Technical Analysis, Spread Trading and Data Snooping Control

Ioannis Psaradellis[a]*, Jason Laws[b], Athanasios A. Pantelous[c], Georgios Sermpinis[d]

Version: 17th June 2020

Abstract

This paper utilizes a large universe of 18,410 technical trading rules (TTRs) and adopts a technique that controls for *false discoveries* to evaluate the performance of frequently traded spreads using daily data over 1990–2016. For the first time, the paper applies an excessive out-of-sample analysis in different subperiods across all TTRs examined. For commodity spreads, the evidence of significant predictability appears much stronger compared to equity and currency spreads. Out-of-sample performance of portfolios of significant rules typically exceeds transaction cost estimates and generates a Sharpe ratio of 3.67 in 2016. In general, we reject previous studies' evidence of a uniformly monotonic downward trend in the selection of predictive TTRs over 1990–2016.

JEL classification: C12, C53, G11, G14, G15

Keywords: Technical Trading Rules; Spread Trading Predictability; False Discovery Rate; Bootstrap Test; Portfolio Performance

---

We thank Stephen J. Brown, Ernest P., Robert Faff, John Fan, Chris Florackis, Po-Hsuan Hsu, Neil Kellard, Paul Lajbcygier, Marcel Prokopczuk, Michalis Stamatogiannis and Chardin Wese Simen; the participants at the 11th International Conference on Computational and Financial Econometrics (CFE 2017), 8th National Conference of the Financial Engineering and Banking Society (FEBS 2017), 24th International Conference of Forecasting Financial Markets (FFM 2017), and Quantitative Finance and Risk Analysis Symposium (QFRA 2017); and the participants at Monash University and the Universities of St Andrews, the ICMA Centre, Liverpool, Shanghai, Xi'an Jiaotong-Liverpool University, and Glasgow for helpful comments. Any remaining errors are our own.

[1] *Corresponding author: School of Economics & Finance, University of St Andrews, St Andrews, U.K.

[b] Management School, University of Liverpool, Liverpool, U.K.

[c] Department of Econometrics & Business Statistics, Monash Business School, Monash University, Melbourne, Australia

[d] Adam Smith Business School, University of Glasgow, Glasgow, U.K.

1. Introduction

Technical trading believed to be one of the longest-established forms of investment analysis, which explores future trading opportunities for financial assets, just by analyzing the asset prices' time-series history. Numerous studies have revisited the predictability of technical analysis across several markets.[1] On long-term systematic investing, successful fund managers and institutional traders employ technical trading to capture major trends and reversals of market price. Recent evidence about hedge fund performance also reveals that technical analysis, when employed by innovative classes of investors, yields higher performance, lower risk, and superior market-timing ability than do other fund strategies, especially in high-sentiment periods (Smith et al. 2016).

On the other hand, spread trading is a relative-value arbitrage strategy with a 35-year plus history. Over that time, the strategy has remained popular among hedge fund and investment managers and it constitutes one of three broad types of strategies used by hedge funds (see Pedersen, 2015).[2] Such a strategy can generate excess returns by going long one asset, while going short another one to yield profits, because of a short-term deviation of their relative valuations from a perceived equilibrium.[3] Practitioners frequently employ technical analysis to measure such a correlated relationship between two assets and trade the spread process (see among others, Vidyamurthy, 2004; Bogomolov, 2013).

In this regard, we try to assess the predictability and genuine out-of-sample (OOS) profitability of disciplined, consistent *technical trading rules* (TTRs) on spread trading. In particular, we investigate the evolution of OOS outperformance of spread trading using daily data, including

[1] See among others, Brock et al. (1992), Neely et al. (1997), Allen and Karjalainen (1999), Sullivan, Timmermann, and White (STW) (1999) Lo et al. (2000), Kavajecz and Odders-White (2004), Bajgrowicz and Scaillet (BS) (2012), Kuang, Schroder and Wang (2014), Hsu, Taylor, and Wang (HTW) (2016). More than half of them plead in favor of technical trading's significant predictability, whereas the rest argue against it.
[2] The other two broad types of hedge fund strategies are equity and macro strategies.
[3] Popular statistical arbitrage studies include also those of Shleifer and Vishny (1997), Hogan et al. (2004), Gatev et al. (2006), Kondor (2009), Avellaneda and Lee (2010), Do and Faff (2010) and Chen et al. (2017).

frequently traded spreads on *commodities*, *equity indices*, and *currencies*, while taking into account their pairwise correlations (Chen et al. 2017), over the period from January 1990 to December 2016. For that purpose, we construct and employ an *up-to-date* and *wider* universe of more than 18,000 TTRs on spreads. Those are capable of capturing major statistical properties through automated rules aimed at identifying overbought/oversold levels (e.g., Bollinger bands and commodity channel indices) and significant trends (e.g., moving averages, filter rules) within the constructed spreads.

In order to assess the significance of technical trading performance, we utilize one of the most powerful and up-to-date multiple-hypothesis testing frameworks to control for data-snooping effects. In particular, *data snooping* becomes a considerable concern when recruiting a big dataset, whose number of variables (i.e., trading rules) is larger than the number of observations. Contrary to the classical statistical inference in which a single hypothesis testing is highly likely to trigger false discoveries, due to the enormous amount of information constantly utilized by investors, we employ the *false discovery rate control* ($FDR^{+/-}$) of Barras, Scaillet, and Wermers (BSW) (2010, 2019) to identify significantly outperforming trading rules based on their Sharpe ratio criterion, while we consider realistic transaction costs. Contrary to its recent computationally intensive counterparts (i.e., BRC test of White, 2000; Step M test of Romano and Wolf, 2005; SSPA test of Hsu, Hsu, & Kuan, 2010), the $FDR^{+/-}$ is a simple, fast approach, while on the other hand it is efficient in terms of yielding only a very small bias when estimating the proportion of rules with zero performance (BSW, 2019). This results in the $FDR^{+/-}$ test being appropriate for finance applications as it is powerful in identifying models with true outperformance even if the *best-model*'s performance is a result of pure lack (see also, Chordia, Goyal and Saretto, 2020). We also use the *manipulation-proof performance measure* (MPPM) of Ingersoll et al. (2007) to validate that OOS performance of significant TTRs is not the result of manipulated time series properties, but true outperformance (see also Goetzman et al., 2002). As far as we are aware, this is the first time for such approach being employed in the field of spread trading.

Except from evaluating the predictability of TTRs over the full sample, we also separate our dataset into five subperiods, of no more than 5 years each, based on major historical events, using each subperiod's last year of daily data as the OOS period. Doing so helps us to optimally distinguish between IS and OOS horizons, to achieve a reasonable balance between *Type I* and *II errors*, while being aligned with the recent findings of Harvey and Liu (2015) on backtesting and true OOS experiments. In this way, we add an *extra* layer of accuracy to the TTRs selection, on top of the established power of the $FDR^{+/-}$ method, while we provide an evaluation of the time-varying predictability of TTRs on spread trading. We create combined market portfolios consisting of commodity, equity, currency spreads, to explore the OOS profitability and diversification benefits for an investor exposed in every single market separately and in all markets together at the same time. In addition, we compare the economic significance of our results with that of the results of TTRs on spreads constructed using a *dynamic hedging* technique for robustness purposes.

Although the major picture of OOS performance of TTRs on individual spreads shows a scattered predictive ability, portfolios of significant rules on *commodity* pairs (i.e., heating oil-gas oil) are still able to consistently achieve healthy Sharpe ratios (e.g., 4.29) over the recent years, whereas specific foreign exchange pairs (i.e., EUR-JPY) yield similar Sharpe ratios over 2016. Their corresponding outstanding MPPMs (above 2%) reveal that such a performance is probably robust rather than manipulation due to unpriced risk. On the contrary, the majority of the equity indices' spreads seem to underperform most of the time except from one case (i.e. CAC-TOPIX).

A *global* portfolio of TTRs traded on all asset classes' spreads, yields consistently positive Sharpe ratios and MPPMs and across all OOS horizons, which validates that our findings are manipulation-*free*.

Additionally, we examine a group of other *stylized facts*, derived from the technical analysis literature (BS; HTW), that state that the predictability of technical analysis has shrunk over time, because of informational efficiency or due to the potential *self-destruction* of price patterns once they have been realized and exploited by a mass of traders. We find that this is not always the case, especially when forecasting and trading commodity spreads or equally-weighted spread portfolios from all asset classes. Moreover, we try to uncover potential drivers of the above significant performance of technical analysis for spread trading. We achieve that by assessing whether the yielded returns can be explained by risk factors, such as those of Carhart's (1997) four-factor model and Asenss's et al. (2013) value and momentum *everywhere* factors. We also evaluate the effect of investors' sentiment, market volatility and funding and market liquidity shocks on spreads' trading performance. The evidence reveals a positive and significant co-movement with market volatility, and so, higher spread trading performance during periods of turmoil. Spread trading also seems to outperform when borrowing is difficult, while its negative loading on market liquidity shocks reveals that its profitability is probably a result of taking contrarian positions to *crowded* trades and providing liquidity to abandoned securities with low price and high expected returns when those type of shocks exist. Finally, we present results on extra robustness checks of a break-even analysis of transaction costs and subperiod analysis of TTRs predictability based on in-sample evidence in our Online Appendix.

The remainder of the paper is organized as follows. Section 2 provides the statistical data analysis and pairs formation. Section 3 and Appendix introduce the TTRs that are implemented.

Section 4 focuses on the transaction costs, and performance metrics employed. Section 5 presents the issue of data-snooping bias. Section 6 presents the empirical findings including the spreads' full sample and OOS performance as well as the robustness checks. Section 7 concludes.

2. Data and descriptive statistics

2.1. Data and pairs formation

We consider daily data on the pairs employed to construct these spreads, which include those between the closing prices of commodities, equity indices, and currencies. In total, we examine 15 pairs, including five spreads from each of the three markets considered. The full list of the examined spreads can be found in Table 1. We use the universe of 45 commodity, equity and currency series of Moskowitz et al. (2012) who examine time series momentum in various asset classes, to pool the spread series under investigation. We then follow the approach of Chen et al. (2017) and compute the pairwise correlations between every single series within the same asset class (e.g., commodities). In particular, we calculate the Pearson correlation coefficients between the returns of a every single series and the rest series within same asset class from January 1, 1990, to December 12, 2016.[4] Finally, we keep the top five spreads with the highest correlations as an evidence that those pairs move closely together, while at the same time they should be popular among statistical arbitrage investors.[5] The employed cross-currency pairs are U.S. dollar denominated (i.e., U.S. dollars per unit of foreign currency), whereas for the equity spreads, we directly use the equity indices instead of any corresponding exchange-traded funds (ETFs),

[4] Except in the case of heating oil-gas oil, where the sample period starts on 1995 due to data availability.

[5] Moskowitz et al. (2012) focus on the most liquid instruments to avoid returns being contaminated by illiquidity or stale price issues. This is to match more closely an implementable strategy at a significant trade size. We also communicated with fund managers and statistical arbitrage investors, who confirmed that our chosen pairs are those frequently advertised by trading websites, or launched by financial market companies, such as the CME and ICE groups.

following the previous literature on technical analysis. Furthermore, for our commodity spreads, we employ the continuous price series of each commodity futures contract, offered by Thomson Reuters Datastream, with the nearest deliverable contract forming the first value in the series.[6] We also consider daily data on short-term interest rates for every currency to calculate currency returns. We used WM/Reuters FX benchmarks to acquire data on the foreign exchange rates and Thomson Reuters Datastream to acquire the closing prices of the commodities futures and equity indices listed above.

To evaluate the technical analysis used in spread trading, it is essential to make sure that the issue of nonsimultaneous pricing that often plagues such simulations doesn't exist. Indeed, all the examined commodities contracts have the same trading hours as they are listed on the CBOT (agriculture), NYMEX (energy and precious metals), and COMEX (precious metals) exchanges, which constitute the CME group derivatives marketplace.[7] Each European equity spread consists of equity indices issued by the same or different stock exchanges (i.e., London and Frankfurt Stock Exchanges and Euronext Paris), which have the same actual closing times after taking into account the 1-hour time difference. The same holds for each of the considered U.S. indices, which are mainly issued by NYSE. In addition to this, because we obtain our foreign exchange rates data from the WM/Reuters FX benchmarks, all of them represent the closing spot rates, fixed daily at 16:00 GMT.

Market participants can construct spreads in various ways, depending on their principal investment goal. In our study, we pair any two assets by subtracting the closing price of one

[6] When the first day of the delivery month is reached, we roll to the nearest deliverable contract. Doing so ensures that the underlying instrument should last longer than the observation period when analyzing the performance of technical trading and that no issues of nonsynchronous trading exist.
[7] For example, the closing times for agricultural futures are 13:20 and 07:45, whereas that for energy and precious metals is 17:00 EST.

underlying leg from the other, because our main aim is to capture their dominant trends and reversals through technical analysis. This means that we allocate an equal proportion of our wealth to each side. Thus, the formation of a pair S_t, in which we go long a risky asset P_1 and short another risky asset P_2 at time t, is $S_t = P_{1,t} - P_{2,t}$. However, in the case of commodity spreads, we must consider whether both commodity futures contracts are traded in different units before we start calculating the spreads, and if not, to adjust for that.[8] For the equity and exchange rate spreads, such an issue does not exist.[9] Hence, we do not follow a specific rule and just calculate the difference between the spot prices of these assets.

2.2. Descriptive statistics and statistical behavior

Table 1 reports the descriptive statistics of the daily returns on all the spreads formed, along with the pairwise correlation regarding the time series of each pair's underlying legs.[10] Regarding the former, the annualized mean and standard deviation along with the *p*-values from the first-order autocorrelation under the Ljung-Box (1978) Q statistics of the daily returns of each spread are reported. Regarding the latter, we calculate the correlation coefficient between every two series forming a spread to assess their co-movement.[11]

---

[8] For example, for the platinum-gold we use a 2:1 ratio, because the gold and palladium futures contract unit is 100 troy ounces, whereas that of the platinum contract is 50 troy ounces. We apply similar transformations for the rest if the two components are traded in different units.

[9] In particular, the order we place the legs here does not make a great deal of difference in our findings. This is because TTRs employed capture trends or reversals on the spreads' series and the order the legs are placed will not change the final outcome.

[10] The pairwise correlations presented here have been calculated based on the full sample time series (i.e., January 1990 - December 2016). For our OOS performance exercise, which separates the full sample in five different subsamples, the pairwise correlations have been calculated again and the spreads selected with respect to the highest correlation coefficients remain the same across all subperiods.

[11] We have also tested for cointegration ranking between every two series forming a spread to further assess their co-movement. We find that, mostly commodity spreads reject the null hypothesis for zero-rank cointegration. The results are available upon request.

[Insert Table 1 around here]

In terms of performance, the mean returns vary across spreads in every asset class. Commodity and currency spreads are dominated by negative returns, while equity spreads by positive returns. In terms of the standard deviation of daily gross returns, not surprisingly commodity and equity spreads are more volatile than the currency spreads. However, there are also considerable deviations of volatility levels even within the same asset class, especially in the case of equity spreads. The Ljung-Box test for residual autocorrelation in daily gross returns indicates persistence in all the cases, at least at the 10% significance level. We translate this into the existence of trends for the majority of the spreads considered. In the cases of the equity and currency pairs the statistical significance is even stronger at the 1% and 5% significance level, respectively. This evidence strongly supports the use of also trend-following TTRs. In terms of pairwise correlations between the two legs of a spread, we find that commodity spreads report the highest correlation coefficients on average, with a couple of cases even having perfect correlation, while the equity spreads follow. The currency spreads present the lowest correlation coefficients on average compared to the rest asset classes, but again the levels of the pairs selected are quite high.

3. TTR universe

We consider *seven* families of TTRs based on past price data of the computed pairs, as they are widely used by commodities, equities, and foreign exchange traders.[12] Those classes of rules are categorized into *momentum/trend-following* rules and *contrarian/mean-reverting* rules, usually

[12] We do not apply rules utilizing volumes of transactions, such as the on-balance volume averages (see STW), because it is difficult to accurately observe the exact volume of a pair of assets, which are not actively traded in the market.

employed by pair traders trying to identify *overbought/oversold levels*. The momentum/trend-following rules include the following: *filter rules*, with the main characteristic of following strong trends by taking long (short) positions accordingly; *moving averages*, which attempt to ride trends and take positions when crossovers occur, between the pair value and its moving average of a given length or between two moving averages of different lengths, signifying a break in the trend; *support and resistance rules*, which try to identify breaches in a pair's price through local maximums (minimums), triggering further price movements in the same direction and leading to long (short) signals; and *channel breakouts*, similar to having time-varying support and resistance levels that form a channel of fixed percentage, leading to a signal when a pair's price penetrates the channel from above or below. The contrarian/mean-reverting rules include the following: *relative strength indicators* (RSIs), which belong to the general family of *overbought/oversold* indicators and attempt to capture a correction in the opposite direction of a pair's extreme price movement; *Bollinger band reversals*, which attempt to identify overbought and oversold market levels, defined as price being a particular distance away from its moving average of a given length, in terms of standard deviations; and *commodity channel index (CCI) rules*, which are similar to a combination of RSIs and Bollinger band reversals and try to quantify the connection between a pair's price, its corresponding moving average, and its standard deviation, but where a specific inverse factor is used to scale the index.

Following previous studies (STW; BS; HTW), we consider numerous variations of the above TTRs and a spectrum of different plausible parameterizations of each variation. These possible TTRs form a large universe, totaling 18,412 including 1,932 filter rules, 7,920 moving averages, 2,310 support and resistance, 2,250 channel breakouts, 730 RSIs, 2,160 Bollinger bands, and 1,110

CCI rules. Online Appendix A presents the exact details of each class of trading rules, their variations, and the various parameterizations examined.

4. Excess returns, transaction costs, and performance metrics

Before we assess the performance of the TTRs, we must first compute the daily excess return obtained from buying and holding each spread (i.e., buying the first underlying asset and selling the second simultaneously), for each prediction period. To estimate the daily gross and therefore the investment performance of spread trading, we employ simple returns rather than logs, because they are additive in the cross section and in pairs formation. For the commodity and equity spreads, the calculation of their daily excess return is the daily gross return of a self-financing portfolio:

$$r_t = \left[\frac{(P_{1,t} - P_{1,t-1})}{P_{1,t-1}} - \frac{(P_{2,t} - P_{2,t-1})}{P_{2,t-1}}\right]. \tag{1}$$

where r_t is the daily gross return from buying and holding the pair for 1 day; $P_{1,t}$ and $P_{2,t}$ are the spot prices of the first and second components, respectively, on day t; and $P_{1,t-1}$ and $P_{2,t-1}$ are the spot prices of the two components on day $t-1$. To calculate the daily excess return for currency spreads, we follow HTW, and take into account the short-term interest rates of each currency. Thus, we consider two parts to construct the excess returns: the simple return of each component, over the holding period and the interest rate carry related to borrowing one unit of domestic currency and lending one unit of foreign currency overnight. We also transform the annualized short-term interest rates into daily rates for our application.

Now, let $s_{j,t-1}$ denote the trading signal generated from the trading rule $j, 1 \leq j \leq l$ (where $l = 18{,}412$) at the end of the prediction period $t-1$ $(\tau \leq t \leq T)$, which depends on the information given, where $s_{j,t-1} = 1, 0,$ or -1 represents a long, neutral, or short position taken at time t. We use the *Sharpe ratio* criterion as the main performance metric for creating portfolios of significant TTRs (see also STW; BS) and the MPPM to assess whether any outperformance is a result of manipulated returns due to unpriced risk or excessive leverage. The Sharpe ratio as being a relative performance criterion SR_j for trading rule j at time t is defined by

$$SR_{j,t} = \frac{\bar{f}_j}{\widehat{\sigma_j}}, \; j = 1, \dots, l, \tag{2}$$

where $\bar{f}_{j,t}$ and $\widehat{\sigma_{j,t}}$ are the mean excess return and the estimated standard deviation of the mean excess return, respectively. The mean excess return criterion $\bar{f}_{j,t}$ for the trading rule j is given by

$$\bar{f}_{j,t} = \frac{1}{N} \sum_{t=R}^{T} s_{j,t-1} r_t \, , \; j = 1, \dots, l, \tag{3}$$

where $N = T - \tau + 1$ is the number of days examined and τ denotes the start date of each subperiod. We consider that some of the TTRs employ lagged values up to 1 year (252 days). Furthermore, the Sharpe ratio is strictly connected to the observable *t*-statistic of the empirical distribution of a strategy's returns, which makes this metric appropriate for our multiple-hypothesis-testing framework (Harvey and Liu, 2015).[13]

Proposed by Ingersoll et al. (2007), the MPPM measures the return on a fund's strategy relative to the amount of accepted risk over a specific time period, while resembling a representative utility function. The measure actually takes into account return distributions, which are far from normal or lognormal and could result in *concave payoffs*. As already mentioned, those payoffs often lead to a performance that looks superior. Hence, the reason for employing such a measure is to verify that potentially generated OOS Sharpe ratios are not inflated. The MPPM is

[13] The *t*-statistic of a given sample of historical returns $(r_1, r_2, \dots r_t)$, testing the null hypothesis that the average excess return is zero, is usually defined as $t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{T}}$, whereas the corresponding Sharpe ratio is given by the formula $SR = \frac{\hat{\mu}}{\hat{\sigma}}$.

strictly increasing and concave in order to prevent manipulation. The measure is also time separable, allowing it to avoid the dynamic manipulation of the estimated statistic, and the measure has a power form, allowing it to be consistent with an economic equilibrium (Ingersoll et al. 2007). The MPPM formula, which represents the annualized continuously compounded excess return certainty equivalent of the portfolio, is defined as

$$MPPM = \frac{1}{(1-\rho)\Delta t} ln\left[\frac{1}{T}\sum_{t=R}^{T}(1+s_{j,t-1}r_t)^{1-\rho}\right], \tag{4}$$

where ρ denotes a constant risk tolerance parameter, which historically takes values between 2 and 4; Δt denotes the length of time between observations (in our case $\Delta t = 1/252$ for daily returns); and r_t denotes each pair's excess return series, which have been calculated above. Morningstar (2006) and Ingersoll et al. (2007) conclude that the level of $\rho = 2$ results in performance metrics consistent with the risk tolerances of typical retail investors. For this reason, we mainly present results based on this level.[14]

So far, we have not considered the impact of transaction costs on the TTRs' performance over the examined spreads. In practice, these costs may be quite significant, especially for statistical arbitrage traders who form long-short portfolios. Additionally, potential predictability of a selected strategy before the implementation of transaction costs can be easily neutralized when those costs are adjusted through the selection process, sometimes because of the impact of frequent signals (Timmerman and Granger 2004). Thus, we handle transaction costs *endogenously* to the selection process. In particular, we subtract the transaction costs every time a buy or sell signal is triggered based on the prediction of the corresponding spread. This comes down separately considering the one-way transaction costs of each component.

[14] We also tested more conservative levels of risk tolerance (i.e., ρ =3 and 4), and we can report that the results were slightly different. The relevant findings are also available on request.

Following Locke and Venkatesh (1997), we consider one-way transaction costs of 3.3 basis points for a position taken on each of the commodity futures to construct the commodity spreads. Furthermore, we assume that an investor funds her position with 100% equity rather than using a margin account, because we measure daily returns as the difference in the relative prices (Miffre and Rallis 2007). In terms of stock indices trading costs, we are aware of the complexity of precisely measuring the transaction costs, which have declined over time. Earlier studies use one-way transaction costs ranging from 10 to 30 basis points for the trading of U.S. stock indices (Allen and Karjalainen 1999). However, recent evidence for the *live* trading costs faced by real-world arbitrageurs are quite a bit lower.[15] Based on the findings of Frazzini et al. (2015), as well as communications with several brokerage firms, we consider 19.73 basis points as the one-way trading costs for stock indices. This level of costs represents the value-weighted mean (i.e., weighted by the dollar value of the trades) of the market impact estimate of a long-short portfolio traded on live data, similar to a pairs trading strategy. By trading currencies, the only transaction costs investors do face arise from the bid-ask spreads in spot exchange rates and interest rates (no fixed brokerage costs). Following Neely and Weller (2013) and HTW, we calculate the one-way transaction costs for each currency from the corresponding bid-ask spread in the forward exchange rates on any particular day. Specifically, we use one-third of the quoted 1-month forward rate bid-ask spread for each currency. Several studies have shown that posted bid-ask spreads are usually larger than the effective ones actually traded (Lyons, 2001; Neely and Weller, 2003). This results in an average one-way transaction cost of 4 basis points for all the developed economies tested.

[15] For example, Frazzini et al. (2015) estimate that market impact, which covers the greatest part of the cost of execution for a large institutional trader, is under 9 basis points, on average, for trades executed during a day, whereas the rest of the costs (e.g., commissions, bid-ask spreads) are almost negligible, especially for large trade sizes, because they do not increase analogously with size.

5. Data-snooping bias

5.1. Defining the data-snooping bias and existing data-snooping methods

Also known as the data-mining issue, the data-snooping bias has become even more urgent as of late because of the use of large datasets by investors and researchers. Large datasets can lead to promising results, sometimes even by chance. Data replication issues lead to chance results because it is quite easy to incorrectly identify an outperforming trading rule. Classical statistical inference focusing on single-hypothesis testing for each strategy, without paying attention to the performance of the remaining ones, can lead to false rejections, or the so-called *Type I error*, due to the extensive specification search. Multiple-hypothesis frameworks, developed to limit such occurrences, are more than necessary nowadays. Recently, Harvey (2017) raised this issue as the *p*-hacking phenomenon (i.e., frequent false significant *p*-values) and explained that new, adjusted *p*-values reflecting the genuine significance of an investment strategy should be defined. Large universes of TTRs provide a breeding ground for testing the power of multiple-hypothesis methods, because it is quite likely that one will discover a rule that works well, even by chance, especially within a specific family of rules (see, among others, STW; Hsu, Hsu, and Kuan, 2010; BS).

Studies trying to control for data snooping bias are divided in two classes, based on the main criterion employed for statistical inference, those using the *family-wise error rate* (FWER) and those using the *false discovery rate* (FDR). Although, in the case of FWER, numerous developments have been made in the financial literature to achieve a good trade-off between true positives (i.e., *Type II errors*) and false selections (i.e., *Type I errors*) (see among others, White, 2000; Hansen, 2005; Romano and Wolf, 2005; Hsu, Hsu, and Kuan, 2010; Hsu et al. 2014) the FWER control is not that suitable for finance applications as it lacks power and is mechanically affected by the number of hypotheses being tested (Chordia, Goyal and Saretto, 2020).[16] For instance, the FWER methods are more conservative and they do not select further rules once they have detected a rule whose performance is due to luck. On the other hand, the FDR, by definition, tolerates a certain proportion of false rejections, so as to improves the power of detecting more significant discoveries, while having an optimal balance in minimizing *Type I* and *II errors* (Abramovich et al. 2006). This feature makes the FDR a more powerful multiple-hypothesis-testing tool than FWER. In finance, BSW propose a modified FDR$^{+/-}$, based on Storey's (2004) FDR approach, that is aimed at discovering significant alphas in mutual fund performance, while allows a separate quantification of false discoveries among funds performing better and worse than the benchmark, respectively. By developing such a framework, BSW try to accommodate what investors do in practice, who usually assess and combine multiple strategies (not just the best strategy) at any given time, to diversify against model risk. BS employ the FDR$^{+/-}$ approach in the context of identifying outperforming TTRs on the DJIA index and their findings confirm the above comparative advantage of the FDR$^{+/-}$ to find outperforming rules over the FWER methods. Another important feature of the FDR method is that it also holds under *weak dependence* conditions, when the number of hypotheses is very large (Benjamini and Yekuteli 2001; Storey 2002; Storey et al. 2004), due to asymptotics. This is a key assumption for our study because the TTRs included in our universe satisfy this feature by being dependent on small blocks, within the same family (e.g., filter rules), but essentially independent across different families (BS).[17] For the

[16] BS and Chordia, Goyal and Saretto (2020) also validate the superiority of FDR against FWER methods by executing numerous Monte Carlo experiments for both approaches. We have also run the same experiment and the above finding remains the same.

[17] Even though the FDR$^{+/-}$ has recently received criticism on its ability to detect the proportions of funds with zero and significant performance, as their generated estimators found to be biased when the number of series' observations is small (e.g. monthly) (see, Andrikogiannopoulou and Papakonstantinou 2019), BSW (2019) have convincingly shown that FDR$^{+/-}$ is still capable of producing efficient and consistent estimators on normal conditions and especially when the number of time series observations is large, such as in our case, at which daily series of multiple years returns are considered.

above reasons, we adopt the FDR$^{+/-}$ test as the most suitable, simple and fast multiple-hypothesis-testing method in the context of minimizing the data-snooping effects arising from the application of our large TTR universe in spread trading.

5.2. FDR methodology

The FDR$^{+/-}$ approach concentrates on estimating the proportion of false discoveries among trading strategies performing better or worse than the benchmark (e.g., the risk-free rate), while displaying genuine performance under a specific level of statistical significance. As a multiple-hypothesis-testing procedure, the value of the test statistic for each rule j, φ_j, defines the null hypothesis in which rule j does *not* outperform the benchmark (i.e., $H_{0j}: \varphi_j = 0$),, while the alternative hypothesis assumes the presence of abnormal performance, either positive or negative (i.e., $H_{Aj}: \varphi_j > 0 \; or \; \varphi_j < 0$).[18] In our case, we consider the annualized Sharpe ratio as test statistic for performing the multiple hypothesis testing. Now let R^+ be the number of significantly positive rules and F^+ the number of erroneous selections among them. The FDR$^+$ concentrating on the rules generating positive returns, is then defined as the expected value of the ratio of false selections to the number of outperforming rules. Thus, the estimate of FDR$^+$ is given by $\widehat{FDR}^+ = \frac{\widehat{F^+}}{\widehat{R^+}}$, where $\hat{F}^+$ and $\hat{R}^+$ are the estimators of $F^+$and R^+, respectively.[19]

[18] Because we use the Sharpe ratio as the performance metric, our benchmark is, by definition, the *risk-free* rate, describing an investor who is out of the market.

[19] Similarly, we can compute a separate estimator of the FDR^- among the rules generating negative returns. Doing so, however, would be outside of the scope of this paper.

For the estimation of the FDR$^+$, we just need to estimate the number of lucky rules, F^+, in the right tail of the distribution of performance metrics, φ_j, at a given significance level γ. This is given by

$$\hat{F}^+ = \pi_0 * l * \gamma/2, \tag{5}$$

where π_0 is the proportion of rules showing no abnormal performance, in the entire universe, of size l, and $\gamma/2$ is the *p*-value cutoff assuming symmetry in terms of the appearance of nongenuine rules based on luck in the two tails. The FDR method actually tries to capture information from the center of the distribution of test statistics (i.e., φ_j), mostly dominated by lucky rules, to correct any luck in the two tails (BSW). For this reason, an accurate estimator of π_0 is the key point for the FDR$^+$ approach. Storey's (2002) main assumption that the true null *p*-values are uniformly distributed over the interval [0,1], whereas the *p*-values of alternative models lie close to zero, in a two-sided setup, helps us to define the estimator of π_0 as

$$\widehat{\pi_0}(\lambda) = \frac{\#\{p_j > \lambda;\ j = 1, \dots, l\}}{l(1-\lambda)}, \tag{6}$$

where $\lambda \in [0,1)$ is a tuning parameter indicating the specific level above which the null *p*-values should appear. After acquiring a conservative estimator for $\hat{F}^+$, and given the number of significantly positive TTRs (i.e., $\#\{p_j \leq \gamma, \varphi_j > 0;\ \ j = 1, \dots, l\}$), we can obtain a conservative estimator for FDR$^+$ under threshold γ of

$$\widehat{FDR}^+(\gamma) = \hat{F}^+/\hat{R}^+ = \frac{1/2\widehat{\pi_0} l\gamma}{\#\{p_j \leq \gamma, \varphi_j > 0;\ j = 1, \dots, l\}}. \tag{7}$$

The FDR$^+$ method requires *p*-values, p_j for $1 \leq j \leq l$, from a two-tailed test, because the main parameter we need to estimate is the proportion of rules with no abnormal performance, π_0, in the total universe. Furthermore, and because we require no prior knowledge of the exact distribution of *p*-values the stationary bootstrap resampling technique is used to obtain the individual *p*-values.

This resampling method is mainly applicable to stationary weakly dependent time-series data as it works by sampling blocks of varying length of consecutive observations of returns (Politis and Romano 1994).[20] Then the corresponding *p*-value of each rule is given by comparing the original Sharpe ratio with the quantiles of simulated Sharpe ratios centered by the original value. Nevertheless, the critical part of the FDR method is to identify the right *p*-value cutoff γ by controlling the FDR^+ at a fixed predetermined level (i.e., 10%), to isolate the genuinely outperforming rules from the total population. We achieve this by employing the *point estimates* approach of Storey et al. (2004) under the weak-dependence condition.

6. Empirical findings

6.1. Portfolio construction and full sample performance

We now try to measure the empirical predictability of technical analysis on spread trading, based on the full 25-year sample period. After utilizing the universe of 18,412 TTRs on every single spread and asset class, we apply the FDR^+ as described in the previous section not only to identify and measure the performance of the best rule, but also to assess its ability as portfolio optimization tool. We construct portfolios of significant rules by setting the $\widehat{FDR^+}$ equal to 10%, as a good trade-off between truly outperforming TTRs and poorly chosen ones, similar to BS. Thus, we build a 10%-FDR^+ portfolio of TTRs for each pair, meaning that 90% of all the portfolio's rules significantly outperform the benchmark. We pool the signals of the chosen rules with equal weight, similarly to a forecast averaging technique (allocating $1 evenly).[21] Finally, we

[20] During our empirical simulations, we set the stationary bootstrap parameters as B = 1,000 realizations and the average block length equal to 0.1 (i.e., q = 10).

[21] We do not attribute more weight to the most outperforming rules, because doing so would result in a deviation from the desired FDR level and likewise the selection of fewer strategies. Equally weighted portfolio strategies are also difficult to be beaten, especially when data snooping bias is controlled (Hsu et al. 2018).

treat the neutral signals as totally liquidating our positions and do not invest a proportion of wealth corresponding to them at the *risk-free* rate. This assumption helps us to measure the true performance of the FDR portfolios.

Table 2 provides evidence for the full sample performance of 10%-FDR^{+} portfolio of significant rules for each spread over the 1990-2016 period. In particular, we present the number of predictive rules, the Sharpe ratio and MPPM of every single portfolio. Additionally, the Sharpe ratio and its corresponding *p*-value (in parenthesis) and the family of the best significant rule found for each spread are also reported in the two last columns.

[Insert Table 2 around here]

In general, technical analysis predictability appears significantly strong for all spreads considered with commodity spreads being the most predictable in terms of significant rules selected. Equity spreads are the next most predictable and currency spreads seems to follow with the least predictive rules in their corresponding portfolios. Considering the performance, the overall picture is the same. Commodity spreads, yield the highest annualized Sharpe ratios and MPPMs, denoted excess returns, compared to the equity and currency families of spreads. In particular, commodities' Sharpe ratios and MPPMs range from 0.72 to 1.06 and 0.22% and 3.98%, respectively. The equity spreads' portfolios Sharpe ratios and MPPMs range from 0.40 to 1.39 and 0.05% to 3.74%, respectively, while relevant performance metrics for currency spreads range from 0.34 to 0.86 and 0.04% to 0.29%, respectively. The top performing spread across all assets considered is the Brent-WTI crude oil or the so-called *crack spread*, with the CAC-TOPIX, equity spread following closely in performance.

Considering the information given for each spread's best performing rules, all of them are statistically significant. Analogous, to the previous evidence, the corresponding best rule for each one of the commodity spreads is statistically significant at the 1% level. For the top rules of equity and currency spreads the picture is almost the same with few cases (i.e., FTSE100-CAC 40, EUR-CHF and EUR-JPY) being significant only at the 10% nominal level. What is interesting here is the comparison of the Sharpe ratio of the portfolios of significant rules with that of the best significant rule. For the majority of cases, the corresponding Sharpe ratios of the FDR^{+} portfolios are better or at least almost equal to those of the best performing rule for each spread. Such a finding reveals the diversification benefits of the FDR^{+} approach as portfolio construction tool. Pinpointing now which TTRs contribute the lion's share of the best predictive rule, the majority of them belong to two contrarian families for the commodity pairs, namely, the RSIs and the Bollinger bands. For the equity spreads, a similar finding is evident, with CCIs being the highest-performing rules. On the contrary, among the currency spreads, we observe a variety of families of best performing rules. For example, moving-averages, support and resistance, channel breakouts, RSIs as and Bollinger bands are among the top ones. In overall, mean-reverting rules, especially those holding the rule's signal for a certain period, seem the best performing for the majority of the spreads examined.

6.2. Out-of-sample performance

To evaluate the genuine performance of the TTRs and to address the issue of data snooping in more realistic conditions, we employ an OOS analysis in the following sections. Doing so helps us to economically evaluate the performance of a portfolio of rules, selected ex ante, similar to how institutional investors would do so in practice.[22] Such an analysis also provides evidence in terms of the performance persistence and the economic evaluation of the TTRs, even in an artificial post-sample period.

We separate the whole sample into five historical subperiods- 1991-1996, 1997-2001, 2002-2007, 2008-2011, and 2012-2016 -for our OOS experiment, where we investigate different market dynamics in a more sensible algorithmic trading application. Although our sample starts from 1990, this year is not included in our first subsample because we require data going back 1 year to generate some of the TTRs. Despite the fact that the above subperiods may be of different sizes, they are closely related to major historical events for all markets considered, namely, the Maastricht Treaty in 1992, the East Asian currency crisis in 1997, the *dot-com* bubble in 1999-2000 and the subsequent 2002 credit crunch, the appearance of the euro in 2002 and the 2003-2007 energy crisis, the global financial crisis of 2008, and, finally, the recent crude oil downturn in 2014.

Before we move forward with the results, another important issue with the OOS estimation, raised by Harvey and Liu (2015), is the splitting of the dataset between the IS and OOS segments. This estimation procedure usually comes down to a trade-off between *Type I* (false discoveries) and *Type II* (missed discoveries) *errors*, a trade-off that is closely related to the testing power of the IS and OOS periods. In particular, the shorter is the IS dataset, the greater will be the chance of missing true discoveries (*Type II errors*), and vice versa. For instance, a 90-10 split of the data will lead to an increase in *Type I errors*, whereas a 50-50 split will lead to an increase in *Type II*

[22] We find also the same spreads as the most correlated when we implementing Chen's et al. (2017) method on the time series universe of Moskowitz et al. (2012) for each subperiod separately.

errors. Although multiple-hypothesis-testing frameworks aim to minimize these types of errors, we adopt a 70-30 split for the IS and OOS intervals to secure a good balance between them.[23]

The FDR procedure is employed during the IS period for each pair in order to select the TTRs for evaluation in the OOS horizon. Specifically, we construct the portfolios of rules by selecting them as in the previous section and we build a 10%-FDR^{+} equally-weighted portfolio of TTRs for each spread, using 70% of each subperiod's historical data. The last 30% is used for the OOS estimation. This approach provides us with almost the whole of the last year as the OOS horizon for every subperiod, whereas the previous years (no more than 4 years) constitute the IS period. Although we appreciate that this is still a stringent OOS evaluation, it better matches what traders do in practice, as opposed to previous studies, which use only a single long-term OOS horizon of many years (see HTW among others).

Table 3 reports the median number of significant rules for each portfolio and across all periods, the OOS performance of the equally weighted FDR portfolios of significant rules based on the Sharpe ratio criterion, and their computed MPPM. The results are reported for each spread examined.

[Insert Table 3 around here]

First, Table 3 supports the ability of the FDR method to select a sufficient number of predictive rules across all subperiods and for each pair. Specifically, the commodities spreads seem to be more predictable, with an average median of their portfolios being close to 407 rules, whereas, for equity and currency spreads, this number is considerably lower, at 24 and 11 rules, respectively. In general, the OOS performance of the TTRs on the commodity pairs is, on average, higher than

[23] Following Harvey and Liu (2015), who mention that 90-10 and 50-50 splits lead to an increase in *Type I* and *II errors*, respectively, we choose a split midway between these two, so as to gain an optimal balance.

the almost equal performance of the TTRs on the equity and foreign exchange pairs. The FDR portfolio can produce a small profit in many cases, with the MPPM being totally consistent with Sharpe ratio findings, leaving no room for manipulating for manipulation of excess returns.

Concentrate on the commodity pairs. Each pair has at least one post-sample period, in which the FDR portfolio of significant TTRs yields a positive Sharpe ratio. The MPPM findings advocate that performance, indicating no manipulation of those ratios, which range from 0.79 to a very healthy 4.29, while the MPPMs range from 1.31 to an outstanding 22.5%. The most promising spread is the heating oil-gas oil, which yields consistently very healthy Sharpe ratios (above 1) and an average MPPM of 15.8% for all the examined post-sample periods. Interestingly, its performance seems to be enhanced over the more recent periods, especially in the last period when a Sharpe ratio of 4.29 and an MPPM of 22.5% are generated. The Brent-crude oil spread follows by providing almost equally positive metrics. However, its positive performance concentrates only in the first two periods. The metal spreads, seem also to yield a positive performance over the last three periods, but of a lower magnitude, which could be still found attractive, while the corn-soybean spread performs worst.

Considering the results for the equity pairs, the overall picture shows a weak OOS performance across all the periods, in general, apart from a few cases. For instance, only the CAC-TOPIX spread yields consistently positive performance, which reaches a peak over the last three subperiods, by reporting an outstanding maximum Sharpe ratio of 3.00 and an MPPM of 10.3%. Then only the FTSE100-CAC 40 in 1996 and the DAX-FTSE100 spread in 2001 seem to demonstrate positive performance metrics, but of a small magnitude. For the rest cases, negative Sharpe ratios and MPPMs are generated across all five post-sample years.

In terms of the OOS performance of currency spreads, the results seem more encouraging for most cases, with portfolios of significant rules yielding positive Sharpe ratios and MPPMs ranging from 0.07 to 1.41 and 0.01 and 3.60% respectively, at least during the first three OOS periods. The EUR-JPY pair, for example, decays in performance through the years, whereas the EUR-CHF pair, as another example, shows considerable Sharpe ratios cyclically, reaching its top performance on 2011, with a Sharpe ratio of 1.00 and an MPPM of 3.60%. Moreover, NOK-SEK seems able to yield a Sharpe ratio of 1.41 in 2011.

Based on these findings, seasonal market inefficiencies seem more appropriate in explaining the overall OOS performance of the TTRs on the corresponding pairs. Another potential reason could be the existence of an unobservable extreme tail risk event, which has been exploited by the TTRs. Spread traders can exploit and arbitrage away returns in specific periods of time. On the other hand, these returns tend to diminish in the subsequent periods, especially when more traders deploy their strategies and alleviating the existing arbitrage opportunities. However, this is not consistent to commodity spreads' findings.

6.3. Robustness checks

6.3.1. Out-of-sample performance under dynamic hedge ratios

For robustness purposes on OOS performance, we exercise another pairs trading technique used by traders. The technique captures information from the co-movement of the two assets, to dynamically create optimal hedge ratios to guard against risk when trading such portfolios. Thus, a pair is now defined as $S_t = P_{1,t} - \beta * P_{2,t}$, where β is the regression coefficient similar to the study of Chen et al. (2017), which actually serves as the optimal hedge ratio[24]. We keep the same

[24] We have also tried another trading approach spread traders do in practice, which sets the lagged values of contrarian rules equal to the half-life of mean-reversion of the spread. However, the findings are poor for the examined spreads. The relevant table is available upon request.

order in the components used in the naïve case, to determine the dependent and independent variables. We describe this procedure as *dynamic hedging*, because it is time dependent in terms of the correlation of the assets employed. We perform the *dynamic hedging* approach for every subperiod separately and before forming our 10%-FDR^{+} portfolios.

Table 4 presents the OOS findings for the TTRs' performance using *dynamic hedging*. We also report the hedge ratio for each spread except from the median size of each spread-portfolio and its corresponding performance metrics.

[Insert Table 4 around here]

Once again, the above evidence suggests that the FDR portfolios present an adequate median number of genuine rules for all the examined spreads across the five post-sample periods. Such a result is also supported by the quite high MPPMs observed. In particular and like in previous sections, the commodities spreads are more predictable when using technical analysis, whereas the equity and currency spreads are almost equally predictable (an average median portfolio of 22 rules). The optimal hedge ratios exceed 1 for the majority of pairs and across all markets considered.

By looking at the overall commodities performance, the dynamic hedging technique slightly improves the performance metrics for the case of oil spreads, relative to the simple 1:1 spread trading technique presented in Table 3. It seems to add value on the fact that the Brent-WTI crude oil spread yields positive Sharpe ratios and MPPMs even during the recent periods. However, this is not exactly the case for the metal spreads, who are profitable in fewer subperiods even though the achieve higher MPPMs in some cases compared to those in Table 3. What is interesting though is the positive performance of corn-soybean spread, this time in three out of five periods. The overall Sharpe ratios are ranging from 0.21 to 4.42, whereas the positive MPPMs span from 0.23% to an outstanding 24.9%.

In terms of the equity spreads, it is evident that the *dynamic hedging* approach performs better than the one without hedging as more spreads have become profitable even though such profits are of small magnitude. For example, the portfolio of TTRs on S&P500-DAX spread, is profitable in almost all subperiods and it achieves a top performance generating a Sharpe ratio and MPPM of 1.69 and 2.15% respectively, in 2007. On the other hand, the outstanding performance of the CAC-TOPIX spread seems to have dropped across all periods.

Finally, for the currency pairs, we notice a similar pattern, but, this time, the positive performance of TTRs has been enhanced in some cases and diminished in others. All the examined spreads denote outperformance for at least one post-sample period, except from 1996. However, this is not stable over time. Despite this, the outperformance of more than half of the currency spreads is concentrated around three specific years, including the most recent periods (i.e., 2001, 2011, and 2016), in which we notice outstanding Sharpe ratios, such as 3.83 for EUR-JPY in 2016 and 2.17 for EUR-CHF in 2011. Their corresponding MPPMs (i.e., 4.88% and 2.43%) also reveal that the ratios are not the result of manipulation due to *concave payoffs*.

6.3.2 OOS performance for market portfolios of pairs

The final OOS simulation involves the performance examination of integrated market portfolios of TTRs across all post-sample periods. Usually, spread traders not only expose themselves to a single market but are constantly searching for arbitrage opportunities across several markets and assets. Thus, we construct four integrated portfolios based on FDR^{+} selections, for commodity, equity and currency pairs, respectively and global one, which includes all. We compare also their performance with identical integrated portfolios constructed under the dynamic hedging technique for robustness purposes. Again, we assume no optimization for the integrated portfolios, by assigning equal proportions of our total wealth to every single rule identified as significant across all spreads. For instance, we invest $1 of our total wealth and distribute it evenly across all spreads' significant rules in a particular portfolio (e.g., the commodities portfolio). We present the relevant performance including also the *compounded annual growth rate* (CAGR), except from the Sharpe ratio and MPPM metrics, of the four market portfolios, across every single OOS subperiod in Table 5.

[Insert Table 5 around here]

In general, both the 1:1 and dynamic hedging approaches demonstrate similar performance, with the former one scoring slightly higher Sharpe rations and excess returns for most cases. Once again, it is obvious that the commodities portfolios outperform the rest asset portfolios by far and provide very attractive investment vehicles for spread traders across all the periods. It is remarkable that even during the latest years the commodities portfolios yield on average annualized Sharpe ratios of almost 3.00 and annualized manipulation-free excess returns of almost 4%. The equity indices portfolios follow, but their profitability concentrates mostly on the recent years, since they report negative performance metrics over the earlier subperiods. Their significant profitability, especially during the recent years, could be driven by the outstanding performance of the CAC-TOPIX spread rather than the rest of equity spreads. The currency portfolios of significant TTRs report the smallest positive performance, which is evident in all post-sample years except from the 1996. However, their corresponding Sharpe ratios and excess returns are quite low and their top performance under both the 1:1 and *dynamic hedging* approaches appears in 2011. Finally, global portfolios retain a uniformly positive performance, mostly attributed to the commodities portfolios, across all post-sample periods. The diversification benefits we obtain are here quite observable and of great interest especially for spread investors who trade on different assets. This time, however, the positive Sharpe ratios and manipulation-free excess returns fall to levels ranging from 0.43 to 3.67 and 0.41% to 2.11%, respectively for the naïve case. The relevant metrics under the *dynamic hedging* approach are quite similar, showing only small variations around different post-sample periods.

Finally, Fig. 1 provides a graphic interpretation of the IS (see Online Appendix) and the above OOS performances of the global market portfolio based on the 1:1 spread formation approach. We focus on this approach, because it yields the highest returns (i.e., compared to the *dynamic hedging*), as indicated in Table5, to give an extra picture of the performance of our most successful approach.[25] In Figure 1, we depict the cumulative returns of a $1 investment in each subperiod separately, based on the CAGR generated from the global portfolio in both IS and OOS. Doing so, it is similar to reinvesting the value of our portfolio on the daily return. We also separate the cumulative return's evolution in the IS and OOS horizons (dashed line) for every subperiod to highlight the pure post sample performance.

[Insert Figure 1 around here]

The diversification benefits for an investor who performs spread trading in all the three markets are clear. The $1 value of the portfolio could have grown to $1.5 at the end of 1996 and 2001. Similarly, the same dollar could have grown to more than $1.1 at then of 2007, 2011 and 2016.

[25] The relevant graphs for commodity, equity and currency portfolio, as well as those corresponding to the no-hedging setup validate their weaker performance, and they are available on request.

The key feature here is that the portfolios maintain the *upward* positive, performance also during the OOS horizons.

Interestingly, the positive performance of technical analysis on spread trading has not been reduced over the years, especially for commodities, at which it remains almost equal if not increased. Such a finding argues with the previous literature (Neely et al. 2009; BS), which emphasizes the decay in outperformance of technical analysis over time, presumably because of the increased hedge fund and trading activity. This is maybe due to the fact that there is limited evidence on the application TTRs on assets' spreads and so more arbitrage opportunities left to be exploited even nowadays. Notwithstanding this, the overall picture shows the commodities market to be more lucrative for a spread trader, as revealed by the post-sample performance of the commodity portfolios over time and compared with the corresponding performance of the rest of the markets studied.

6.3.2 Drivers of performance

We assess the spreads' risk-adjusted performance, its factor exposures and its relationship with liquidity, market volatility and the level of investor's sentiment on Table 6. Firstly, in Panel A, we employ the Carhart's (1997) four-factor model and we run monthly time series regressions of our 1:1 global portfolio's returns on the size (SMB), value (HML) and cross-sectional momentum (UMD) factors over the 1990-2016 period.[26][27] We construct the global portfolio over the full sample period in a similar manner to the one over the OOS period, while we transform the daily

[26] All factors employed have been downloaded by Kenneth's, R., French website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).
[27] We use monthly regression over the full sample period since most of the factors and indices are available on a monthly basis and by just running the regressions over the OOS periods, we would have left with only few observations to produce robust findings.

returns to monthly by taking the average return of the portfolio for each month. We use the Carhart's (1997) four-factor model because it also includes the momentum factor, due to the existence of trend-following apart from contrarian rules in the total universe of TTRs. Panel B reports the results of a similar regression, but this time on the value and momentum *everywhere* cross-sectional factors of Asness et al. (2013). Finally, we assess in Panel C how the global portfolio returns co-vary with the time series levels and monthly changes (first differences) of the investor's sentiment index of Baker and Wurgler (2006, 2007), the time series log values of VIX index as a measure of market volatility, especially in periods of turmoil, and the levels of funding and marker liquidity shocks in separate regressions.[28][29] In a similar manner to Asness et al. (2013), we use as funding liquidity proxy the negative of the Treasury-Eurodollar (TED) spread and as market liquidity proxy the Pastor and Stambaugh (2003) liquidity factor (i.e., the innovations in aggregate liquidity)[30]. In order to obtain the funding and market liquidity shocks, we compute the residuals from the AR(2) process applied on those proxies.[31]

[Insert Table 6 around here]

Panel A of Table 6 demonstrates significantly negative loadings with the market index and the HML as well as the UMD factors. However, those are of a very small magnitude to fully explain the spread portfolio's returns, a finding which is also supported by the small value of adjusted-R^2

[28]The investor's sentiment index of Baker and Wurgler (2006, 2007) has been downloaded from Jeffrey Wurgler's website (http://people.stern.nyu.edu/jwurgler/).

[29] We have run the same time series regressions for commodity, equity and currency spread portfolios of TTRs. The relevant results are available upon request.

[30] The TED spread data has been downloaded from Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/TEDRATE), while the Pastor and Stambaugh (2003) liquidity factor from Robert F. Stambaugh's website (http://finance.wharton.upenn.edu/~stambaug/).

[31] We have also applied the Fama and French three-factor model including each of the liquidity variables on the global portfolio returns and the estimated coefficients are similar.

(i.e., 6.73%). Our global portfolio also delivers a significant alpha of 0.04%. The Panel B reveals a similar picture. The betas of the market index and value and momentum *everywhere* factors are also significantly negative, but of low levels, supporting the previous findings. The alpha of the model remains almost at the same level (i.e., 0.05%), but the corresponding adjusted-R^2 is slightly higher at 8.01%. In terms of the relationship between the global portfolio of TTRs' returns and the investor's sentiment index, there is no significant effect in both levels and first differences. However, this is not the case between the portfolio's returns and VIX index. There seems to be a significantly positive relationship of a small magnitude. In other words, spread trading could be more profitable during periods of high market volatility. When it comes to the relationship of global portfolio returns and liquidity shocks, this is significantly negative for both funding and liquidity risk, with the funding liquidity having a greater impact on spread trading. Hence, spread trading performs better when funding liquidity drops or, in other words, when borrowing is not easy. The negative loading on market liquidity shocks indicates no positive liquidity risk premium given as compensation for taking such a type of risk and so its interpretation seems more of a puzzle. A potential explanation could also support that the most profitable rules are contrarian. Pastor and Stambaugh (2003) realize a significantly positive relation between U.S. equity momentum returns and their liquidity shocks. Towards this direction, contrarian trading strategies usually provide liquidity to demand pressure that could be triggered by strategies like momentum, which put more price pressure on *crowded* trades during liquidity shocks. In such conditions contrarian rules provide liquidity by buying low and selling high, a strategy which can yield higher returns (see also, Pedersen, 2015).

7. Conclusion

We investigate a hedge fund trading strategy based on the high correlation of two assets, while employing technical analysis to predict the price movements of the constructed spreads. For that purpose, we conducted large-scale research on the full sample and OOS performance of TTRs across a set of commodity, equity, and currency spreads being actively traded by statistical arbitrageurs over the 1990 - 2016 period. Our analysis involves quite a large number of TTRs split between generic trend-following and contrarian classes.

To mitigate the data mining problem arising from the usage of such a large pool of predictive rules, we adopt a recently developed, simple and efficient multiple hypothesis testing method, namely $FDR^{+/-}$ of BSW (2010, 2019), which allows us to create statistical inferences to generate new, adjusted thresholds for significant *t*-statistics. Additionally, we employ an MPPM to assess whether OOS performance is illusory because of some unpriced risk or because of a product of pure skill.

Our findings reveal that technical trading still yields significant Sharpe ratios for many of the spreads considered. Those ratios are consistently followed by positive MPPMs of almost analogous magnitude. Commodity pairs consistently outperform the equity and currency ones. The OOS analysis, conducted across five different subperiods, reveals that technical analysis performance hasn't worsened over time, with commodity and currency spread displaying outperformance even in recent periods. Hence, increased hedge fund activity, through which a mass exercise of trading rules, have not squeezed out potential returns. Combined in a global portfolio, we demonstrate a consistent positive performance over the OOS periods examined. The economic significance of the returns achieved using TTRs on certain spreads and periods may be a compensation for short-term market inefficiencies. We try to shed more light on the main drivers of the above performance by running time series regressions between the returns of the global portfolio and famous risk factors, such as those of Fama and French (1993) and Carhart (1997) as well as the very recent value and momentum *everywhere* factors of Asness et al. (2013). The findings report a significantly negative relationship between global portfolio's returns and the momentum factor. We perform similar regressions by also considering the investor's sentiment index of Baker and Wurgler (2006, 2007), VIX index and funding and market liquidity shocks. The evidence suggests that TTRs performance on spread trading is significantly driven by the market volatility and possibly related to its contrarian nature of buying less liquid securities with lower prices and higher expected returns compared to more *crowded securities during market liquidity shocks.*

References

Abramovich, F., Benjamini, Y., Donoho, D.L. & Johnstone, I.M. (2006). Special invited lecture: Adapting to unknown sparsity by controlling the false discovery rate. Annals of Statistics, 34, 584–653. DOI: 10.1214/009053606000000074

Allen, F. & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51, 245–271. https://doi.org/10.1016/S0304-405X(98)00052-X

Andrikogiannopoulou, A. & F. Papakonstantinou (2019), Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? Journal of Finance, Replications and Corrigenda, 75, 2667-2688. https://doi.org/10.1111/jofi.12784

Asness, C.S., Moskowitz, T.J. & Pedersen, L.H. (2013). Value and Momentum Everywhere. Journal of Finance, 68, 929–985. https://doi.org/10.1111/jofi.12021

Avellaneda, M. & Lee, J.H. (2010). Statistical arbitrage in the US equities market. Quantitative Finance, 10, 761–782. https://doi.org/10.1080/14697680903124632

Bajgrowicz, P. & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106, 473–491. https://doi.org/10.1016/j.jfineco.2012.06.001

Baker, M. & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M. & Wurgler, J. (2007). Investor sentiment in the stock market. Journal of Economic Perspectives 21, 129–157. DOI: 10.1257/jep.21.2.129

Barras, L., Scaillet, O. & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance, 65, 179–216. https://doi.org/10.1111/j.1540-6261.2009.01527.x

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

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics, 29, 1165–1188. doi:10.1214/aos/1013699998

Bogomolov. T. (2013). Pairs trading based on statistical variability of the spread process, Quantitative Finance, 13, 1411-1430. DOI: 10.1080/14697688.2012.748934

Brock, W., Lakonishok, J. & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47, 1731–1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

Carhart, M.M. & (1997). On persistence in mutual fund performance. Journal of Finance, 52, 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chen, H., Chen, S., Chen, Z. & Li, F. (2017). Empirical investigation of an equity pairs trading strategy. Management Science, 65, 19–31. https://doi.org/10.1287/mnsc.2017.2825

Chordia, T., Goyal, A. & Saretto, A. (2020). Anomalies and False Rejections, The Review of Financial Studies, 33, 5, 2134–2179. https://doi.org/10.1093/rfs/hhaa018

Do, B. & Faff, R. (2010). Does simple pairs trading still work? Finance Analyst Journal, 66, 83, 95. https://doi.org/10.2469/faj.v66.n4.1

Frazzini, A., Israel, R. & Moskowitz, T.J. (2015). Trading costs of asset pricing anomalies. Working paper, Chicago University.

Gatev, E., Goetzmann, W.N. & Rouwenhorst, K.G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies, 19, 797–827. https://doi.org/10.1093/rfs/hhj020

Goetzmann, W., Ingersoll, J., Spiegel, M. I. & Welch, I. (2002). Sharpening Sharpe ratios. Working paper, Yale School of Management.

Hansen, P.R. (2005). A test for superior predictive ability. Journal of Business & Economic Statistics, 23, 365–380. https://doi.org/10.1198/073500105000000063

Harvey, C.R. (2017). The scientific outlook in financial economics. Journal of Finance, 72, 1399–1440. https://doi.org/10.1111/jofi.12530

Harvey, C.R. & Liu, Y. (2015). Backtesting. Journal of Portfolio Management, 42, 13–28. https://doi.org/10.3905/jpm.2015.42.1.013

Hogan, S., Jarrow, R., Teo, R. & Warachka, M. (2004). Testing marking efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial Economics, 73 525-565. https://doi.org/10.1016/j.jfineco.2003.10.004

Hsu, P.H., Han, Q., Wu, W. & Cao, Z. (2018). Asset allocation strategies, data snooping, and the 1 / N rule. Journal of Banking & Finance, 97, 257-269. https://doi.org/10.1016/j.jbankfin.2018.09.021

Hsu, P.H., Hsu, Y.C. & 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. https://doi.org/10.1016/j.jempfin.2010.01.001

Hsu, P.H., Taylor, M.P. &Wang, Z. (2016). Technical trading: Is it still beating the foreign exchange market? Journal of International Economics, 102, 188-208. https://doi.org/10.1016/j.jinteco.2016.03.012

Hsu, Y.C., Kuan, C.M., & Yen, M.F. (2014). A Generalized Stepwise Procedure with Improved Power for Multiple Inequalities Testing. Journal of Financial Econometrics, 12, 730-755. https://doi.org/10.1093/jjfinec/nbu014

Ingersoll, J., Spiegel, M., Goetzmann, W., & Welch, I. (2007), Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies, 20, 1503–1546. https://doi.org/10.1093/rfs/hhm025.

Kavajecz, K. & Odders-White, E.R. (2004). Technical analysis and liquidity provision. Review of Financial Studies, 17(4), 1043–1071. https://doi.org/10.1093/rfs/hhg057

Kondor, P. (2009). Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading. Journal of Finance, 64, 631-655. https://doi.org/10.1111/j.1540-6261.2009.01445.x

Kuang, P., Schroder, M. & Wang, Q. (2014). Illusory profitability of technical analysis in emerging foreign exchange markets. International Journal of Forecasting, 30, 192–205. 10.1016/j.ijforecast.2013.07.015

Ljung, G.M. & Box, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika, 65, 297–303. https://doi.org/10.1093/biomet/65.2.297

Lo, A.W., Mamaysky, H. & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. Journal of Finance 55, 1705–1765. https://doi.org/10.1111/0022-1082.00265

Locke, P.R. & Venkatesh, P.C. (1997). Futures market transaction costs. Journal of Futures Markets, 17, 229–245. https://doi.org/10.1002/(SICI)1096-9934(199704)17:2<229::AID-FUT5>3.0.CO;2-L

Lyons, R.K. (2001). The microstructure approach to exchange rates. MIT Press, Cambridge, MA.

McLean, R.D. & Pontiff, J. (2016). Does academic research destroy stock return predictability? Journal of Finance, 71, 5-32. https://doi.org/10.1111/jofi.12365

Miffre, J. & Rallis, G. (2007). Momentum strategies in commodity futures markets. Journal of Banking & Finance, 31, 1863–1886. https://doi.org/10.1016/j.jbankfin.2006.12.005

Morningstar (2006). The Morningstar Rating™ Methodology. Report.

Moskowitz, T.J., Ooi, Y.H. & Pedersen, L.H. (2012). Time series momentum. Journal of Financial Economics, 104, 228-250. https://doi.org/10.1016/j.jfineco.2011.11.003

Neely, C. & Weller, P. Dittmar, R., (1997). Is technical analysis in the foreign exchange market profitable? A genetic programming approach. Journal of Financial and Quantitative Analysis, 32, 405–426. https://doi.org/10.2307/2331231

Neely, C.J. & Weller, P. (2013). Lessons from the evolution of foreign exchange trading strategies. Journal of Banking & Finance, 37, 3783–3798. https://doi.org/10.1016/j.jbankfin.2013.05.029

Neely, C.J. & Weller, P.A. (2003). Intraday technical trading in the foreign exchange market. Journal of International Money & Finance, 22(2), 223–237. https://doi.org/10.1016/S0261-5606(02)00101-8

Neely, C.J., Weller, P.A. & Ulrich, M. (2009). The adaptive markets hypothesis: Evidence from the foreign exchange market. Journal of Financial and Quantitative Analysis, 44, 467–488. https://doi.org/10.1017/S0022109009090103

Pastor, L. & Stambaugh, F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111, 642-685. https://doi.org/10.1086/374184

Pedersen, L. (2015). Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined. Princeton University Press.

Politis, D. & Romano, J. (1994). The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313. https://doi.org/10.1080/01621459.1994.10476870

Romano, J.P. & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. Econometrica, 73, 1237–1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x

Shleifer, A. & Vishny, R. (1997). The limits of arbitrage. Journal of Finance, 52, 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Smith, M.D., Wang, N., Wang, Y., & Zychowicz, J.E. (2016). Sentiment and the effectiveness of technical analysis: Evidence from the hedge fund industry. Journal of Financial and Quantitative Analysis, 53, 1991–2013. https://doi.org/10.1017/S0022109016000843

Storey, J., (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B 64, 479–498. https://doi.org/10.1111/1467-9868.00346

Storey, J., Taylor, J. & Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. Journal of the Royal Statistical Society, Series B 66, 187–205. https://doi.org/10.1111/j.1467-9868.2004.00439.x

Sullivan, R., Timmermann, A. & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance, 54, 1647–1691. https://doi.org/10.1111/0022-1082.00163

Vidyamurthy, G. (2004). Pairs Trading: quantitative methods and analysis. John Wiley & Sons, 217.

Weisman, A. B. (2002). Informationless investing and hedge fund performance measurement bias. Journal of Portfolio Management, 28, 80–91. https://doi.org/10.3905/jpm.2002.319857

White, H. (2000). A reality check for data snooping. Econometrica, 68, 1097–1126. https://doi.org/10.1111/1468-0262.00152

Table 1. Descriptive statistics and statistical behavior of spreads' daily spot prices and returns.

Spreads	**Ann. Mean (%)**	**Ann. SD**	**1st autoc. (*p*-value)**	**Pairwise Correlation**
Commodities				
Brent-WTI crude oil	-0.79%	22.07%	0.00	0.99
Heatoil-Gasoil	1.16%	31.20%	0.00	0.99
Platinum-Gold	0.15%	19.10%	0.03	0.86
Gold-Silver	-3.90%	20.30%	0.08	0.95
Corn-Soybean	-0.27%	22.80%	0.00	0.92
Equities				
FTSE100-CAC 40	-0.48%	11.91%	0.00	0.87
AMST-CAC 40	-1.35%	10.80%	0.00	0.92
S&P500-DAX	1.35%	19.80%	0.00	0.97
CAC-TOPIX	4.90%	25.90%	0.00	0.91
DAX-FTSE100	4.04%	14.70%	0.00	0.91
Currencies				
EUR-CHF	-1.92%	6.51%	0.00	0.69
EUR-JPY	-0.08%	12.04%	0.02	0.55
AUD-CAD	-0.06%	10.16%	0.00	0.91
AUD-NZD	0.72%	7.80%	0.01	0.92
NOK-SEK	0.21%	7.31%	0.00	0.91

We present the descriptive statistics and pairwise correlation of daily returns on holding spreads of different asset classes for the period 1990-2016. The descriptive statistics are the annualized mean return and volatility as well as the *p*-value testing the null of no autocorrelation. The pairwise correlation is assessed by reporting the correlation coefficients between the spot prices of the underlying components of each spread.

Table 2. Predictive ability and outperformance of TTRs for the full 25-year sample period

	# predictive rules	Sharpe Ratio Port.	MPPM Port. (%)	Highest ratio (p-value)	Best Rule
Commodities					
Brent-WTI crude oil	563	0.77	3.98	1.20 (0.00)	BB2
Heatoil-Gasoil	147	0.84	1.07	0.73 (0.00)	SR2
Platinum-Gold	12	0.77	0.22	0.61 (0.00)	RSI2
Gold-Silver	140	0.72	0.86	0.60 (0.00)	SR1
Corn-Soybean	34	1.06	1.34	1.24 (0.00)	RS1
Equities					
FTSE100-CAC 40	27	0.39	0.12	0.31 (0.07)	CCI3
AMST-CAC 40	28	0.65	0.05	0.34 (0.00)	CCI3
S&P500-DAX	29	0.40	0.40	0.40 (0.00)	CCI3
CAC-TOPIX	88	1.39	3.74	1.49 (0.00)	BB1
DAX-FTSE100	8	0.59	0.09	0.36 (0.00)	F1
Currencies					
EUR-CHF	9	0.34	0.12	0.34 (0.06)	MA1
EUR-JPY	16	0.86	0.29	0.34 (0.08)	CB1
AUD-CAD	18	0.65	0.13	0.49 (0.00)	SR2
AUD-NZD	10	0.41	0.08	0.34 (0.00)	RSI2
NOK-SEK	9	0.51	0.03	0.34 (0.00)	BB2

We examine the performance of a total of 18,412 TTRs over the 1990-2016 period after imposing transaction costs. We implement the FDR^+ at a fixed predetermined level (i.e., 10%) to select technical rules providing significantly *positive* performance under the Sharpe ratio performance metric. *#predictive rules* denotes the number of TTRs that provide significantly positive Sharpe ratios. *Highest ratio* denotes the best rule's Sharpe ratio, with p-values in parentheses. *Sharpe Ratio port.* and *MPPM port.* report the performance of the equally-weighted portfolio of predictive rules under the Sharpe ratio and MPPM criteria respectively. The best rules are reported in the *Best rule* section. All MPPMs and Sharpe ratios are annualized.

Table 3. Out-of-sample annualized Sharpe ratios and MPPMs of the FDR^{+} portfolio, of significant in-sample rules.

		1996		**2001**		**2007**		**2011**		**2016**	
	Median Port.	Sharpe ratio	MPPM (%)	Sharpe ratio	MPPM (%)	Sharpe ratio	MPPM (%)	Sharpe ratio	MPPM (%)	Sharpe ratio	MPPM (%)
Commodities											
Brent-WTI crude oil	1,204	**1.14**	**11.5**	**1.72**	**18.37**	-0.01	-0.43	0.00	0.00	0.00	0.00
Heatoil-Gasoil	700	-	-	**1.72**	**10.1**	**3.76**	**13.38**	**3.65**	**17.55**	**4.29**	**22.5**
Platinum-Gold	23	-0.59	-0.15	-0.69	-1.08	-0.25	-0.28	**1.83**	**2.64**	**0.63**	**0.21**
Gold-Silver	77	-0.11	-0.07	-1.18	-1.45	**0.85**	**2.05**	**1.09**	**2.21**	**1.05**	**1.22**
Corn-Soybean	33	-0.41	-1.31	-0.28	-0.08	**0.79**	**1.31**	-1.12	-2.75	0.63	1.81
Equities											
FTSE100-CAC 40	**12**	**1.28**	**0.53**	-0.91	-0.31	-0.81	-0.10	-1.47	-1.25	-1.21	-0.23
AMST-CAC 40	19	-1.51	-0.37	-0.35	-0.15	-0.10	-0.01	-1.61	-0.91	-0.73	-0.06
S&P500-DAX	22	-2.27	-2.11	-0.73	-1.59	0.73	1.34	-3.07	-3.11	-1.55	-1.00
CAC-TOPIX	51	**0.41**	**0.11**	**0.78**	**0.68**	**3.00**	**10.3**	**1.31**	**7.00**	**2.57**	**9.55**
DAX-FTSE100	15.5	-0.94	-0.26	**0.66**	**0.24**	-0.55	-0.13	-0.16	-0.31	-	-
Currencies											
EUR-CHF	**11**	**1.32**	**0.07**	-0.07	-0.02	**0.73**	**0.01**	**1.00**	**3.60**	**0.15**	**0.01**
EUR-JPY	**13**	**1.01**	**0.23**	**0.85**	**0.59**	**0.47**	**0.16**	-0.92	-0.20	**0.70**	**0.11**
AUD-CAD	4	-0.64	-0.66	**0.82**	**0.66**	**0.81**	**0.35**	-0.26	-0.33	0.00	0.00
AUD-NZD	**15**	-1.36	-0.84	-1.74	-0.28	**0.23**	**0.22**	-0.56	-0.05	-1.13	-0.34
NOK-SEK	10	-1.28	-0.41	**0.24**	**0.05**	**0.07**	**0.06**	**1.41**	**0.37**	-2.22	-1.01

We report out-of-sample annualized Sharpe ratios and MPPMs for the last year of each subperiod based on the 10% -FDR^{+} portfolios of significant rules. The rules are those selected in the in-sample horizon, which covers 70% of each subperiod's observations. We impose historical transaction costs on the computations.

Table 4. Out-of-sample annualized Sharpe ratios and MPPMs of the FDR$^+$ portfolio of significant in-sample rules under the dynamic hedge ratio approach.

	Hedge ratio	Median Port.	**1996** Sharpe ratio	MPPM (%)	**2001** Sharpe ratio	MPPM (%)	**2007** Sharpe ratio	MPPM (%)	**2011** Sharpe ratio	MPPM (%)	**2016** Sharpe ratio	MPPM (%)
Commodities												
Brent-WTI crude oil	1.05	269	**1.70**	**15.4**	**1.42**	**8.72**	**0.33**	**1.03**	-0.32	-2.34	**2.02**	**0.82**
Heatoil-Gasoil	0.03	642	-	-	**1.79**	**10.8**	**3.82**	**13.7**	**3.25**	**15.1**	**4.42**	**25.9**
Platinum-Gold	1.15	37	-0.17	-0.02	-0.49	-1.71	-1.34	-2.34	**1.24**	**4.15**	-0.78	-1.24
Gold-Silver	30.1	13	**1.86**	**0.44**	**0.52**	**0.23**	-0.16	-0.31	**0.91**	**4.58**	-0.68	-0.64
Corn-Soybean	2.35	65	**2.25**	**1.89**	-0.21	-0.33	**0.30**	**0.51**	-0.45	-2.24	**0.21**	**0.94**
Equities												
FTSE100-CAC 40	1.27	17	-0.42	-0.27	-0.49	-0.33	-0.66	-0.26	-0.61	-0.62	**0.35**	**0.78**
AMST-CAC 40	11.02	5	-1.21	-0.56	**0.57**	**0.07**	-0.17	-0.08	**0.51**	**0.08**	-2.71	-0.58
S&P500-DAX	4.41	39	**0.58**	**0.21**	**0.71**	**1.92**	**1.69**	**2.15**	-0.75	-2.66	**0.56**	**0.37**
CAC-TOPIX	3.16	30	-1.44	-3.71	**0.77**	**0.89**	**0.76**	**0.28**	**0.27**	**0.91**	**2.17**	**6.87**
DAX-FTSE100	0.91	8	-0.76	-0.54	**0.81**	**0.22**	-0.54	-0.11	-	-	-1.01	-0.12
Currencies												
EUR-CHF	1.39	35	-1.56	-0.37	**0.67**	**0.66**	-0.01	-0.21	**2.17**	**4.88**	-0.67	-0.08
EUR-JPY	1.26	24	-0.53	-0.17	**0.37**	**0.34**	0.02	0.01	**0.43**	**0.24**	**3.83**	**2.43**
AUD-CAD	1.04	21	-0.37	-0.18	-0.58	-0.44	-0.31	**-0.23**	**1.74**	**0.31**	**0.91**	**0.19**
AUD-NZD	1.17	14	-1.48	-0.35	**0.93**	**0.07**	**0.15**	**0.03**	-1.05	-0.31	-1.76	-0.17
NOK-SEK	1.11	2	-1.47	-0.35	**1.09**	**0.23**	-	-	0.59	0.06	-1.63	-0.99

We report out-of-sample annualized Sharpe ratios and MPPMs for the last year of each subperiod based on the 10%-FDR$^+$ portfolios of significant rules, under the *dynamic hedging* approach. The rules are those selected in the in-sample horizon, which covers 70% of each subperiod's observations. We impose historical transaction costs on the computations.

Table 5. Out-of-sample performance of technical trading portfolios based on the annualized Sharpe ratio, CAGR and MPPM.

	1996			2001			2007			2011			2016		
FDR method	SR[36]	CAGR %	MPPM %	SR	CAGR %	MPPM %	SR	CAGR %	MPPM %	SR	CAGR %	MPPM %	SR	CAGR %	MPPM %
Commodities *port*	**0.98**	**2.82**	**2.75**	**1.82**	**5.58**	**5.39**	**1.46**	**2.71**	**2.66**	**3.01**	**4.51**	**4.39**	**3.24**	**4.75**	**4.63**
Equities *port*	-1.91	-0.42	-0.42	-0.40	-0.21	-0.21	**2.62**	**2.33**	**2.32**	**0.33**	**0.35**	**0.34**	**2.34**	**2.08**	**2.06**
Currencies *port*	-1.31	-0.31	-0.32	**0.77**	**0.21**	**0.21**	**0.65**	**0.15**	**0.15**	**0.91**	**0.72**	**0.72**	-2.65	-0.41	-0.41
Global *port*	**0.43**	**0.41**	**0.41**	**1.78**	**1.83**	**1.82**	**2.55**	**1.73**	**1.71**	**2.59**	**1.84**	**1.82**	**3.67**	**2.12**	**2.11**
DHR FDR															
Commodities *port*	**1.91**	**4.53**	**4.41**	**1.65**	**3.53**	**3.45**	**1.21**	**1.74**	**1.72**	**1.78**	**3.87**	**3.77**	**2.96**	**4.01**	**3.94**
Equities *port*	-1.66	-0.95	-0.96	**0.88**	**0.57**	**0.57**	**1.31**	**0.39**	**0.39**	-0.35	-0.31	-0.32	**2.05**	**1.36**	**1.35**
Currencies *port*	-1.98	-0.28	-0.28	**0.55**	**0.18**	**0.18**	-0.22	-0.03	-0.03	**2.21**	**1.03**	**1.01**	**1.49**	**0.27**	**0.27**
Global *port*	**1.23**	**1.07**	**1.06**	**1.86**	**1.42**	**1.41**	**1.41**	**0.71**	**0.69**	**1.59**	**1.45**	**1.44**	**3.23**	**1.88**	**1.86**

We report the out-of-sample annualized Sharpe ratio, CAGR, and MPPM for the last year of each subperiod and for market portfolios comprising the corresponding FDR significant rules, the FDR portfolios constructed under the *1:1* spread formation, and the *dynamic hedging* approach, for each pair, selected from the in-sample horizon that covers 70% of each subperiod's observations. We impose historical transaction costs in the computations.

[36] For portfolios' returns yielding positive Sharpe ratios, we also checked their skewness, and we report that for all those cases, the skewness derived was positive rather than negative, and thus we are confident that the Sharpe ratios are not the result of manipulation (see also Goetzmann et al. 2002). The results are available on request.

Table 6. Factor exposures, liquidity, volatility and investor's sentiment of the FDR+ global portfolio of significant rules.

Panel A	**Carhart (1997) 4-Factor Models**							
		MSCI World	SMB	HML	UMD	Intercept	R^2	Adj. R^2
Global port.	Coefficient	-0.003***	0.0004	-0.003***	-0.0018***	0.04%***	7.92%	6.73%
	(t-Stat)	(-4.21)	(0.39)	(-2.88)	(-2.55)	(13.46)		

Panel B	**Asness, Moskowitz, and Pedersen (2013) factors**							
		MSCI World	VAL Everywhere	MOM Everywhere		Intercept	R^2	Adj. R^2
Global port.	Coefficient	-0.003***	-0.007***	-0.006***		0.05%***	8.91%	8.01%
	(t-Stat)	(-3.32)	(-2.61)	(-3.11)		(12.76)		

Panel C	**Market volatility, liquidity and sentiment**					
		TED spread	Pastor-Stambaugh (2003)	Sentiment	Change in Sentiment	lnVIX (monthly)
Global port.	Coefficient	-0.046***				
	(t-Stat)	(-3.35)	-0.002***			
			(-4.30)			
				-0.0006		
				(-1.51)	-0.0004	
					(-0.98)	0.001***
						(5.84)

We run monthly time series regressions of the returns of the global portfolio on the factors of SMB, HML and UMD of Carhart (1997) model and the Asness et al. (2013) value and momentum *everywhere* factors over the 1990-2016 period. We also run similar regressions of global portfolio's returns, on the funding liquidity (TED spread, market liquidity (Pastor-Stambaugh, 2003), the investor's sentiment index of Baker and Wurgler (2006, 2007) and the VIX index separately.

Figures

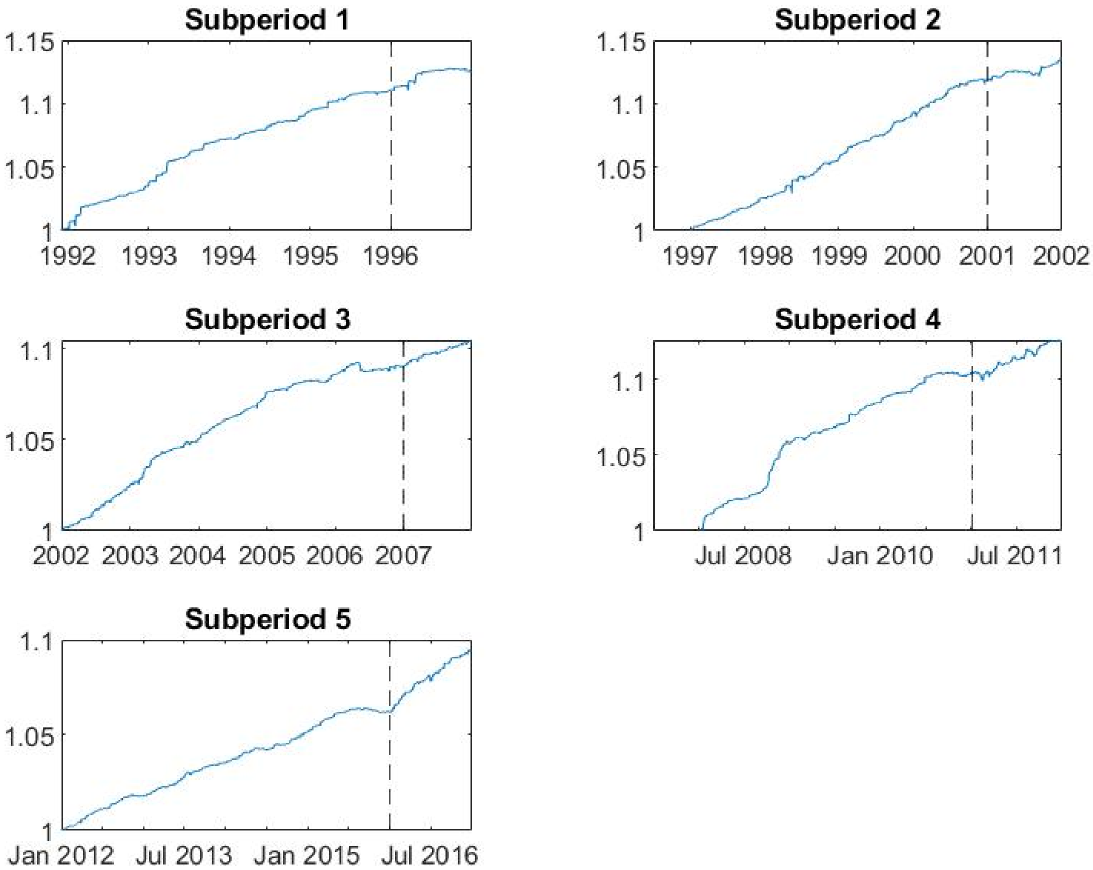

Fig. 1. Compounded cumulative excess return of the global portfolio of outperforming rules based on the Sharpe ratio metric for the IS and OOS horizons in each subperiod tested. The OOS horizon starts from the dashed vertical line. We assume $1 invested on January 1, 1991, in an equally weighted portfolio of all the in-sample outperforming rules for each pair, while also testing the portfolio's performance out of sample. The best-performing technical rules were selected using the FDR method during every in-sample period. We also include historical transaction costs.

Online Appendix

Technical Analysis, Spread Trading and Data Snooping Control

A. Details of technical trading rule parameters

In this section, we describe in precise detail the entire universe of our technical trading rules (TTRs), following the previous studies of Sullivan Timmermann and White (STW) (1999) and Hsu, Taylor, and Wang (2016).

A.1. Filter rules

The filter rule allows the initiation of a spread's trader position only in response to major price trends. Therefore, an investor *buys* a spread if its price *increases* by a fixed percentage from a previous *low*, and he *sells* if the price *decreases* by a fixed percentage from a previous *high*. We assume three different filter rule variations as described below, and we set the previous low (high) as the most recent spread value between the daily closing prices of two assets that is less (greater) than the *n* previous daily spread values, for a given value of *n.*

F1: *If the daily spread value between the daily closing spot prices of two assets increases (decreases) by at least* x% *from its previous low (high) and remains so for* d *days, then go long (short) the spread.*

The second variant also allows for neutral positions in lieu of always being either long or short according to the basic filter rule:

F2: *If the daily spread value increases by at least* x% *from its previous low and remains so for* d(x) *days, then go long the spread until its daily value decreases by at least* y% *from its* *subsequent high and remains so for* d(y) *days, at which time liquidate the long position. If the daily spread value decreases by at least* x% *from its previous high and remains so for* d(x) *days, then go short the spread until its daily value increases by at least* y *percent from its subsequent low and remains so for* d(y) *days, at which time liquidate the short position.*

The third variation assumes a position is held for a fixed number of periods, ignoring all other signals:

F3: *If the daily spread value increases (decreases) by at least* x% *from its previous low (high) and remains so for* d *days*, *then go long (short) the spread for* c *days and neutralize the position.*

$n = 1, 2, 5, 10, 15, 20$ [6 values]

$x = 0.05, 0.1, 0.5, 1.0, 5.0, 10.0, 20.0$ as a percentage [7 values]

$y = 0.05, 0.1, 0.5, 1.0, 5.0, 10.0, 20.0$ as a percentage [7 values]. Twenty-one (x, y) combinations are possible when $y < x$.

$d = 0, 1, 2, 5$ [4 values]. Six combinations are possible when $(d(x), d(y))$.

$c = 1, 5, 10, 15, 20, 25$ [6 values]

The total number of rules is $(x * d * n) + (x, y) combinations * (d(x), d(y))\ combinations * n + (x * d * c * n) = 168 + 756 + 1{,}008 = 1{,}932$.

A.2. Moving averages

Moving averages (MA) are the most popular and actively traded classes of TTRs. To initiate a trade a crossover between the spread value and a moving average of a given length, or between a short and a long moving average of different lengths, should happen. These upside (downside) penetrations of a moving average help an investor to discover the emergence of new trends and maintain his position as long as the crossover remains. We provide four different variations as follows:

MA1: *If the daily spread value moves at least* x% *above (below) the moving average, that is,* MA(n), *and remains so for* d *days, then go long (short) the spread until its closing value moves at least* x% *below (above)* MA(n) *and remains so for* d *days, at which time go short (long) the spread.*

Similar to the *filter rules*, a variant of a *moving average rule*, assuming we hold a position for a fixed number of periods ignoring all other signals, would be as described below:

MA2: *If the daily spread value moves at least* x% *above (below)* MA(n), *and remains so for* d *days, then go long (short) the spread for* c *days and neutralize the position.*

The next two paragraphs describe a short/long double-moving average, as well as a similar one including a fixed holding period:

MA3: *If the short moving average, that is,* MA(m), *moves at least* x% *above (below) the long moving average, that is,* MA(n), *and remains so for* d *days, then go long (short) the spread until* MA(m) *moves at least* x% *below (above)* MA(n) *and remains so for* d *days, at which time go short (long) the spread.*

MA4: *If* MA(m) *moves at least* x% *above (below)* MA(n) *and remains so for* d *days, then go long (short) the spread for* c *days and neutralize the position.*

$n = 2, 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [11 values]

$m = 2, 5, 10, 15, 20, 25, 50, 100, 150, 200$ [10 values]. Fifty-five combinations are possible when $m < n$.

$x = 0, 0.05, 0.1, 0.5, 1.0, 5.0$ as a percentage [6 values].

$d = 0, 2, 3, 4, 5$ [5 values].

$c = 5, 10, 25$ [3 values]

The total number of rules is $(n * x * d) + (n * x * d * c) + \big((m, n)\ combinations * x * d\big) + \big((m, n)\ combinations * x * d * c\big) = 330 + 990 + 1{,}650 + 4{,}950 = 7{,}920$.

A.3. Support and resistance

Like filter TTRs, support and resistance rules try to discover major price movements beyond certain levels that are difficult to breach, rather than a more recent high or low. The intuition behind this rule is that, usually, investors think that sooner or later the movement in the spread's price will tend to stop and the price will return to a certain level. However, if the price breaks through a certain resistance or support level by a certain amount, it is more likely to continue moving in the same direction until it finds a new level. In this way, a long or a short signal is generated depending on the rule's construction. Again, we predefine the support and resistance levels as the minimum and maximum closing values of a spread over the previous *n* closing values, respectively.

SR1: *If the daily spread value rises above (below) by at least* x% *the local maximum (minimum) over the* n *previous spread values, and remains so for* d *days, then go long (short) the spread.*

In addition to the above, we impose a holding-period filter:

SR2: *If the daily spread value rises above (below) by at least* x% *the local maximum (minimum) over the* n *previous spread values, and remains so for* d *days, then go long (short) the spread for* c *days and neutralize the position.*

$n = 2, 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [11 values].

$x = 0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10.0$ as a percentage [7 values].

$d = 0, 1, 2, 3, 4, 5$ [6 values].

$c = 1, 5, 10, 25$ [4 values].

The total number of rules is $(n * x * d) + (n * x * d * c) = 462 + 1{,}848 = 2{,}310$.

A.4. Channel breakouts

Similar to the case of the time-varying support and resistance rule, a trading channel occurs when the highest value of a spread over a prespecified number of previous days is within a fixed percentage (*b%*) of the lowest value over those days. The graphical representation of a price channel is equal to a spread of parallel trend lines drifting together within a certain width. As soon as one of these trend lines is *broken*, a buy or a sell signal is generated. Thus, an investor goes long (short) when the price moves above (below) the channel. The above time-varying support and resistance levels represent the lower and upper bounds of the channel, whereas the difference between those bounds and the high and low, respectively, over the previous prespecified days, does not exceed *b%*.

CB1: *If a* b% *trading channel occurs and if the daily spread value rises above (below) by at least* x% *the upper (lower) bound over the previous* n *days and remains so for* d *days, then go long (short) the spread.*

We also consider a holding period for each position triggered:

CB2: *If a* b% *trading channel occurs and if the daily spread value rises above (below) by at least* x% *the upper (lower) bound over the previous* n *days and remains so for* d *days, then go long (short) the spread for* c *days and neutralize the position.*

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [10 values].

$b = 0.1, 0.5, 1.0, 5.0, 10.0$ as a percentage [5 values].

$x = 0.05, 0.1, 0.5, 1.0, 5.0$ as a percentage [5 values]. Fifteen (x, b) combinations are possible when $x < b$.

$d = 0, 1, 2$ [3 values].

$c = 1, 5, 10, 25$ [4 values].

The total number of rules is $(n * (x, b)\ combinations * d) + (n * (x, b)\ combinations * d * c) = 450 + 1{,}800 = 2{,}250$.

A.5. Relative strength indicator rules

Introduced by Levy (1967), relative strength indictor (RSI) rules belong to the general family of *overbought/oversold* indicators, also commonly called *oscillators*, from which we also pool the rest of our proposed *reversal* TTRs (i.e., Bollinger bands, CCIs). As already mentioned, RSIs attempt to reveal upcoming price corrections in the opposite direction to extreme upward or downward movements, in which a short or a long signal is executed accordingly. The generic formula of an RSI is

$$RSI_t(n) = 100 - \frac{100}{1+\frac{U_t(n)}{D_t(n)}} = 100\left[\frac{U_t(n)}{U_t(n)+D_t(n)}\right],$$

where $U_t(n)$ and $D_t(n)$ represent the cumulated upward and downward trends, calculated as the sum of the first differences between monotonically increasing or decreasing closing prices in absolute terms over the previous *n* days. In other words, $U_t(n)$ denotes the total gains from a potential upward movement, and $D_t(n)$ denotes the total losses of a potential downward movement over the previous *n* days. Thus, normalized to the scale of 100, the RSI estimates the dominance of an upward relative to the dominance of a downward trend. In its simplest version, an RSI of value 70 characterizes a specific spread as overbought, whereas a value of 30 rates the spread as oversold.[37] As well as this naïve RSI variant, we also consider two more modifications introduced by Hsu et al. (2016):

RSI1: *If* $\mathrm{RSI_t(n)}$ *rises above 70, then go short the spread. Alternatively, if* $\mathrm{RSI_t(n)}$ *falls below 30, then long the spread.*

RSI2: *If* $\mathrm{RSI_t(n)}$ *rises above* $50 + \mathrm{k}$ *for at least* d *days and then falls below* $50 + \mathrm{k}$, *go short the spread. Alternatively, if* $\mathrm{RSI_t(n)}$ *falls below* $50 - \mathrm{k}$ *for at least* d *days and then rises above* $50 - \mathrm{k}$, *go long the spread.*

RSI3: *If* $\mathrm{RSI_t(n)}$ *rises above* $50 + k$ *for at least d days and then falls below* $50 + \mathrm{k}$, *go short the spread for* c *days and neutralize the position. Alternatively, if* $\mathrm{RSI_t(n)}$ *falls below* $50 - \mathrm{k}$ *for at*

[37] This is the most naïve RSI rule, and it can be found on several trading websites and platforms.

least d *days and then rises above* $50 - \mathrm{k}$, *go long the spread for c days and neutralize the position*

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [10 values].

$k = 10, 15, 20, 25$ [4 values].

$d = 1, 2, 5$ [3 values].

$c = 1, 5, 10, 20, 25$ [5 values].

The total number of rules is $n + (n * k * d) + (n * k * d * c) = 10 + 120 + 600 = 730$.

A.6. Bollinger bands

Developed by the famous technical trader John Bollinger in the 1980s, Bollinger bands are volatility indicators aimed at taking advantage of unjustifiably high or low prices and their imminent corrections. To achieve this, they consider upper and lower bands of a spread's price in terms of standard deviations away from a moving average over a prespecified number of previous days. Considering a spread's moving average and its moving standard deviation over the *n* previous days, we have

$$MA_t(n) = \frac{1}{n}\sum_{i=1}^{n} P_{t-1+i}\,,\ \sigma_t(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_{t-1+i} - P_{t-1}}{P_{t-1}}\right)^2}.$$

We define the upper and lower bands of a given width *z* as $MA_t(n) \pm z * \sigma_t(n)$, which are the specific moving average plus or minus *z* times the specific moving standard deviation.

Almost always, the price of a spread trades between the two bands, except in cases in which extreme conditions occur. Thus, any breakout above or below the bands is a major event. Many investors believe the closer the prices move to the upper band, the more overbought the market and vice versa. This can lead to a pullback of prices captured by a *reversal* TTR. We consider rules based on prices leaving the bands triggering a position and possibly the subsequent crossing of the moving average neutralizing the position.

BB1: *If the daily closing spot price (spread) moves above the upper band of a given width* z *and remains so for* d *days, go short the spread until it returns to the moving average, at which time neutralize the short position. If the daily spread value moves below the lower band of a given width* z *and remains so for* d *days, go long the spread until it returns to the moving average, at which time neutralize the position.*

BB2: *If the daily spread value moves above the upper band of a given width* z *and remains so for* d *days, go short the spread for c days and then neutralize the position. If the daily spread value moves below the lower band of a given width* z *and remains so for* d *days, go long the spread for* c *days and then neutralize the position.*

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [10 values].

$z = 0.5, 1, 1.5, 2, 2.5, 3$ [6 values].

$d = 0, 1, 2, 3, 4, 5$ [6 values].

$c = 1, 5, 10, 20, 25$ [5 values].

The total number of rules is $(n * z * d) + (n * z * d * c) = 360 + 1{,}800 = 2{,}160$.

A.7. Commodity channel index rule

Introduced by Donald Lambert in 1980, the commodity channel index (CCI) also belongs to the family of oscillators, attempting to capture cyclical trends and so determine *overbought/oversold* levels. The CCI was initially developed to discover such levels in the commodities market, but its prominent applicability soon encouraged technical traders to use it in the equities and currencies markets as well. Like Bollinger bands, the CCI not only uses extreme upper and lower bands to trigger long/short signals but also takes into account the volatility of a spread. The CCI is defined as

$$CCI_t(n) = \frac{P_t - MA_t(n)}{0.015 * \sigma_t(n)},$$

where P_t is the price of a spread at a specific time *t*, and $MA_t(n)$ and $\sigma_t(n)$ denote the spread's moving average and standard deviation over the previous *n* days, calculated like in the case of the Bollinger bands. Thus, the CCI measures the current price level relative to an average price level over a specific period of time and is fairly high when prices are far above the moving average and vice versa. The constant 0.015 just ensures that the majority of CCI values will lie between -100 and +100, which represent the upper and lower bounds of this TTR.

As a *reversal* indicator, the CCI searches overbought (i.e., greater than +100) or oversold (i.e. less than -100) conditions to foretell a mean reversion. Similarly, bullish and bearish divergences can be used to detect early momentum shifts and anticipate trend reversals. We employ two simple *reversal* variants of the CCI as well as a CCI that detects bullish/bearish divergence breakouts.

CCI1: *If* $\mathrm{CCI_t(n)}$ *remains above* $(+100 + \mathrm{k})$ *for at least* d *days and subsequently moves below* $+100$, *go short the spread. If* $\mathrm{CCI_t(n)}$ *remains below* $(-100 - \mathrm{k})$ *for at least* d *days and subsequently moves above* -100, *go long the spread.*

Assuming a holding period *c*, we have

CCI2: *If* $\mathrm{CCI_t(n)}$ *remains above* $(+100+\mathrm{k})$ *for at least* d *days and subsequently moves below* $+100$, *go short the spread for* c *days and then neutralize the position. If* $\mathrm{CCI_t(n)}$ *remains below* $(-100-\mathrm{k})$ *for at least* d *days and subsequently moves above* -100, *go long the spread for* c *days and then neutralize the position.*

We finally consider a special case of a CCI and divergence breakout. Divergences can foresee a potential trend reversal point as they usually reflect a change in momentum. We examine two types of divergence, bullish and bearish. A bullish divergence appears when the spread performs a lower low (i.e., support break) and the CCI shapes a higher low, over the previous *n* days, indicating less downside momentum. A bearish divergence appears when the spread performs a higher high (i.e., resistance break) and the CCI forms a lower high over the previous *n* days, indicating less upside momentum. In other words, we are looking for a breach in the support and resistance levels of a spread's price, while searching for a direction change in the CCI. In particular, we have:

CC3: *If the daily spread value moves below by at least* x% *the local minimum over the* n *previous spread values and the CCI remains below* $(-100-\mathrm{k})$, *while its local minimum over the same n days moves above its previous value, then go long the spread. If the daily spread value rises above by at least* x% *the local maximum over the* n *previous spread values, and the CCI remains above* $(+100+\mathrm{k})$, *while its local maximum over the same* n *days moves below its previous value, then go short the spread.*

$n = 5, 10, 15, 20, 25, 50, 100, 150, 200, 250$ [10 values].

$k = 0, 50, 100$ [3 values].

$d = 1, 2, 3, 4, 5$ [5 values].

$c = 1, 5, 10, 20, 25$ [5 values].

$x = 0.05, 0.1, 0.5, 1.0, 2.5, 5.0, 10.0$ as a percentage [7 values].

The total number of rules is $(n * k * d) + (n * k * d * c) + (n * x * k) = 150 + 750 + 210 = 1{,}110$.

B. Robustness checks, break-even transaction costs and subperiod analysis

B.1. Robustness checks

B.1.1. Break-even transaction costs

So far, we have reported the results of TTR performance, while handling transaction costs (brokerage fees and forward rate bid-ask spreads) as *endogenous* to the selection process. Moving one step forward and following previous studies (Bessembinder and Chan, 1998; Neely and Weller, 2013; Bajgrowicz and Scaillet, 2012 (BS)), we identify the break-even transaction costs of the most significant TTR for each corresponding spread. Typically, one-way break-even transaction costs represent the level of transaction costs that would neutralize the outperformance generated by a TTR and so minimize the outperformance exactly to zero. Thus, we compute the break-even transaction costs by increasing them to such a level that the most predictive TTR cannot generate positive performance under the FDR framework. This approach tackles the exogeneity problem of transaction costs described by BS, by deriving *ex ante* break-even transaction costs computed endogenously. The overall procedure helps us to identify the robustness of the results generated in the previous section, by comparing the break-even transaction costs to the actual transaction costs we consider. The greater the difference, the more robust the performance of a TTR may be deemed to be.

Table B.1 displays the break-even and the actual one-way transaction costs employed in our study (in basis points) for the full sample. We also report the number of trades for the most significant TTRs. The first column corresponds to the actual transaction costs used for each spread, that is, the brokerage fees for the commodity and equity spreads and the mean of the estimated forward rate bid-ask spread for the currency spreads. The next two columns relate to the one-way break-even transaction costs, as described above, and the number of trades triggered for the best significant TTR, selected under the Sharpe ratio criterion, over the full sample period.

[Insert Table B.1 around here]

Break-even transaction costs surpass, by far, the actual transaction costs in most cases because the number of trades they produce is quite small compared with the full sample of trading days. For instance, the break-even transaction costs of the best-performing rules selected for commodity spreads range from 25 (platinum-gold) to 77 (Brent-WTI crude oil and heatoil-gasoil). The equity indices' break-even transaction costs exceed, on average, by far (more than double) even conservative (i.e., high) historical estimates of actual transaction. A similar pictured is revealed by looking the break-even costs of TTRs performance traded on the currency spreads. Again, the computed break-even costs exceed the actual ones at least by 10 basis points. It is also worthy to mention that the equity spreads' rules achieve the highest break-even transaction costs, on average, followed by those for the commodity and currency spreads.

In terms of the number of trades, the TTRs of each of the three markets examined are split into three different groups. The equity spreads' significant rules tend to trade on a higher frequency, whereas the commodity and currency spreads' rules trade on medium and low frequencies, respectively. However, given that we build our analysis on daily data covering a 25-year period (i.e., January 1990 to December 2016), which corresponds to a total of 7,045 trading days (except in the case of the heating oil –gas oil spread), the overall trading activity shows that the genuine

TTRs are quite prudent overall. Only for the cases of S&P500-DAX and CAX-TOPIX, we realize a couple of thousands of trades. This picture stems from treating the transaction costs as endogenous to the selection process, such that rules generating fewer signals are chosen. In other words, considerable transaction costs can offset the performance of TTRs that trigger more frequent signals. In addition, the nature of the Sharpe ratio in capturing the average excess return per unit of total risk is probably the reason for the chosen rules triggering fewer trades as most significant, because the aim is to minimize the total risk of the investment.

In overall, it is noteworthy that technical predictability can be transformed into outperformance given a fair level of transaction costs in spreads trading, at least in a *backtesting* framework of genuinely selected rules under the FDR^{+} method. This robustness check also highlights that transaction costs do not necessarily eradicate the possibility of yielding significant outperformance in spreads trading using technical analysis.

B.1.2. Subperiod analysis

In this subsection, we focus on shorter periods of time to assess the time-varying predictability of TTRs on our corresponding spreads as they were defined in Section 6 of the main body of the paper. We treat each subperiod as whole here, and so we do not split them between IS and OOS horizons. Previous empirical studies on the performance of technical analysis reveal a considerable decay in the outperformance and the evolution of predictability over recent years, a sign of informational efficiency improvements across investors. For example, STW and BS provide relevant evidence from equities markets, whereas Menkhoff and Taylor (2007) and Neely, Weller, and Ulrich (2009) demonstrate the lower performance of technical analysis over time in the foreign exchange market. Thus, the 25-year historical dataset allows us to revisit the evolution of the predictability of TTRs, but this time on spreads traded in different markets.

Fig. B.1 is a scatterplot of the families of TTRs against the total number or rules selected across all spreads and subperiods, as identified using the FDR^+ method. Specifically, it provides information on the decomposition of each of the 10%-FDR^+ portfolios of predictive rules formed during the five subperiods, for each spread, into their corresponding families.

[Insert Figure 1 around here]

Before we describe the findings we must mention that the total number of significant rules detected across all spreads and periods exceeds the 10,000. At a first glance we notice a specific pattern in the selection of significant TTRs across almost every spread and subperiod. Interestingly, this pattern involves a considerable concentration on momentum rules, especially support and resistance and channel breakout rules except for contrarian ones, which seem to dominate in numbers. For instance, support and resistance and channel breakout rules appear to be more predictable from the rest trend-following classes of TTRs. However, this doesn’t mean that the most profitable rules in OOS are those belonging to the trend-following classes, because we only assess here the time-varying predictability of the selected rules. The OOS profitability is thoroughly examined in Section 6 of the main body of the paper. Also, the trend-following rules examined constitute a large proportion of the total universe of rules examined (i.e., 14,412 out of 18,412) compared to the contrarian rules. So, it is expected that a considerable number of them will demonstrate a good predicting ability. In the meantime, Bollinger bands and CCIs seem to outperform the RSIs with respect to the contrarian families, while being of the most predictable rules among the whole universe of examined TTRs. Another important finding provided by Fig. B.1, is the very sporadic appearance of moving averages among the significant outperforming rules, across all spreads and subperiods.

Table B.2 presents the numbers of significant rules, in terms of predictability and after minimizing the data-snooping bias, for every spread separately. We carry out the FDR^{+} procedure under the Sharpe ratio criterion in every subperiod and for the TTR population.

[Insert Table B.2 around here]

In the first panel of the table, we notice that technical analysis predicts the commodity spreads, across almost all subperiods. However, the number of significant rules differs from period to period. The Brent-WTI crude oil case highlights the decay of rules predictability in more recent decades, but this is not always the case for the rest of the commodity spreads. The overall picture indicates that technical predictability is equally strong during the more recent subperiods and for specific spreads. For instance, the platinum-gold and corn-soybean spreads are more predictable as we move toward the last year of our dataset, than in earlier years. In the case of platinum-palladium, we observe a more stable performance of TTRs over all the examined subperiods. The picture seems even more diverse for the equities spreads in the second panel. For all the spreads considered, there are times of considerable predictability, when the number of significant rules is larger, and times of weak performance, but no specific pattern is revealed. Only in the DAX-FTSE100 spread can we detect a downward pattern in technical analysis predictability as we move toward the last two subperiods. Thus, previous evidence, such as that described by Brock, Lakonishok, and LeBraron (1992), STW and Bajgrowicz and Scaillet, (2012), is not consistent with that for spreads trading on equities markets. Finally, we observe a very similar picture in the exchange rate case. They display both upward and downward trends in terms of predictability from period to period, with the EUR-JPY consistently exhibiting the biggest number of predictable rules across all subperiods. Overall, the commodity spreads appear to be more predictable using the TTRs than do both the equity and the exchange rate spreads, across all subperiods. The second most predictable spreads are those of the equity market.

The most important finding of Table B.2 lies in the fact that the predictability of technical analysis for spreads trading has different characteristics than that for single assets. We show that, mainly, there is no uniformly monotonic downward trend in the performance of TTRs for the specific spreads. As we mentioned above, this is opposite of the findings of other relevant studies focusing on single commodities, equity indices, and currencies markets. The evidence above generally supports temporary market anomalies. Nevertheless, because model predictability is not always synonymous with excess profitability, we investigate the validity of such interim market efficiencies in our OOS analysis section. Finally, the subperiod analysis also emphasizes the greater power and flexibility of the FDR^+ method in selecting significant rules, even in short periods of time, compared with previous approaches.

References

Allen, F. & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51, 245–271. https://doi.org/10.1016/S0304-405X(98)00052-X

Bajgrowicz, P. & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106, 473–491. https://doi.org/10.1016/j.jfineco.2012.06.001

Bessembinder, H. & Chan, K. (1998). Market efficiency and the returns to technical analysis. Financial Management, 27, 5–17. DOI: 10.2307/3666289

Brock, W., Lakonishok, J. & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47, 1731–1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

Hsu, P.H., Taylor, M.P. & Wang, Z. (2016). Technical trading: Is it still beating the foreign exchange market? Journal of International Economics, 102, 188-208. https://doi.org/10.1016/j.jinteco.2016.03.012

Levy, R. A. (1967). Relative Strength as a criterion for investment selection. Journal of Finance, 22, 595-610. https://doi.org/10.1111/j.1540-6261.1967.tb00295.x

Menkhoff, L. & Taylor, M.P. (2007). The obstinate passion of foreign exchange professionals: Technical analysis. Journal of Economic Literature, 45, 936–972. DOI: 10.1257/jel.45.4.936

Neely, C.J. & Weller, P.A. (2003). Intraday technical trading in the foreign exchange market. Journal of International Money & Finance, 22(2), 223–237. https://doi.org/10.1016/S0261-5606(02)00101-8

Neely, C.J., Weller, P.A. & Ulrich, M. (2009). The adaptive markets hypothesis: Evidence from the foreign exchange market. Journal of Financial and Quantitative Analysis, 44, 467–488. https://doi.org/10.1017/S0022109009090103

Sullivan, R., Timmermann, A. & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance, 54, 1647–1691. https://doi.org/10.1111/0022-1082.00163

Table B.1. Break-even transaction costs for predictive TTRs for the full 25-year sample period

	Cost (bps)	**Break-even cost (bps)**	**# trades**
Commodities			
Brent-WTI crude oil	6.6	77	434
Heatoil-Gasoil	6.6	77	28
Platinum-gold	6.6	25	291
Gold-Silver	6.6	66	38
Corn-Soybean	6.6	26	705
Equities			
FTSE100-CAC 40	20	121	3
AMST-CAC 40	20	46	2,183
S&P500-DAX	20	54	25
CAC-TOPIX	20	56	1,245
DAX-FTSE100	20	110	7
Currencies			
EUR-CHF	6.1	69	5
EUR-JPY	3.4	14	3
AUD-CAD	8	28	6
AUD-NZD	12.2	66	8
NOK-SEK	14	29	44

We report the highest one-way break-even transaction costs (in basis points) that will reduce the performance metrics of the most predictive rules to zero. The mean excess return and Sharpe ratio are the performance metrics. *#trades* denotes the number of trades triggered by each TTR over the sample period. - denotes that, for a given spread and performance metric, no significantly profitable TTR exists.

Table B.2. Number of technical rules with significantly positive Sharpe ratios in five subperiods

Subperiod	1991–1996	1997–2001	2002–2007	2008–2011	2012–2016
Commodities					
Brent-WTI crude oil	793	86	104	14	11
Heatoil-Gasoil	-	1248	826	575	298
Platinum-gold	6	9	42	51	15
Gold-Silver	129	59	232	77	11
Corn-Soybean	7	51	26	63	33
Equities					
FTSE100-CAC 40	12	22	12	7	18
AMST-CAC 40	37	19	87	10	5
S&P500-DAX	28	6	1	22	30
CAC-TOPIX	44	30	51	113	96
DAX-FTSE100	103	27	4	2	-
Currencies					
EUR-CHF	22	3	2	15	1
EUR-JPY	26	3	17	7	18
AUD-CAD	6	1	8	28	0
AUD-NZD	5	23	9	8	9
NOK-SEK	14	10	2	3	12

This table presents the number of TTRs, of a total of 18,142, that provide significantly positive Sharpe ratios based on the FDR$^+$ test over the total sample's five subperiods: 1991–1996, 1997–2001, 2002–2007, 2008–2011, and 2012–2016. We design the subperiods based on historical events including the Maastricht Treaty in 1992, the East Asian currency crisis in 1997, the *dot-com* bubble in 1999–2000 and subsequent 2002 credit crunch, the appearance of the euro in 2002 and the 2003–2007 energy crisis, the global financial crisis of 2008, and, finally, the recent crude oil downturn in 2014. Historical transaction costs are imposed on the generated returns.

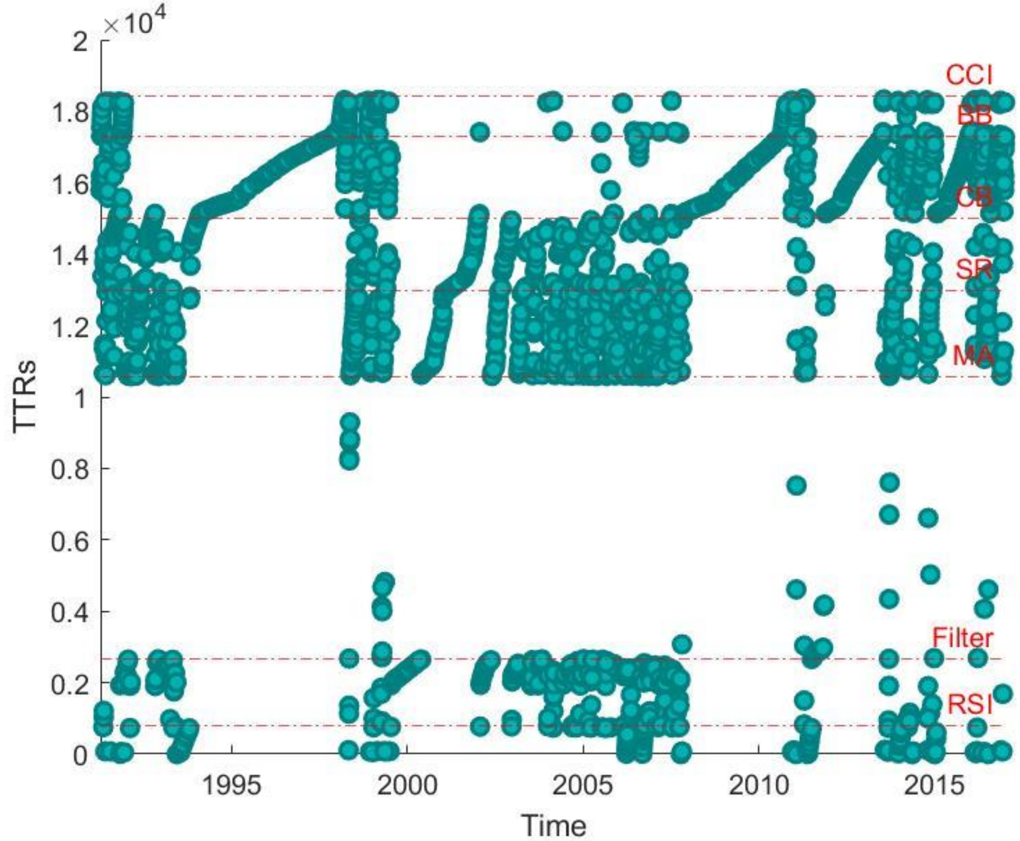

Fig. B.1. FDR portfolio decomposition for each pair and across all subperiods. The horizonal lines split up the different families of TTRs, which add to 18,412 TTRs in total. We display the categories of TTRs in the following order: RSIs, filter rules, moving averages, support and resistance rules, channel breakouts, Bollinger bands, and CCIs.