Technical Analysis, Spread Trading, and Data Snooping Control

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

Version: 10th September 2021

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

---

[a]**Corresponding author:* School of Economics & Finance, University of St Andrews, The Scores, St Andrews, KY169AR U.K., *Email address:* **ioannis.psaradellis@st-andrews.ac.uk ,** *Tel.*: +44 (0)1334 462427

[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 is believed to be one of the longest-established forms of investment analysis, which explores future trading opportunities for financial assets by analyzing the asset prices' time-series history. Numerous studies have studied the profitability of technical analysis across several financial markets. Half of them plead in favor of technical analysis profitability (see Neely et al., 1997; Sullivan, Timmermann, and White, 1999 (STW); Lo et al., 2000; Hsu, Taylor and Wang, 2016 (HTW)), whereas the rest argue against it at least in the recent past (see, Allen and Karjalainen, 1999; Marshall et al., 2008a, 2008b; Bajgrowicz and Scaillet, 2012 (BS); Shynkevich, 2012a, 2012c, 2012b; Kuang, Schroder and Wang, 2014, Shynkevich, 2016). However, out of the studies that reveal technical analysis' profitability, only a few control for data snooping bias via a statistical inference specification (see STW and HTW).

Nevertheless, technical analysis remains extremely popular among practitioners; see, for example, the surveys and studies of Bogomolov (2013) for the equities market, Gehrig and Menkhoff (2004) for the forex market and Levine and Pedersen (2016) for the commodities market. For example, Levine and Pedersen (2016) support that trend-following strategies such as time-series momentum, moving averages, and filters are some of the principal investment styles followed by hedge funds and commodity trading advisers (CTAs). So far, studies using advanced statistical inference tools did not investigate realistic technical trading rules (TTRs) widely used by practitioners, which might be a possible reason for the observed *argument* on technical analysis' profitability. Academics support that implicitly, all TTRs are not profitable once controlling for luck, even when they test specific rules in their studies. However, those rules might not be the same as or a subset of the ones used extensively by the practitioners.

Our study aim is to contribute to this debate and offer a convincing answer. We assess the predictability and OOS profitability of an *up-to-date* and *broader* universe of more than 18,000 disciplined TTRs on spread trading, combined with the *false discovery rate control* ($FDR^{+/-}$) of Barras, Scaillet, and Wermers (BSW) (2010, 2019) to control for data-snooping effects. For this purpose, we consider frequently traded spreads on several markets from January 1990 to December 2016. We report that technical analysis still displays significant predictability on spread trading across different markets while controlling for data snooping bias. Our findings are consistent with statistical arbitragers' and active traders' reports that successfully employ technical trading to capture significant trends and reversals of the market price.

Spread trading is a relative-value arbitrage strategy with a 35-year plus history. Over that time, the strategy has remained popular among fund managers as it constitutes one of three broad types of strategies used by hedge funds (see Pedersen, 2015). Such a strategy can generate excess returns by going long one asset while going short another to yield profits because of a short-term deviation of their relative valuations from a perceived equilibrium. [1] 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), but this is the first time an academic study coherently focuses on assessing the predictive ability of technical analysis on spread trading of different assets.

Data snooping becomes a considerable concern when recruiting an extensive dataset whose number of variables is larger than the number of observations. This issue results in false discoveries, especially when classical statistical inference is used in which investors repeatedly test the same single hypothesis without adapting a rejection region. Contrary to its recent computationally

[1] Popular statistical arbitrage studies also include those of Gatev et al. (2006), Kondor (2009), and Chen et al. (2017).

intensive counterparts, the $FDR^{+/-}$ is a simple approach, while on the other hand, it is efficient in terms of yielding only a minimal bias when estimating the proportion of rules with zero performance, and it identifies models with true outperformance even if the best model's performance results from pure luck (BSW, 2019). We also use the *manipulation-proof performance measure* (MPPM) of Ingersoll et al. (2007) to validate that the OOS performance of significant TTRs is not the result of manipulated time-series properties but true outperformance. As far as we know, this is the first time such an approach is employed in spread trading.

In addition, to evaluating the predictability of TTRs over the full sample, we also separate our dataset into five subperiods, based on major historical events, using each subperiod's last year as the OOS period. 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., Heat oil-Gasoil) 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. Such a finding contrasts with the literature (BS; HTW), stating that the predictability of technical analysis has shrunk over time because of informational efficiency. Their corresponding outstanding MPPMs reveal that such a performance is probably robust rather than a result of manipulation due to unpriced risk. On the contrary, most of the equity indices' spreads seem to underperform most of the time except from one case (i.e., CAC-TOPIX).

Additionally, 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 Carhart's (1997) four-factor model and Asness et al.'s (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. A *global* market portfolio also seems to perform well when borrowing is difficult, while the portfolio's 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 types of shocks exist. Finally, we present a break-even analysis of transaction costs and OOS profitability of combined market portfolios, among other robustness checks 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, while 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, including 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 complete list of the examined spreads can be found in Table 1. We use the universe of 45 commodities, equities, and currencies 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 every single series and the rest series within the same asset class from January 1, 1990, to December 12, 2016.[2] Finally, we keep the top five spreads with the highest correlations as evidence that those pairs move closer together while simultaneously being popular among statistical arbitrage investors. [3] The employed cross-currency pairs are U.S. dollar-denominated (i.e., U.S. dollars per unit of foreign currency). We use the equity indices directly instead of any corresponding exchange-traded funds (ETFs) for the equity spreads, 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.[4] 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 ensure that the issue of nonsimultaneous pricing that often plagues such trading strategies does not 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.[5] Each European equity spread consists of equity indices issued by the same or different stock exchanges (i.e., London and Frankfurt Stock

[2] Except in Heat oil-Gasoil, where the sample period starts in 1995 due to data availability.
[3] Moskowitz et al. (2012) also focus on the most liquid instruments to avoid contaminated returns by illiquidity or stale price issues. We also communicated with fund managers and statistical arbitrage investors, who confirmed that our chosen pairs are frequently advertised by trading websites or launched by financial market companies, such as the CME and ICE groups.
[4] 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 technical trading performance and that no nonsynchronous trading issues exist.
[5] For example, the closing times for agricultural futures are 13:20 and 07:45, whereas for energy and precious metals is 17:00 EST.

Exchanges and Euronext Paris), which have the exact actual closing times after considering the 1-hour time difference. The same holds for each of the considered U.S. indices, which the NYSE mainly issues. In addition to this, because we obtain our foreign exchange rates data from the WM/Reuters FX benchmarks, all 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 underlying leg from the other because our main aim is to capture their dominant trends and reversals through technical analysis. This approach 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 calculating the spreads, and if not, to adjust for that.[6] For the equity and exchange rate spreads, such an issue does not exist. 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.[7] Regarding the former, the annualized mean and standard deviation and the *p*-values from the first-order au-

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

[7] 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 whole sample into five different subsamples, the pairwise correlations have been calculated again, and the spreads selected concerning the highest correlation coefficients remain the same across all subperiods.

tocorrelation 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.[8]

[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 currency spreads. However, there are also considerable deviations of volatility levels even within the same asset class, especially in 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. The statistical significance is even more substantial in the equity and currency pairs cases at the 1% and 5% significance levels, respectively. This evidence strongly supports the use of also trend-following TTRs. Regarding 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 correlation levels of the pairs selected are pretty high.

3. TTR universe

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

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. Following previous studies, we also consider numerous variations of the above TTRs (see STW; BS; HTW). Those rules are categorized into *momentum/trend-following* and *contrarian/mean-reverting* rules, usually employed by pair traders to identify *overbought/oversold levels*.

Filter rules: Follow strong trends by taking long (short) positions accordingly. They allow the initiation of a spread's trader position only in response to significant price trends. Therefore, an investor buys a spread if its price increases by a fixed percentage from a previous low, and she sells it if the price decreases by a fixed percentage from a previous high. We assume three different filter rule variations, which consider a certain number of previous days to define high/low values, allow for neutral positions and hold a position for a fixed number of days.

Moving averages: Attempt to ride trends and take positions when crossovers occur, between the spread value and moving average of a given length or between two moving averages of different lengths, signifying a break in the trend. This upside (downside) penetration of a moving average helps an investor discover new trends and maintain her position as long as the crossover remains. We adopt four moving average rule variations, which consider a certain number of previous days to define the crossover between the spread and a moving average, a certain number of previous days to define the crossover between a short and a long moving average, where in this case the number of formation days of short and long moving averages is different, and hold a position for a fixed number of days for both previous variants.

Support and resistance rules: 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. 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 equilibrium. However, if the price breaks through a specific resistance or support level by a certain amount, it is more likely to continue moving in the same direction until it finds a new equilibrium. We assume two variants of support and resistance rules. For the first one, a prespecified number of previous closing values is adopted to define the local maximums/minimums, while for the second one, a holding-period filter is also added, similar to the previous families.

Channel breakouts: Similar to having time-varying support and resistance levels that form a channel of a fixed percentage, leading to a signal when a pair's price penetrates the channel from above or below. A buy or sell signal is generated as soon as one of these trend lines is *broken*. Thus, an investor goes long (short) when the price moves above (below) the channel. The graphical representation of a price channel is equal to a spread of parallel trend lines drifting together within a certain width. We consider two variants of channel breakouts, as follows, a prespecified number of previous days to construct a channel of a certain percentage as the difference between the local maximums and minimums, holding the same position for a specific number of days.

Relative strength indicators (RSIs): They belong to the general family of *overbought/oversold* indicators and attempt to capture a correction of a pair’s extreme price movement in the opposite direction. The RSI estimates the dominance of an upward movement relative to the dominance of a downward movement. In its simplest version, an RSI of value 70 characterizes a specific spread as overbought, and so, a sell signal is triggered, whereas a value of 30 rates the spread as oversold, in which case a buy signal is generated. Three RSI variations are assumed in our study. A simple version of RSI based on the benchmark overbought/oversold levels (i.e., 70 /30), an RSI variant with different values of overbought/oversold levels, and an alternative one in which neutral positions are taken after a certain number of days from which the initial position is triggered.

Bollinger band reversals: Attempt to identify overbought and oversold market levels, defined as the price is a particular distance away (in terms of standard deviations) from its moving average of a given length. Thus, any breakout above or below moving average bands of a given width concerning each spread's standard deviation is a significant event. Investors believe the closer the prices move to the upper (lower) band, the more overbought (oversold) the market, and a short (long) position is taken. This can lead to a pullback of asset prices. We assume two Bollinger band variations, one which takes a neutral position when the spread value reverts to its moving average value after a long/short position is taken and one which holds the long/short position for a certain number of days before neutralizing.

Commodity channel index (CCI) rules: Similar to a combination of RSIs and Bollinger band reversals, they try to quantify the connection between a pair’s price, its corresponding moving average, and its standard deviation, but a specific inverse factor is used to scale the index in this rule. The CCI measures the current price level relative to an average price level over a specific period and is relatively high when prices are far above the moving average and vice versa. As a *reversal* indicator, the CCI searches overbought (i.e., $CCI > +100$) or oversold (i.e., $CCI < -100$) conditions to foretell a mean reversion. Similarly, bullish and bearish divergences can also detect early momentum shifts, leading to trend reversals. Three variants of CCI are employed. The first one uses different values above +100 and -100 to trigger short and long signals, respectively. The second one holds the same position for a certain number of days, while the third variant is a particular case of a CCI and divergence breakout. Divergences can foresee a potential trend reversal point as they usually reflect a change in momentum.[9]

The above TTRs along with the variants of each family of rules and a spectrum of different plausible parameterizations of each variation 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 the variations and the various parameterizations of the families of trading rules 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 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 one 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 consider the short-term interest rates of each currency. Thus, we consider two parts to construct the excess returns: the simple return of each component,

[9] A detailed description of a divergence breakout can be found in our Online Appendix.

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 primary performance metric for creating portfolios of significant TTRs (see also STW; BS) and the MPPM to assess whether any outperformance results from 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

$$SR_{j,t} = \frac{\overline{f}_j}{\hat{\sigma}_j},\ j = 1, \dots, l, \tag{2}$$

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

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

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, making this metric appropriate for our multiple-hypothesis-testing framework (Harvey and Liu, 2015).[10]

[10] 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 $SR = \frac{\hat{\mu}}{\hat{\sigma}}$.

Ingersoll et al. (2007) proposed that the MPPM measures the return on a fund's strategy relative to the amount of accepted risk over a specific period while resembling a representative utility function. The measure considers return distributions 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, employing such a measure is to verify that potentially generated OOS Sharpe ratios are not inflated. The MPPM is strictly increasing and concave 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

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

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 has been calculated above. 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.[11]

So far, we have not considered the impact of transaction costs on the TTRs' performance over the examined spreads. These costs may be pretty high in practice, especially for statistical arbitrage traders who form long-short portfolios. Additionally, the potential predictability of a selected strategy before implementing transaction costs can be easily neutralized when those costs are adjusted

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

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 considering the one-way transaction costs of each component separately.

Following Locke and Venkatesh (1997), we consider one-way transaction costs of 3.3 basis points for a position taken on each 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 to trade U.S. stock indices (Allen and Karjalainen 1999). However, recent evidence shows that *live* trading costs faced by real-world arbitrageurs are pretty lower.[12] Based on the findings of Frazzini et al. (2015) and communications with several brokerage firms, we consider 19.73 basis points as the one-way trading costs for stock indices. This cost level 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. The only transaction costs investors face when trading currencies 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

[12] For example, Frazzini et al. (2015) estimate that market impact, which covers the most significant part of the execution cost for a large institutional trader, is under nine basis points, on average, for trades executed during a day. In contrast, 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.

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 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.

5. Data-snooping bias

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

Data-snooping bias has become even more urgent as of late because of the use of large datasets involving data replication issues by investors and researchers. Classical statistical inference focusing on single-hypothesis testing for each set of series, 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 pretty 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 et al., 2010; HTW; BS, Shynkevich, 2012a; 2012b, 2012c, 2016).

Studies trying to control data snooping bias are divided into two classes, based on the primary criterion employed for statistical inference, those using the *family-wise error rate* (FWER) and those using the *false discovery rate* (FDR). Their difference is mainly intuitive. The FDR is defined as the proportion of false discoveries among the total rejections, while the FWER estimates the probability of making at least one false rejection.

Starting with studies using the latter criterion, White (2000) introduces the so-called *bootstrap reality check* (BRC), which focuses on the statistical significance only of the *best* performing strategy, drawn from several strategies, while contemporaneously tests whether the significance of all strategies is less than the nominal significance level. Following studies have also tried to reassess the power of the BRC test concerning TTR's predictability in different markets (see Hsu and Kuan, 2005; Marshall et al., 2008a, 2008b). Focusing also on the *maximal* trading rule, Hansen (2005) proposes the *superior predictive ability* (SPA) test to correct drawbacks of the BRC test by utilizing studentized test statistics, while setting fewer weights to the test statistics of rules showing poor performance. Studies that exploit TTRs performance on equities markets via the SPA test also include those of Hsu and Kuan (2005) and Shynkevich (2012a; 2012b, 2012c).

Incorporating the assumption that investors are keen on discovering all the statistically significant TTRs showing positive performance instead of investing their total wealth in the maximal one, Romano and Wolf (2005) suggest their *stepwise multiple testing* (*StepM*) method as an improvement to the *single-step* BRC test of White (2000). They follow a stepwise structure, in which individual *p*-values are placed in an acceding order, after bootstrapping the empirical distribution of each rule. After comparing each *p*-value with a nominal significance level during the first step, they replicate the same mechanism in the second step, after excluding the statistically significant rules of the first step. Similarly, Hsu et al. (2010) develop a stepwise extension of the SPA test of Hansen (2005) to minimize the influence of poor performers on the power of the test. HTW (2016) and Shynkevich (2016) use the same extension to evaluate the profitability of TTRs in foreign exchange and bond portfolios, respectively.

Although the above developments aim to achieve a good trade-off between *Type I errors* and *Type II errors*, the FWER control is not well qualified for finance applications as it lacks power and is mechanically affected by the number of hypotheses being tested (Chordia, Goyal and Saretto, 2020). 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 to improve the power of detecting more significant discoveries while having an optimal balance in minimizing *Type I* and *II errors* (Abramovich et al., 2006).

In finance, BSW (2010) propose a modified $FDR^{+/-}$, based on Storey's (2004) FDR approach, which aims to discover significant alphas in mutual fund performance while allowing a separate quantification of false discoveries among funds. By developing such a framework, BSW (2010) try to accommodate what investors do in practice, who usually assess and combine multiple strategies 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 comparative advantage of the $FDR^{+/-}$ over the FWER. 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 crucial assumption for our study because the TTRs included in our universe satisfy this feature by being dependent on small blocks within the same family but essentially independent across different families.

On the other hand, the $FDR^{+/-}$ approach has been reviewed on its ability to accurately detect the proportions of mutual funds with zero and significant performance. Andrikogiannopoulou and Papakonstantinou (2019) found that the zero-alpha proportion of funds in BSW (2010) to be upward biased and so over-conservative, especially when the number of series' observations is small (e.g., monthly). In their reply, BSW (2019) have convincingly shown that $FDR^{+/-}$ is still capable of producing efficient and consistent estimators especially when the number of time series observations is large, such as in our case, at which daily series observations of multiple years returns are considered. Also, our study is mainly a prediction exercise, contrary to the skill detection of funds generating significant alphas. This means that even slight over-conservativeness in estimating lucky rules is not a big issue as we mostly care about the most significant predictors among TTRs. For the above reasons, we adopt the $FDR^{+/-}$ test as the most suitable multiple-hypothesis-testing method for our experiment.

5.2. FDR methodology

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}: $\varphi_j > 0 \; or \; \varphi_j < 0$).[13] In our case, we consider the annualized Sharpe ratio as a 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 $\widehat{FDR}^+ = \frac{\hat{F}^+}{\hat{R}^+}$, where $\hat{F}^+$ and $\hat{R}^+$ are the estimators of $F^+$and R^+, respectively.[14]

[13] 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.

[14] 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.

To estimate 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 γ is the p-value cutoff, multiplied by $1/2$ since we assume symmetry in the appearance of lucky rules in the two tails of the empirical distribution. The FDR method tries to capture information from the center of the distribution of test statistics (i.e., φ_j), mostly dominated by rules with no significant performance (either positive or negative), to correct any luck in the two tails (BSW) (2010). 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

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda;\ j = 1, \ldots, l\}}{l(1-\lambda)}, \tag{6}$$

where $\lambda \in [0,1)$ is a tuning parameter indicating the specific level above which the null p-values should appear. After acquiring the estimator of π_0, and given the number of significantly positive TTRs (i.e., $\#\{p_j \leq \gamma, \varphi_j > 0;\ j = 1, \ldots, l\}$), we can obtain a conservative estimator for FDR^+ under threshold γ as

$$\widehat{FDR^+}(\gamma) = \hat{F}^+/\hat{R}^+ = \frac{1/2\hat{\pi}_0 l\gamma}{\#\{p_j \leq \gamma, \varphi_j > 0;\ j = 1, \ldots, l\}}. \tag{7}$$

Finding γ is essential for controlling FDR^+ at an acceptable level of erroneous selections, allowing us to select rules with p-values $\leq \gamma$.

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 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 lengths of consecutive observations of returns (Politis and Romano 1994). [15] 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 correct *p*-value cut off 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. In this regard, we start by placing the *p*-values of the rules with a positive performance in ascending order. Then we compute FDR^+ using equation (7) for the rule with the smallest *p*-value. We continue by adding the rule with the second-smallest *p*-value and recompute FDR^+. We repeat the same procedure until the desired FDR^+ target is achieved (i.e., 10%).

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 entire 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 to identify and

[15] 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).

measure the performance of the best rule and assess its ability as a 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).[16] Finally, we avoid investing the proportion of our wealth in a savings account (i.e., at the risk-free rate) when a neutral signal is triggered. This assumption helps us measure the actual returns generated by the FDR portfolios of TTRs without being augmented with the risk-free rate of return.

Table 2 provides evidence for the full sample performance of the 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. 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 seem to follow by including 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,

[16] 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.

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 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 commodities spreads is statistically significant at the 1% level. For the top rules of equity and currency spreads, the picture is almost identical 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 most 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 a portfolio construction tool. Pinpointing now which TTRs contribute the lion's share of the best predictive rule, most 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 top-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, and Bollinger bands are among the top ones. Overall, mean-reverting rules, especially those holding the rule's signal for a certain period, seem to be the best performing for the majority of the spreads examined.

6.2. Out-of-sample performance

To evaluate the actual performance of the TTRs and address the issue of data snooping in more realistic conditions, we employ an OOS analysis in the following sections. Doing so helps us economically evaluate the performance of a portfolio of rules, selected ex-ante, similar to how institutional investors would do so in practice.[17] Such an analysis also provides evidence 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, investigating 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 one year to generate some TTRs. Even though 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 critical issue with the OOS estimation, raised by Harvey and Liu (2015), is splitting 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 closely related to the testing power of the IS and OOS periods. In particular, the shorter is the IS dataset, the greater the chance of missing true discoveries (*Type II errors*), and vice versa. For instance, a 90-10 split of the data will increase *Type I errors*,

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

whereas a 50-50 split will increase *Type II 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.

The FDR procedure is employed during the IS period for each pair 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 four years) constitute the IS period. Although we appreciate that this is still a stiff OOS evaluation, it better matches what traders do in practice than 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 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 consistent with Sharpe ratio findings, leaving no room 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, ranging 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 Heat oil-Gasoil, 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-WTI crude oil spread follows by providing almost equally positive metrics. However, its positive performance concentrates only on the first two periods. The metal spreads also seem to yield a positive performance over the last three periods but of a lower magnitude, which could still be 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.

Regarding 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 in 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.

We now compare the above performance of the 10% – FDR^+ portfolio of TTRs with a naïve benchmark strategy to assess the usefulness of FDR^+ specification in accurately selecting the out-performing rules for each spread. We use the same IS and OOS horizons as above, but this time we evaluate the performance of all rules found with positive performance IS in the five OOS periods without considering their statistical significance. In particular, we create equally weighted portfolios of all TTRs with positive Sharpe ratios IS and trade them OOS. Table 4 reports the related findings, such as the median number of rules with positive Sharpe ratios for each portfolio, the OOS Sharpe ratio of the equally weighted portfolios of *positive* rules, and their computed MPPM.

[Insert Table 4 around here]

Generally, the *naïve* portfolio of rules with positive performance underperforms the 10% – FDR^+ portfolio of TTRs, as shown in Table 3, for most spreads and OOS periods examined. The former portfolio seems to generate negative returns most of the time, with only a few positive exemptions in specific OOS years, such as those of Brent-WTI, Heat oil-Gasoil, Platinum- Gold, CAC-TOPIX, AUD-CAD. But again, the 10% – FDR^+ portfolio yields higher Sharpe ratios and MPPMs than the benchmark portfolio for most of those spreads. There are only a couple of cases (i.e., Brent-WTI in 1996 and 2007; Platinum-Gold in 2016; AUD-CAD in 2016), where the *naïve* portfolio outperforms the portfolio constructed with the FDR^+ specification. However, this outperformance is highly likely due to pure luck as it happens sporadically in specific years. Another important finding is the superiority of the FDR^+ to control for losses. We observe that even in cases where negative returns are generated for both portfolios, the $10\% - FDR^+$ reports far lower negative Sharpe ratios and MPPMs than the benchmark portfolio. A particular case is that of Brent-WTI in 2016, in which the benchmark portfolio yields negative returns, but the FDR^+ portfolio selects no rules and remains out of the market. We also compute and compare the maximum drawdown of the examined portfolios to illustrate that the FDR^+ selected rules suffer less extreme losses compared to the in-sample positive rules only in one of the following sections.

6.3 Drivers of performance

We assess the spreads' risk-adjusted performance, their factor exposures, and their relationship with liquidity, market volatility, and the level of investor sentiment in Table 5. Firstly, in Panel A, we employ Carhart's (1997) four-factor model, and we run monthly time-series regressions of an equally-weighted global portfolio's returns on the size (SMB), value (HML), and cross-sectional momentum (UMD) factors over the 1990-2016 period.[18] We construct the global portfolio based on FDR^+ selections, across all three markets, over the full sample period, while we transform the daily returns to monthly by taking the average portfolio return for each month. We use Carhart's (1997) four-factor model because it also includes the momentum factor due to trend-following apart from contrarian rules in the total universe of TTRs. Panel B reports the results of a similar

[18] All factors employed have been downloaded by Kenneth's, R., French website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). 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 a few observations to produce robust findings.

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 (2007), the time-series log values of the VIX index as a measure of market volatility, especially in periods of turmoil, and the levels of funding and market liquidity shocks in separate regressions.[19][20] 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)[21]. We compute the residuals from the AR(2) process applied to those proxies to obtain the funding and market liquidity shocks.

[Insert Table 5 around here]

Panel A of Table 5 demonstrates significantly negative loadings with the market index and the HML and the UMD factors. However, those are of a tiny magnitude to fully explain the spread portfolio's returns, a finding which is also supported by the small value of adjusted-R^2 (i.e., 6.73%). Our global portfolio also delivers a significant alpha of 0.04%. 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%.

---

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

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

[21] 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/).

Regarding 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 the 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 more significant 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 relationship between U.S. equity momentum returns and liquidity shocks. In this direction, contrarian trading strategies usually provide liquidity to demand pressure 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 that 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. Positive MPPMs of almost analogous magnitude consistently follow those ratios. Commodity pairs consistently outperform the equity and currency ones. The OOS analysis, conducted across five different subperiods, reveals that technical analysis performance has not 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, has not squeezed out potential returns. The economic significance of the returns achieved using TTRs on certain spreads and periods may compensate 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 a global portfolio of spreads 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 the *global* portfolio's returns and the momentum factor. We perform similar regressions by considering the investor's sentiment index of Baker and Wurgler (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 than 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.

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

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.

Asness, C.S., Moskowitz, T.J. & Pedersen, L.H. (2013). Value and Momentum Everywhere. Journal of Finance, 68, 929–985.

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

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

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

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

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

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

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

Chen, H., Chen, S., Chen, Z. & Li, F. (2017). Empirical investigation of an equity pairs trading strategy. Management Science, 65, 19–31.

Chordia, T., Goyal, A. & Saretto, A. (2020). Anomalies and False Rejections, The Review of Financial Studies, 33, 5, 2134–2179.

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.

Gehrig, T. and Menkhoff, L. (2004) The use of flow analysis in foreign exchange: exploratory evidence. Journal of International Money and Finance 23: 573–594.

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

Harvey, C.R. (2017). The scientific outlook in financial economics. Journal of Finance, 72, 1399–1440.

Harvey, C.R. & Liu, Y. (2015). Backtesting. Journal of Portfolio Management, 42, 13–28.

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

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.

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.

Ingersoll, J., Spiegel, M., Goetzmann, W., & Welch, I. (2007), Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies, 20, 1503–1546.

Kondor, P. (2009). Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading. Journal of Finance, 64, 631-655.

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

Levine, A. & Pedersen, L., (2016). Which Trend Is Your Friend? Financial Analysts Journal, 72, 51 – 66.

Ljung, G.M. & Box, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika, 65, 297–303.

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.

Locke, P.R. & Venkatesh, P.C. (1997). Futures market transaction costs. Journal of Futures Markets, 17, 229–245.

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

Marshall, B.R., Cahan, R.H. & Cahan, J.M., (2008a). Can commodity futures be profitably traded with quantitative market timing strategies? Journal of Banking and Finance 32, 1810-1819.

Marshall, B.R., Cahan, R.H. & Cahan, J.M., (2008b). Does intraday technical analysis in the U.S. equity market have value? Journal of Empirical Finance 15, 199-210.

Miffre, J. & Rallis, G. (2007). Momentum strategies in commodity futures markets. Journal of Banking & Finance, 31, 1863–1886.

Moskowitz, T.J., Ooi, Y.H. & Pedersen, L.H. (2012). Time series momentum. Journal of Financial Economics, 104, 228-250.

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.

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

Neely, C.J. & Weller, P.A. (2003). Intraday technical trading in the foreign exchange market. Journal of International Money & Finance, 22(2), 223–237.

Pastor, L. & Stambaugh, F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111, 642-685.

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.

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

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

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.

Shynkevich, A. (2012a). Global equity portfolios and short-term predictability of returns: Is it there? Journal of Financial Markets 15, 438-466.

Shynkevich, A. (2012b). Short-term predictability of equity returns along two style dimensions. Journal of Empirical Finance 19, 675-685.

Shynkevich, A. (2012c). Performance of technical analysis in growth and small cap segments of the U.S. equity market. Journal of Banking and Finance 36, 193-208.

Shynkevich, A. (2016). Predictability in bond returns using technical trading rules. Journal of Banking and Finance 70, 55-69.

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

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

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

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
Heat oil-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
Heat oil-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. The *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**	**18.3**	-0.01	-0.43	0.00	0.00	0.00	0.00
Heat oil-Gasoil	700	-	-	**1.72**	**10.1**	**3.76**	**13.38**	**3.65**	**17.5**	**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 all rules with positive performance

		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	2,400	**1.35**	**15.1**	**0.73**	**6.64**	**0.83**	**4.76**	-1.39	-7.15	-1.27	-1.31
Heat oil-Gasoil	1,862	-	-	**1.01**	**5.14**	**2.78**	**7.77**	**1.66**	**6.93**	**2.01**	**7.91**
Platinum-Gold	1,319	-3.13	-1.85	-1.39	-4.31	-1.19	-4.46	-0.38	-2.02	**1.74**	**2.52**
Gold-Silver	1,802	-0.89	-1.96	-2.25	-8.09	-0.64	-4.78	-1.24	-5.08	-2.01	-1.61
Corn-Soybean	1,464	-1.76	-4.28	-2.28	-2.53	-0.54	-1.51	-1.09	-3.75	-0.93	-3.04
Equities											
FTSE100-CAC 40	271	-2.22	-0.49	-3.84	-2.67	-3.71	-1.58	-2.51	-1.21	-1.84	-0.83
AMST-CAC 40	303	-2.66	-0.73	-3.73	-1.24	-2.41	-0.16	-4.19	-1.32	-1.82	-0.27
S&P500-DAX	450	-4.51	-1.56	-0.56	-1.32	-1.84	-0.19	-2.10	-2.25	-1.41	-2.11
CAC-TOPIX	454	-2.84	-0.70	-1.46	-0.95	**1.15**	**1.85**	**0.17**	**0.54**	**1.72**	**3.98**
DAX-FTSE100	407	-1.48	-0.33	-0.65	-0.51	-6.68	-1.75	-2.70	-3.00	-2.86	-0.54
Currencies											
EUR-CHF	186	-2.71	-0.35	-1.81	-0.01	-0.73	-0.01	0.00	0.00	-2.62	-0.48
EUR-JPY	2,124	-2.91	-2.94	-2.88	-2.59	-1.21	-0.68	-1.41	-5.29	-3.06	-6.27
AUD-CAD	1,388	-1.53	-3.40	-2.42	-0.34	-0.87	-0.25	-2.78	-3.11	**0.35**	**0.11**
AUD-NZD	500	-3.55	-0.76	-0.94	-0.29	-0.45	-0.22	-2.17	-0.49	-2.22	-0.57
NOK-SEK	298	-2.40	-5.34	-1.80	-0.58	-1.00	-0.11	-1.78	-0.36	-5.60	-2.31

We report out-of-sample annualized Sharpe ratios and MPPMs for the last year of each subperiod based on equally weighted portfolios of all positive rules found in-sample. 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. 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's (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 (2007), and the VIX index separately.