio conditional on filter rules is the most profitable OOS before transaction cost and remains so alongside with moving average after transaction cost. Second, the fact that the break-even points corresponding with filter and support-resistance categories are lower than those of channel breakout and moving average ones implies that the former ones are trading in higher frequencies, and thus generate more trading cost. The implication is that, in the time when transaction cost is high, the traders should avoid using or containing the filter and support-resistance rules in the portfolio. Third, the RSI category performs worst with negative profit for all sub-samples and thus for whole sample period. This implies that the performance of the selected out-performing RSI rules is not persistent. Thus, if traders construct a $mfFDR$ portfolio based on all rules, they should exclude the RSI rules from the pool.

The profitability of four out of five categories and the facts that those rules are prevalent and can be selected ex-ante allow us to argue that technical analysis still has value on the FX market.

6. Conclusion

We introduce the $mfFDR$, which estimates the FDR as a function of multiple covariates, for detecting the false null hypotheses with control of FDR. We show that the method works well in controlling FDR and gain a considerable higher power than existing methods in detecting false null hypotheses both under independent and weak dependent settings. The performance is also robust when the covariates are biased estimated, and there is a certain level of dependency among them.

Empirically, we apply the $mfFDR$ on a large universe of technical trading rules to detect truly profitable ones with control of data snooping. We first study the trading rules on each currency individually. With the use of multiple informative covariates, the results show that the $mfFDR$ based portfolio is much more powerful than the existing methods that are not using covariates. After combining, the portfolio of selected rules can generate positive profit, which is higher than the data snooping control methods with and without using a sole covariate. We then study 30 currencies together where more than 600 thousand trading rules are generated. We implement the method on this set of rules to construct a portfolio that generates a Sharpe ratio of roughly one for roughly 50 years. Given that the profitable portfolio of trading rules is constructed based on only past exchange spot rate data, the results show the value of technical rules in the foreign exchange market. Further analyses additionally provide insight information on which categories of trading rule traders should take into consideration when constructing a portfolio of trading rules.

The development of the $mfFDR$ framework will contribute to complex problems in Finance, Economics and other fields of Science that are plagued by multiple competing models and hypotheses. It is a powerful framework, robust to noise, that can assist in decision making in the era of big data.

References

Allen, H. and Taylor, M. P. (1990) Charts, noise and fundamentals in the london foreign exchange market. *The Economic Journal*, **100**, 49–59.

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

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

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300.

Burnside, C., Eichenbaum, M. S. and Rebelo, S. (2011) Carry trade and momentum in currency markets. Tech. rep., National Bureau of Economic Research.

Chen, X., Robinson, D. G. and Storey, J. D. (2021) The functional false discovery rate with applications to genomics. *Biostatistics*, **22**, 68–81.

Chinn, M. D. and Meese, R. A. (1995) Banking on currency forecasts: How predictable is change in money? *Journal of International Economics*, **38**, 161–178.

Cialenco, I. and Protopapadakis, A. (2011) Do technical trading profits remain in the foreign exchange market? evidence from 14 currencies. *Journal of International Financial Markets, Institutions and Money*, **21**, 176–206.

DeMiguel, V., Garlappi, L. and Uppal, R. (2007) Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, **22**, 1915–1953.

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

Geenens, G. (2014) Probit transformation for nonparametric kernel estimation on the unit interval. *Journal of the American Statistical Association*, **109**, 346–358.

Goodman, S. H. (1979) Foreign exchange rate forecasting techniques: Implications for business and policy. *The Journal of Finance*, **34**, 415–427.

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

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**, 471–484.

Hsu, P.-H., Kyriakou, I., Ma, T. and Sermpinis, G. (2021) Informative covariates, false discoveries and mutual fund performance. *Working paper.*

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.

Loader, C. (1999) *Local Regression and Likelihood.* Springer.

Lustig, H., Roussanov, N. and Verdelhan, A. (2011) Common Risk Factors in Currency Markets. *The Review of Financial Studies*, **24**, 3731–3777.

Meese, R. A. and Rogoff, K. (1983) Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, **14**, 3–24.

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

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

Neely, C. J., Weller, P. A. and Ulrich, J. M. (2009) The adaptive markets hypothesis: Evidence from the foreign exchange market. *The Journal of Financial and Quantitative Analysis*, **44**, 467–488.

Newton, M. A., Noueiry, A., Sarkar, D. and Ahlquist, P. (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, **5**, 155–176.

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

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

Storey, J. D., Akey, J. M. and Kruglyak, L. (2005) Multiple locus linkage analysis of genomewide expression in yeast. *PLOS Biology*, **3**, 1380–1390.

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

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

Appendix A. The multivariate functional false discovery rate

To complement Section 2.1 of the manuscript, we present the procedures for estimating the $\pi_0(\boldsymbol{z})$, $f(p,\boldsymbol{z})$ and $q-$value.

To estimate $\pi_0(\boldsymbol{z})$ one can extend the "bin" approach presented in HKMS, where we partition m tests into some groups based on covariates and estimate $\pi_0(\boldsymbol{z})$ as a constant in each group. In this study, we additionally utilize another approach which is based on density estimation as following. Firstly, for some $\lambda \in [0,1)$ we define:

$$\pi_0(\boldsymbol{z},\lambda) = \frac{\mathbb{P}(P > \lambda | \boldsymbol{Z} = \boldsymbol{z})}{1-\lambda}. \tag{A.1}$$

Conditional on $\boldsymbol{Z} = \boldsymbol{z}$, $\mathbb{P}(P > \lambda) \geq \mathbb{P}(P > \lambda | h = 0).\mathbb{P}(h = 0)$. Thus, we have

$$\begin{aligned}
\mathbb{P}(P > \lambda | \boldsymbol{Z} = \boldsymbol{z}) &\geq \mathbb{P}(P > \lambda | \boldsymbol{Z} = \boldsymbol{z}, h = 0).\mathbb{P}(h = 0 | \boldsymbol{Z} = \boldsymbol{z}) \\
&= (1-\lambda)\mathbb{P}(h = 0 | \boldsymbol{Z} = \boldsymbol{z}) \\
&= (1-\lambda)\pi_0(\boldsymbol{z})
\end{aligned} \tag{A.2}$$

where the second step comes from the uniform distribution of $P|(\boldsymbol{Z} = \boldsymbol{z}, h = 0)$. This turns out $\pi_0(\boldsymbol{z},\lambda) \geq \pi_0(\boldsymbol{z})$, i.e. $\pi_0(\boldsymbol{z},\lambda)$ is an conservative estimate of $\pi_0(\boldsymbol{z})$. If we express $\pi_0(\boldsymbol{z},\lambda)$ as

$$\pi_0(\boldsymbol{z},\lambda) = \mathbb{P}(\boldsymbol{Z} = \boldsymbol{z} | P > \lambda).\frac{\mathbb{P}(P > \lambda)}{1-\lambda} \tag{A.3}$$

then, the first term $\frac{\mathbb{P}(P>\lambda)}{1-\lambda}$ is conservatively estimated by $\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{n(1-\lambda)}$, here $\#$ returns the number of elements in the set, as in Storey (2002). The remain $\mathbb{P}(\boldsymbol{Z} = \boldsymbol{z} | P > \lambda)$, which is the density function of $\boldsymbol{Z}$ conditional on $p-$value $> \lambda$, will be estimated as a function $\hat{h}_\lambda(\boldsymbol{z})$ such that

$$\hat{\pi}_0(\boldsymbol{z},\lambda) = \hat{h}_\lambda(\boldsymbol{z}).\hat{\pi}_0(\lambda) \tag{A.4}$$

is a conservative estimate of $\pi_0(\boldsymbol{z})$.

Next, to estimate the density functions $\hat{h}_\lambda(\boldsymbol{z})$ and $f(p,\boldsymbol{z})$ we use a local likelihood kernel density estimation (KDE) approach in which a probit transformation in Geenens (2014) is adopted. More specifically, let $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$ and Φ^{-1} its inverse. We transform the variables $(p,\boldsymbol{z})$ to $(\tilde{p},\tilde{\boldsymbol{z}})$ by using $\tilde{p}_i = \Phi^{-1}(p_i)$ and $\tilde{z}_i^k = \Phi^{-1}(z_i^k)$, $k = 1,\ldots,d$. We denote by $\tilde{f}(\tilde{p},\tilde{\boldsymbol{z}})$ and $\tilde{h}_\lambda(\tilde{\boldsymbol{z}})$, respectively, the joint density function of $(\tilde{p},\tilde{\boldsymbol{z}})$ and the conditional density function of $\tilde{z}$ on the group of null hypotheses having $p-$value $< \lambda$, $\{i | P_i > \lambda\}$. We estimate them by using the local likelihood KDE method in Loader (1999). When the number of variables in the estimating function is greater than two (i.e., $d \geq 3$ for $\hat{h}_\lambda(\tilde{\boldsymbol{z}})$ and $d \geq 2$ for $f(p, \tilde{\boldsymbol{z}})$), we drop the cross-product terms from the local model. This allows us to overcome the curse of dimensionality as we are working with one-dimensional integrals instead of a multivariate one. The estimation can be implemented easily via the freely available R package `locfit`. When the dimension is less than three, the bandwidth of the KDE is chosen locally via a k-Nearest-Neighbor approach using generalized cross-validation similar to CRS.

The desired density functions $\hat{h}_\lambda(\boldsymbol{z})$ and $f(p, \boldsymbol{z})$, respectively, are then estimated, as $\hat{h}_\lambda(\boldsymbol{z}) = \frac{\tilde{h}_\lambda(\tilde{\boldsymbol{z}})}{\Pi_{k=1}^d \phi(\tilde{z}^k)}$ and $\hat{f}(p, \boldsymbol{z}) = \frac{\tilde{f}(\tilde{p}, \tilde{\boldsymbol{z}})}{\phi(\tilde{p}) \Pi_{k=1}^d \phi(\tilde{z}^k)}$ where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

For each λ, the $\hat{h}_\lambda(\boldsymbol{z})$ is plugged to (A.4) to acquire the estimate $\hat{\pi}_0(\boldsymbol{z}, \lambda)$. The λ is chosen in the set $\{0.4, \dots, 0.8\}$ such that the mean integrated squared error defined as in CRS is minimal.

As argued in CRS, if the $f(p, z)$ is non-increasing with respect to p for each fixed $\boldsymbol{z}$, then we should adjust the $\hat{f}(p, z)$ so that it carries this property. In practice, this can be acquired by resetting the value of $\hat{f}(p_i, z_i)$

$$\hat{f}(p_i, z_i) := \min\left\{\hat{f}(p_i, z_i), \min\left\{\hat{f}(p_l, z_l) | p_l < p_i, \|z_i - z_l\| \leq \epsilon\right\}\right\}$$

where $\|.\|$ is the Euclid distance and ϵ is a small positive number which is set at 0.01 in this study.

Finally, we calculate $\hat{r}(p, \boldsymbol{z}) = \hat{\pi}_0(\boldsymbol{z}) / \hat{f}(p, \boldsymbol{z})$ and estimate the so-called $q-$value function (see CRS) as

$$\hat{q}(p_i, \boldsymbol{z}_i) = \frac{1}{\# S_i} \sum_{k \in S_i} \hat{r}(p_k, \boldsymbol{z}_k), \tag{A.5}$$

where $S_i = \{k | \hat{r}(p_k, \boldsymbol{z}_k) \leq \hat{r}(p_i, \boldsymbol{z}_i)\}$, p_i is the $p-$value of test i and $\boldsymbol{z}_i = (z_i^1, \dots, z_i^d)$ is the covariate bundle of strategy i.[11]

We recall here the "positive false discovery rate", a type I error introduced in Storey (2003), $pFDR = \mathbb{E}\left(\frac{V}{R} \middle| R > 0\right)$ where R is the number of rejected null hypotheses in n tests and V the wrongly rejected ones. For a given target $\tau \in [0, 1]$ of $pFDR$, a null hypothesis $H_{0,i}$ is rejected if and only if $\hat{q}(p_i, \boldsymbol{z}_i) \leq \tau$. One can replicate the arguments in CRS to show that the $mfFDR$ procedure controls at the target τ of $pFDR$. We emphasize that control for $pFDR$ is equivalent to control for the FDR at the same level when n is large which is the case in this study.

[11] This estimation is proposed by Newton *et al.* (2004) and Storey *et al.* (2005) and subsequently adopted in CRS.

Appendix B. Performance of $mfFDR$ under noisy covariates

As mentioned in Section 2.3, the covariates are estimated quantities and thus, have inherited some noise that might affect the power of our method. In this section, we address this concern via considering a simple setting where the input covariates contains noise. More specifically, we use the covariates $u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n)$ as in our previous simulations in Section 2.2 to generate the simulated data, but now we use inputs of $mfFDR$ the observed $u' = (u'_1, \ldots, u'_n)$ and $v' = (v'_1, \ldots, v'_n)$ defined as

$$u'_i = u_i + \eta_i \tag{B.1}$$

and

$$v'_i = v_i + \epsilon_i \tag{B.2}$$

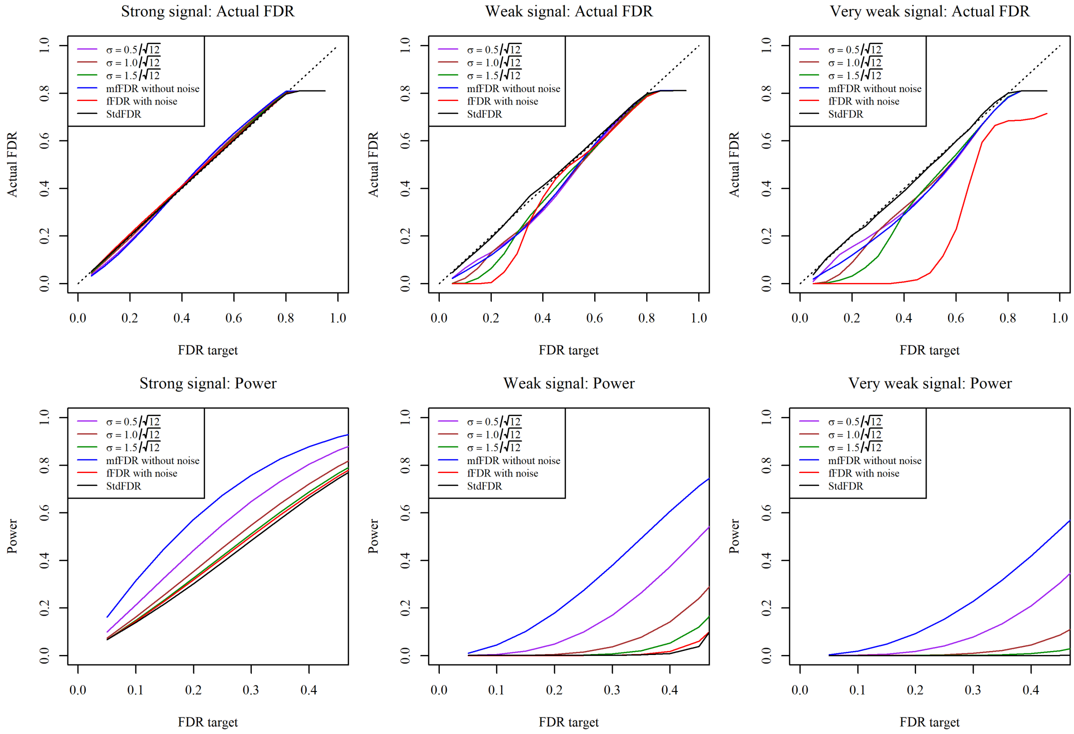
where η_i and ϵ_i are noise generated independently from a normal distribution $N(0, \sigma^2)$, $i = 1, \ldots, n$. To study the performance of the $mfFDR$, we consider three different values of σ including $\sigma_1 = 0.5/\sqrt{12}$, $\sigma_2 = 1.0/\sqrt{12}$ and $\sigma_3 = 1.5/\sqrt{12}$. The determination of those values is based on the fact that the covariates $u, v \sim U[0, 1]$, which has a standard deviation of $1/\sqrt{12}$.

The magnitude of the noise in this setting reflects the level of informativeness of the covariates (the higher variation of the noise is, the less informative the covariates are) or the strength of the relationship between the observed covariates and the true status of hypotheses (the higher variation of the noise is, the weaker the relationship is).

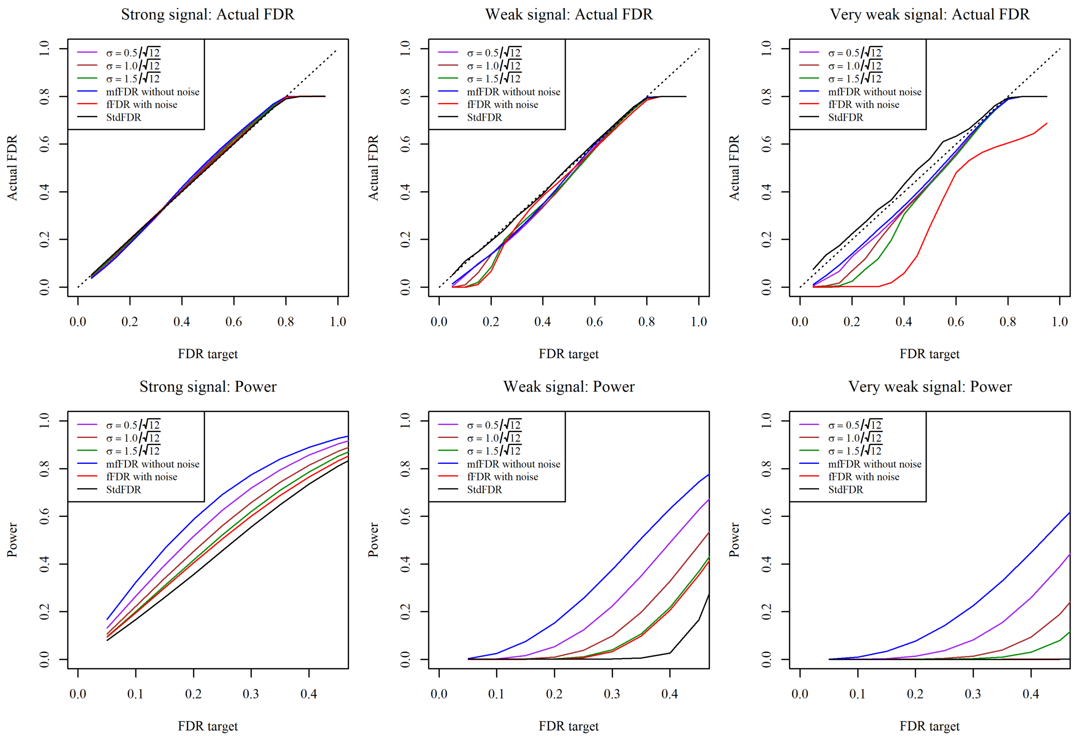
In Figure B.8, we present the performance of $mfFDR$ in terms of FDR control and its power (at FDR target up to 0.45). We see that the FDR is controlled well at any given target from 0.05 to 0.95. To have a complete picture on the impact of the noise on the power of the covariate augmented methods, we add the performance of the $fFDR$ when its covariate is u' and compare to the $mfFDR$ using original (u, v) and the $StdFDR$. Evidently, the power of the $mfFDR$ with noise in covariates is lower than the case with original ones but still remarkably higher than that of the $fFDR$ (with noised covariate u') and $StdFDR$.

Figure B.8: Performance comparison of the $mfFDR$ and $fFDR$ with noisy covariates and the Standard FDR of Storey ($SdtFDR$) procedures. Here the the input covariates for the $mfFDR$ are $u' = u + \varepsilon, v' = v + \eta$, where $\varepsilon, \eta \sim N(0, \sigma^2)$ and $\sigma \in \{0.5/\sqrt{12}, 1.0/\sqrt{12}, 1.5/\sqrt{12}\}$, whereas the $fFDR$ the u'. Panel A shows the performance when the $\pi_0(u, v)$ has a sine form whereas the Panel B the monotonic one.

Panel A: $\pi_0(u, v)$ is a sine function.



Panel B: $\pi_0(u, v)$ is monotonic with respect to each covariate.



Internet Appendix

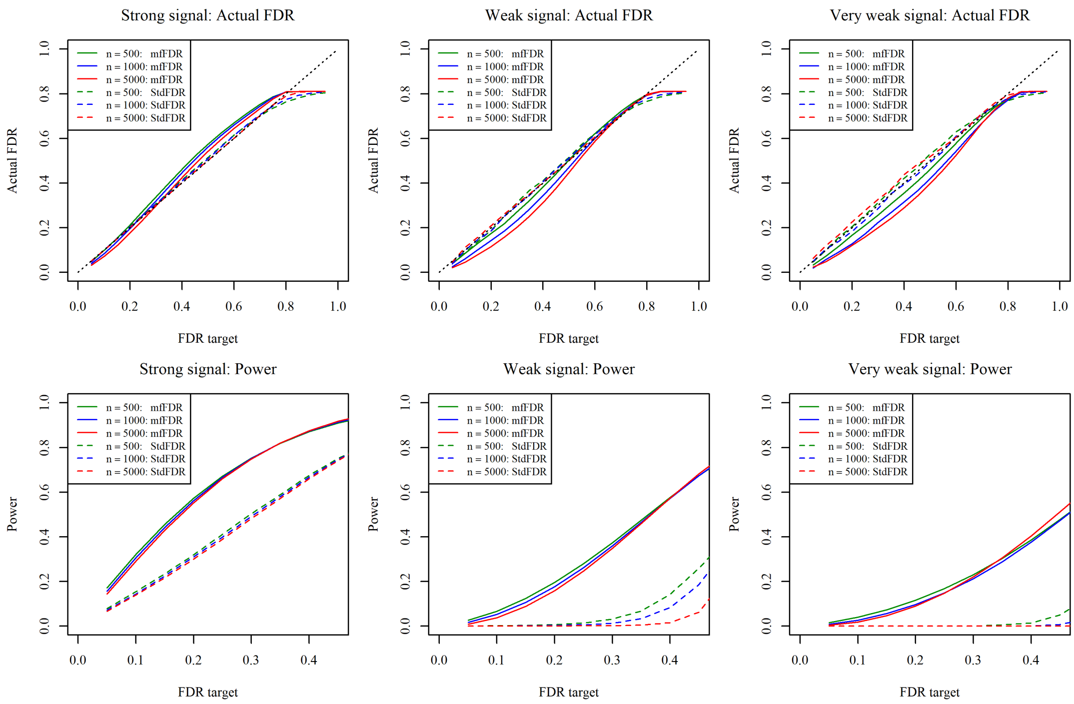
IA. Performance of $mfFDR$ under varying number of tests

In the main manuscript, we conduct simulation studies with population of n =10,000 tests. This number is chosen since it is close to that of the actual input of $mfFDR$ in the empirical experiments. Aiming for a wider range in application of the $mfFDR$ framework, we additionally conduct a robustness check with use of smaller numbers of tests. Particularly, we repeat the simulation with: $n = 500, 1000$ and 5000 tests.

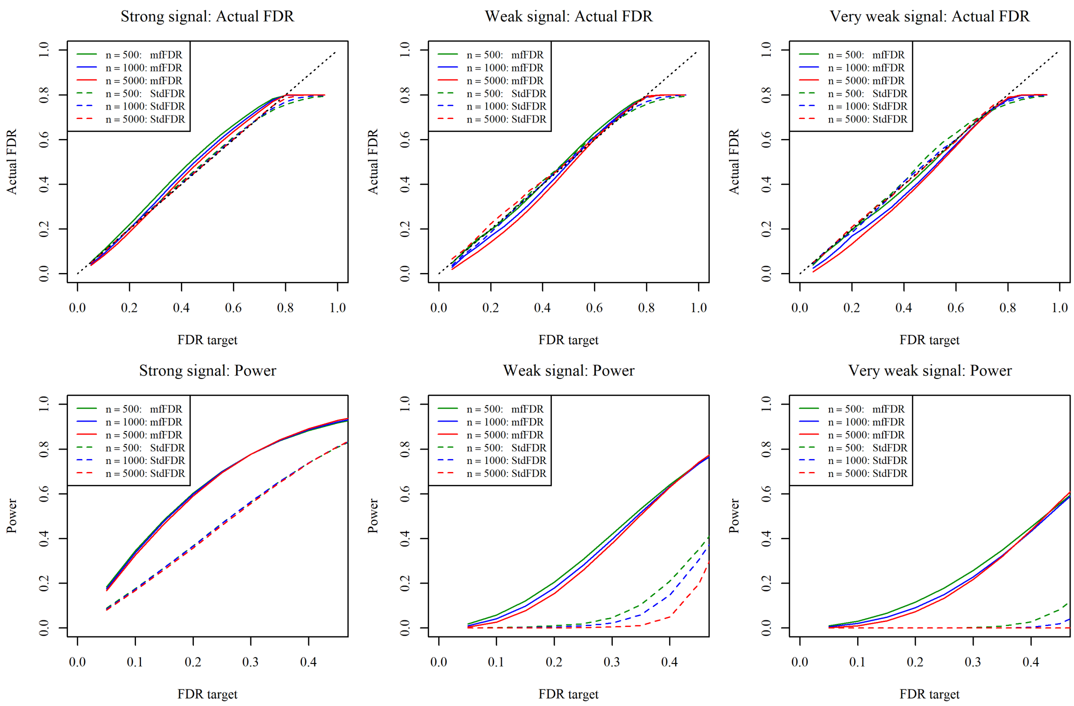
In Table I, we present the performance of the $mfFDR$ against its benchmark, the $StdFDR$, in terms of FDR control and power. First, under the weak and very weak signal cases, the $mfFDR$ controls well for any given FDR targets. Second, for strong signal data, the $mfFDR$ slightly violates the FDR control at high targets, especially when the number of tests is small. More specifically, when $n = 500$, the $mfFDR$ strictly controls well for FDR at targets up to 0.2. When $n = 3000$ and $n = 5000$, these numbers are 0.3 and 0.4. Thus, if the number of considering tests is smaller than 500 and the aim is controlling for a FDR target higher than 0.2, the method should be used on weak or very weak signal rather than strong signal data. Although, it should be noted that controlling FDR at a high target with a small number of tests, has no practical value in finance, economics and most fields of science. Last, in terms of power, the $mfFDR$ is superior to that of the $StdFDR$ in all cases regardless the number of tests.

Figure I: Performance comparison of the $mfFDR$ against the Standard FDR of Storey ($SdtFDR$) with varying population size (n). Panel A shows the performance when the $\pi_0(u,v)$ has a sine form whereas in Panel B is the monotonic one.

Panel A: $\pi_0(u,v)$ is a sine function.



Panel B: $\pi_0(u,v)$ is monotonic with respect to each covariate.



IB. Performance of $\boldsymbol{mfFDR^+}$ based portfolios with various FDR targets

In this section, we present the performance of the $mfFDR^+$ based portfolios both when it is applied on individual currency as well as on all currencies on various FDR targets. More specifically, we repeat the experiment in the main manuscript with FDR target τ varying from 10% to 40%. The results for the former portfolios are presented in Table I while those of all currencies together one are exhibited in Table II. The results corresponding to $\tau = 0.2$ is represented for conveniences in comparison. Overall, the performance of the portfolios are stable across the considered targets. When all currencies are examined together, we observe higher Sharpe ratio and smaller annual return for higher target. This observation implicitly indicates that the volatility of the portfolio's daily return is higher when we control FDR at small target.

Table I: Performance of $mfFDR$ based portfolios on individual currency with varying FDR target. The table shows annualized Sharpe ratios of the $mfFDR$ based portfolio with FDR target $\tau = \{0.1, 0.2, 0.3, 0.4\}$ based on portfolios' returns before (left side) and after transaction cost (right side). The final row shows the average Sharpe ratio across 30 portfolios corresponding to the 30 currencies. The numbers in parentheses are the corresponding $p-$values. "*", "**" and "***" respectively indicate statistical significance at levels of $10\%, 5\%$ and 1%.

Countries	Before transaction cost				After transaction cost			
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$
Australia	0.20 (0.14)	0.21 (0.10)*	0.20 (0.10)*	0.20 (0.10)*	0.16 (0.21)	0.17 (0.19)	0.16 (0.21)	0.16 (0.20)
Canada	0.19 (0.16)	0.17 (0.20)	0.16 (0.22)	0.17 (0.22)	0.13 (0.33)	0.11 (0.43)	0.10 (0.44)	0.10 (0.44)
Germany/E.U.	0.46 (0.00)***	0.47 (0.00)***	0.46 (0.00)***	0.46 (0.00)***	0.43 (0.01)***	0.44 (0.01)***	0.43 (0.01)***	0.43 (0.01)***
Japan	0.47 (0.00)***	0.46 (0.00)***	0.46 (0.00)***	0.45 (0.00)***	0.44 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.42 (0.00)***
New Zealand	0.34 (0.01)***	0.36 (0.00)***	0.36 (0.00)***	0.36 (0.00)***	0.28 (0.01)***	0.30 (0.01)***	0.31 (0.01)***	0.31 (0.01)***
Norway	0.21 (0.12)	0.20 (0.16)	0.19 (0.18)	0.19 (0.17)	0.16 (0.29)	0.15 (0.31)	0.14 (0.32)	0.14 (0.32)
Sweden	0.39 (0.00)***	0.41 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.35 (0.01)***	0.37 (0.01)***	0.37 (0.01)***	0.37 (0.01)***
Switzerland	0.29 (0.04)**	0.29 (0.04)**	0.29 (0.05)**	0.28 (0.05)**	0.25 (0.06)*	0.25 (0.06)*	0.25 (0.07)*	0.25 (0.07)*
U.K.	0.29 (0.06)*	0.29 (0.07)*	0.28 (0.07)*	0.28 (0.07)*	0.26 (0.09)*	0.25 (0.10)*	0.24 (0.11)	0.24 (0.11)
Argentina	0.49 (0.00)***	0.49 (0.00)***	0.50 (0.00)***	0.50 (0.00)***	0.44 (0.00)***	0.44 (0.00)***	0.45 (0.00)***	0.45 (0.00)***
Columbia	0.55 (0.00)***	0.58 (0.00)***	0.61 (0.00)***	0.61 (0.00)***	0.48 (0.01)***	0.51 (0.00)***	0.54 (0.00)***	0.54 (0.00)***
India	0.36 (0.04)**	0.33 (0.06)*	0.35 (0.05)**	0.35 (0.05)**	0.32 (0.07)*	0.29 (0.10)*	0.31 (0.08)*	0.31 (0.08)*
Indonesia	0.33 (0.02)**	0.34 (0.02)**	0.34 (0.02)**	0.33 (0.02)**	0.24 (0.09)*	0.24 (0.09)*	0.25 (0.09)*	0.25 (0.09)*
Israel	0.98 (0.00)***	1.01 (0.00)***	1.01 (0.00)***	1.01 (0.00)***	0.86 (0.00)***	0.89 (0.00)***	0.89 (0.00)***	0.89 (0.00)***
Philippines	0.44 (0.02)**	0.62 (0.00)***	0.43 (0.02)**	0.44 (0.02)**	0.30 (0.10)*	0.46 (0.01)***	0.30 (0.11)	0.30 (0.11)
Romania	0.22 (0.24)	0.23 (0.22)	0.25 (0.18)	0.25 (0.19)	0.12 (0.51)	0.12 (0.50)	0.13 (0.46)	0.14 (0.45)
Russia	0.52 (0.01)***	0.54 (0.00)***	0.43 (0.05)**	0.43 (0.05)**	0.49 (0.01)***	0.51 (0.01)***	0.41 (0.06)*	0.42 (0.06)*
Slovak	0.15 (0.39)	0.18 (0.31)	0.18 (0.31)	0.19 (0.29)	0.11 (0.53)	0.13 (0.45)	0.13 (0.44)	0.14 (0.42)
Brazil	0.39 (0.04)**	0.35 (0.06)*	0.41 (0.03)**	0.41 (0.03)**	0.36 (0.05)**	0.32 (0.08)*	0.38 (0.04)**	0.38 (0.04)**
Chile	0.42 (0.03)**	0.48 (0.02)**	0.38 (0.04)**	0.38 (0.04)**	0.35 (0.07)*	0.39 (0.06)*	0.31 (0.11)	0.32 (0.11)
Czech	0.07 (0.67)	0.12 (0.52)	0.14 (0.45)	0.14 (0.43)	0.02 (0.92)	0.06 (0.73)	0.08 (0.67)	0.08 (0.64)
Hungary	-0.12 (0.48)	-0.10 (0.56)	-0.09 (0.61)	-0.09 (0.62)	-0.17 (0.34)	-0.16 (0.38)	-0.15 (0.43)	-0.15 (0.43)
Korea	0.32 (0.06)*	0.30 (0.09)*	0.31 (0.08)*	0.31 (0.08)*	0.27 (0.13)	0.24 (0.17)	0.26 (0.14)	0.26 (0.13)
Mexico	0.10 (0.49)	0.10 (0.48)	0.10 (0.51)	0.10 (0.51)	0.05 (0.77)	0.05 (0.76)	0.04 (0.78)	0.04 (0.78)
Poland	-0.02 (0.99)	0.00 (0.91)	0.02 (0.84)	0.03 (0.80)	-0.07 (0.82)	-0.04 (0.90)	-0.03 (0.98)	-0.02 (1.00)
Singapore	0.25 (0.09)*	0.31 (0.04)**	0.29 (0.05)**	0.29 (0.05)**	0.08 (0.64)	0.14 (0.37)	0.12 (0.44)	0.12 (0.44)
South Africa	0.17 (0.24)	0.24 (0.12)	0.22 (0.14)	0.22 (0.14)	0.05 (0.71)	0.11 (0.43)	0.09 (0.51)	0.10 (0.51)
Taiwan	0.81 (0.00)***	0.83 (0.00)***	0.81 (0.00)***	0.81 (0.00)***	0.73 (0.00)***	0.76 (0.00)***	0.74 (0.00)***	0.74 (0.00)***
Thailand	0.42 (0.02)**	0.46 (0.01)***	0.42 (0.01)***	0.42 (0.02)**	0.34 (0.06)*	0.38 (0.04)**	0.34 (0.05)**	0.34 (0.06)*
Turkey	0.64 (0.00)***	0.65 (0.00)***	0.65 (0.00)***	0.65 (0.00)***	0.58 (0.00)***	0.59 (0.00)***	0.59 (0.00)***	0.59 (0.00)***
Average	0.34	0.36	0.35	0.35	0.28	0.3	0.29	0.29

Table II: Performance of $mfFDR^+$ based portfolio with various FDR targets. The table shows the annualized Sharpe ratios and mean returns (before and after transaction cost) by implementing the $mfFDR^+$ on all strategies in all currencies to control the FDR at 10% (Panel A), 20% (Panel B), 30% (Panel C) and 40% (Panel D). The selected out-performing strategies then are combined by currency. The fund is allocated to trade on each of currencies having out-performing strategies with weighted by size of the selected out-performing trading rules. The first row presents the numbers of whole sample period while the rests are those of sub-periods.

Period	Excess Sharpe Ratio	Net Sharpe Ratio	Excess Return	Net Return
Panel A: FDR target of 10%				
Whole period	1.02	0.92	4.45%	4.04%
1973-1980	1.66	1.58	5.91%	5.63%
1981-1990	2.02	1.93	7.77%	7.40%
1991-2000	0.85	0.77	5.12%	4.61%
2001-2010	0.60	0.45	2.35%	1.75%
2011-2020	0.37	0.30	1.45%	1.17%
Panel B: FDR target of 20%				
Whole period	1.09	0.98	3.84%	3.47%
1973-1980	1.69	1.60	5.26%	5.00%
1981-1990	1.99	1.89	6.86%	6.51%
1991-2000	0.95	0.85	4.22%	3.76%
2001-2010	0.65	0.50	2.20%	1.67%
2011-2020	0.33	0.25	0.99%	0.75%
Panel C: FDR target of 30%				
Whole period	1.13	1.02	3.55%	3.19%
1973-1980	1.70	1.62	4.93%	4.68%
1981-1990	1.94	1.84	6.31%	5.97%
1991-2000	1.03	0.91	3.83%	3.39%
2001-2010	0.68	0.52	2.09%	1.60%
2011-2020	0.34	0.25	0.87%	0.65%
Panel D: FDR target of 40%				
Whole period	1.14	1.02	3.33%	2.99%
1973-1980	1.67	1.58	4.64%	4.39%
1981-1990	1.90	1.80	5.96%	5.62%
1991-2000	1.08	0.95	3.58%	3.15%
2001-2010	0.70	0.53	2.04%	1.56%
2011-2020	0.31	0.22	0.75%	0.54%

IC. Performance of $mfFDR$-based portfolios with use of mean excess return as the testing performance metric

As a robustness check, we repeat all experiments presented in main manuscript with use of mean excess return as performance metric in hypothesis testing (ϕ). Table III present the OOS performance of the $mfFDR$-based portfolios, both before and after transaction cost, when implementing the method on individual currencies to control FDR at targets of 10%, 20%, 30% and 40%.

Table III: Performance of $mfFDR$ based portfolios on individual currency with varying FDR target when mean return is used as the testing performance metric. The table shows annualized Sharpe ratios of the $mfFDR$ based portfolio with FDR target $\tau = \{0.1, 0.2, 0.3, 0.4\}$ based on portfolios' returns before (left side) and after transaction cost (right side). The final row shows the average Sharpe ratio across 30 portfolios corresponding to the 30 currencies. The numbers in parentheses are the corresponding $p-$values. "*", "**" and "***" respectively indicate statistical significance at levels of $10\%, 5\%$ and 1%.

Countries	Before transaction cost				After transaction cost			
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$
Australia	0.20 (0.11)	0.20 (0.10)*	0.20 (0.11)	0.19 (0.11)	0.16 (0.19)	0.16 (0.20)	0.15 (0.21)	0.15 (0.22)
Canada	0.19 (0.15)	0.18 (0.17)	0.18 (0.17)	0.18 (0.17)	0.13 (0.31)	0.12 (0.39)	0.12 (0.41)	0.12 (0.38)
Germany/E.U.	0.45 (0.00)***	0.46 (0.00)***	0.45 (0.00)***	0.45 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.41 (0.01)***
Japan	0.46 (0.00)***	0.45 (0.00)***	0.45 (0.00)***	0.45 (0.00)***	0.43 (0.00)***	0.41 (0.01)***	0.42 (0.01)***	0.41 (0.01)***
New Zealand	0.33 (0.01)***	0.35 (0.00)***	0.35 (0.00)***	0.36 (0.00)***	0.27 (0.03)**	0.29 (0.02)**	0.30 (0.02)**	0.30 (0.02)**
Norway	0.21 (0.09)*	0.19 (0.14)	0.18 (0.14)	0.18 (0.14)	0.16 (0.20)	0.14 (0.27)	0.13 (0.29)	0.13 (0.29)
Sweden	0.37 (0.01)***	0.40 (0.01)***	0.40 (0.01)***	0.40 (0.01)***	0.32 (0.01)***	0.35 (0.01)***	0.35 (0.01)***	0.35 (0.01)***
Switzerland	0.30 (0.04)**	0.30 (0.02)**	0.30 (0.03)**	0.29 (0.03)**	0.26 (0.06)*	0.26 (0.06)*	0.26 (0.07)*	0.26 (0.07)*
U.K.	0.32 (0.02)**	0.30 (0.04)**	0.30 (0.05)**	0.30 (0.05)**	0.28 (0.05)**	0.27 (0.06)*	0.26 (0.07)*	0.26 (0.07)*
Argentina	0.22 (0.29)	0.24 (0.25)	0.24 (0.23)	0.24 (0.23)	0.18 (0.37)	0.19 (0.33)	0.20 (0.31)	0.20 (0.31)
Columbia	0.52 (0.01)***	0.57 (0.00)***	0.57 (0.00)***	0.57 (0.00)***	0.45 (0.03)**	0.50 (0.01)***	0.51 (0.01)***	0.50 (0.01)***
India	0.37 (0.02)**	0.35 (0.03)**	0.35 (0.03)**	0.35 (0.03)**	0.33 (0.05)**	0.31 (0.06)*	0.31 (0.06)*	0.30 (0.06)*
Indonesia	0.33 (0.02)**	0.34 (0.02)**	0.33 (0.02)**	0.33 (0.02)**	0.24 (0.10)*	0.24 (0.10)*	0.24 (0.10)*	0.24 (0.10)*
Israel	0.30 (0.04)**	0.31 (0.03)**	0.31 (0.03)**	0.31 (0.03)**	0.12 (0.38)	0.13 (0.32)	0.13 (0.33)	0.13 (0.33)
Philippines	0.66 (0.00)***	0.64 (0.00)***	0.63 (0.00)***	0.63 (0.00)***	0.49 (0.01)***	0.47 (0.01)***	0.46 (0.02)**	0.47 (0.02)**
Romania	0.16 (0.39)	0.20 (0.28)	0.20 (0.25)	0.20 (0.26)	0.05 (0.78)	0.08 (0.65)	0.09 (0.62)	0.09 (0.62)
Russia	0.40 (0.09)*	0.41 (0.08)*	0.42 (0.08)*	0.42 (0.08)*	0.38 (0.10)*	0.39 (0.08)*	0.40 (0.08)*	0.40 (0.08)*
Slovak	0.14 (0.42)	0.15 (0.40)	0.16 (0.37)	0.17 (0.36)	0.10 (0.58)	0.10 (0.56)	0.11 (0.54)	0.12 (0.50)
Brazil	0.35 (0.08)*	0.35 (0.09)*	0.34 (0.09)*	0.34 (0.09)*	0.32 (0.11)	0.31 (0.11)	0.31 (0.11)	0.31 (0.11)
Chile	0.44 (0.03)**	0.40 (0.04)**	0.40 (0.05)**	0.40 (0.05)**	0.36 (0.08)*	0.32 (0.11)	0.31 (0.12)	0.31 (0.12)
Czech	0.07 (0.64)	0.11 (0.51)	0.12 (0.47)	0.12 (0.46)	0.01 (0.88)	0.05 (0.71)	0.06 (0.67)	0.06 (0.64)
Hungary	-0.13 (0.49)	-0.12 (0.54)	-0.12 (0.55)	-0.12 (0.57)	-0.18 (0.35)	-0.18 (0.36)	-0.17 (0.37)	-0.17 (0.38)
Korea	0.32 (0.12)	0.30 (0.14)	0.29 (0.15)	0.29 (0.15)	0.26 (0.19)	0.24 (0.21)	0.24 (0.21)	0.24 (0.21)
Mexico	0.08 (0.55)	0.08 (0.55)	0.08 (0.57)	0.08 (0.56)	0.03 (0.78)	0.03 (0.80)	0.03 (0.81)	0.03 (0.80)
Poland	-0.05 (0.78)	-0.01 (0.98)	-0.01 (0.98)	0.00 (0.98)	-0.10 (0.59)	-0.05 (0.78)	-0.06 (0.78)	-0.05 (0.81)
Singapore	0.27 (0.05)**	0.32 (0.03)**	0.32 (0.03)**	0.32 (0.03)**	0.11 (0.54)	0.15 (0.35)	0.14 (0.37)	0.14 (0.37)
South Africa	0.17 (0.24)	0.21 (0.16)	0.23 (0.12)	0.24 (0.11)	0.05 (0.71)	0.09 (0.53)	0.11 (0.46)	0.11 (0.45)
Taiwan	0.78 (0.00)***	0.77 (0.00)***	0.77 (0.00)***	0.77 (0.00)***	0.70 (0.00)***	0.70 (0.00)***	0.69 (0.00)***	0.69 (0.00)***
Thailand	0.46 (0.02)**	0.47 (0.01)***	0.47 (0.01)***	0.47 (0.01)***	0.37 (0.05)**	0.39 (0.04)**	0.39 (0.04)**	0.38 (0.04)**
Turkey	0.48 (0.00)***	0.48 (0.00)***	0.48 (0.00)***	0.48 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.42 (0.00)***	0.42 (0.00)***
Average	0.31	0.31	0.31	0.31	0.24	0.24	0.24	0.24

Table IV exhibits the results when implementing the method on all currencies together.

Table IV: Performance of $mfFDR^{+}$ based portfolio with various FDR targets when mean return is used as the testing performance metric. The table shows the annualized Sharpe ratios and mean returns (before and after transaction cost) and break-even point (bps) by implementing the $mfFDR^{+}$ on all strategies in all currencies to control the FDR at 10% (Panel A), 20% (Panel B), 30% (Panel C) and 40% (Panel D). The selected out-performing strategies then are combined by currency. The fund is allocated to trade on each of currencies having out-performing strategies with weighted by size of the selected out-performing trading rules. The first row presents the numbers of whole sample period while the rests are those of sub-periods.

Period	Excess Sharpe Ratio	Net Sharpe Ratio	Excess Return	Net Return	Break-even
Panel A: FDR target of 10%					
Whole Period	1.00	0.90	4.48	4.04	65
1973-1980	1.43	1.34	5.01	4.69	78
1981-1990	2.09	1.99	8.36	7.96	141
1991-2000	0.85	0.76	5.27	4.71	89
2001-2010	0.63	0.47	2.51	1.89	34
2011-2020	0.35	0.28	1.37	1.08	11
Panel B: FDR target of 20%					
Whole Period	1.08	0.97	3.87	3.47	60
1973-1980	1.45	1.36	4.47	4.18	75
1981-1990	2.07	1.97	7.33	6.95	128
1991-2000	0.95	0.84	4.31	3.80	75
2001-2010	0.69	0.53	2.35	1.80	34
2011-2020	0.34	0.26	1.03	0.78	10
Panel C: FDR target of 30%					
Whole Period	1.12	1	3.55	3.17	57
1973-1980	1.47	1.38	4.24	3.97	75
1981-1990	2.02	1.91	6.71	6.34	119
1991-2000	1.02	0.89	3.85	3.37	68
2001-2010	0.72	0.55	2.24	1.73	34
2011-2020	0.34	0.25	0.88	0.65	9
Panel D: FDR target of 40%					
Whole Period	1.13	1	3.36	2.99	55
1973-1980	1.46	1.36	4.06	3.80	75
1981-1990	1.99	1.87	6.33	5.97	113
1991-2000	1.08	0.94	3.62	3.16	65
2001-2010	0.73	0.56	2.18	1.68	33
2011-2020	0.31	0.22	0.76	0.54	8