Performance of technical trading rules: Evidence from the crude oil market[*]

Ioannis Psaradellis[†], Jason Laws[‡], Athanasios A. Pantelous[*], Georgios Sermpinis[§]

Abstract

This study investigates the debatable success of technical trading rules, through the years, on the trending energy market of crude oil. In particular, the large universe of 7846 trading rules proposed by Sullivan et al. (1999), divided into five families (filter rules, moving averages, support and resistance rules, channel breakouts, and on-balance volume averages), is applied to the daily prices of West Texas Intermediate (WTI) light, sweet crude oil futures as well as the United States Oil (USO) fund, from 2006 onwards. We employ the *k-familywise error rate* (*k*-FWER) and *false discovery rate* (FDR) techniques proposed by Romano and Wolf (2007) and Bajgrowicz and Scaillet (2012) respectively, accounting for data snooping in order to identify significantly profitable trading strategies. Our findings explain that there is no persistent nature in rules performance, contrary to the in-sample outstanding results, although tiny profits can be achieved in some periods. Overall, our results seem to be in favor of interim market inefficiencies.

Keywords: Crude Oil; Technical Trading; Data Snooping; Transaction Costs; Persistence; Market Efficiency

JEL classification: C12, C15, G11, G14

[*]The authors would like to acknowledge the gracious support for this work provided by the EPSRC and ESRC Centre for Doctoral Training on Quantification and Management of Risk & Uncertainty in Complex Systems & Environments (EP/L015927/1). We would like to thank participants at the Econometrics and Financial Data Science (EFDS 2017, ICMA centre) workshop Forecasting Financial Markets (FFM 2016), Quantitative Finance and Risk Analysis (QFRA 2016), and WFMA Xian Jiatong-Liverpool University conferences, and at the Universities of Liverpool and Glasgow, as well as Carol Alexander, Neil Kellard, Marcel Prokopczuk, Michalis Stamatogiannis, Andrea Roncoroni, Ehud Ronn, Ernest P. Chan, Chris Florackis, Vasileios Kallinterakis, and Michael A. H. Dempster for helpful comments. Any remaining errors are ours.

[†] School of Economics & Finance, University of St Andrews, The Scores, KY16 9AR, St Andrews, UK

[*] Department of Econometrics and Business Statistics, Monash Business School, Monash University, Clayton VIC 3800, Australia. Corresponding author: Athanasios.Pantelous@monash.edu

[‡] Management School, University of Liverpool, Chatham Street, L69 7ZH, Liverpool, UK

[§] Adam Smith Business School, University of Glasgow, University Avenue, G12 8QQ, Glasgow, UK

Performance of technical trading rules: Evidence from the crude oil market

This version: 16th of September, 2018

Abstract

This study investigates the debatable success of technical trading rules, through the years, on the trending energy market of crude oil. In particular, the large universe of 7846 trading rules proposed by Sullivan et al. (1999), divided into five families (filter rules, moving averages, support and resistance rules, channel breakouts, and on-balance volume averages), is applied to the daily prices of West Texas Intermediate (WTI) light, sweet crude oil futures as well as the United States Oil (USO) fund, from 2006 onwards. We employ the *k-familywise error rate* (*k*-FWER) and *false discovery rate* (FDR) techniques proposed by Romano and Wolf (2007) and Bajgrowicz and Scaillet (2012) respectively, accounting for data snooping in order to identify significantly profitable trading strategies. Our findings explain that there is no persistent nature in rules performance, contrary to the in-sample outstanding results, although tiny profits can be achieved in some periods. Overall, our results seem to be in favor of interim market inefficiencies.

Keywords: Crude Oil; Technical Trading; Data Snooping; Transaction Costs; Persistence; Market Efficiency

JEL classification: C12, C15, G11, G14

1. Introduction

Technical analysis (sometimes referred to as *chartism*) is believed to be one of the longest-established forms of investment analysis, being a set of graphical or mathematical techniques exploring future trading opportunities for financial assets just by analyzing the time-series history of their asset prices, volume data, and a summary of securities statistics. Brock et al. (1992) mention that technical trading rules "*beating the market*" is supposed to be as old as the U.S. stock market itself. Nowadays, investment funds, brokerage firms, and trading platforms from all over the world utilize numerous types of technical indicators and oscillators as prospective moneymaking tools.

On the other hand, despite the undisputable popularity of technical analysis among practitioners, academia has been rather skeptical for a long time now about its merits, and there is an ongoing debate as to whether the generated profits are just lucky. On the *effectiveness* of this form of analysis and its power to yield profits, Malkiel (1981) describes it pertinently as the "*anathema*" of the academic world, which usually loves to pick on it. This argument is derived from the Efficient Market Hypothesis (EMH), which expresses that security prices reveal all the available information to investors. However, since the 1960s, prominent academics and practitioners have claimed that predictable patterns do exist in returns (especially in certain periods of time), which can lead to *abnormal* profits.[1] In this regard, even Keynes (1936) outlines that most traders' decisions can be deemed a consequence of "*animal spirits*".

This fruitful debate has culminated in a large number of empirical studies employing technical trading rules in several markets and for different indices. Some have found results to support the notion that trading strategies are able to deliver superior returns, at least in certain time periods (Neftci 1991; Brock et al. 1992; Neely et al. 1997; Conrad and Kaul 1998; Sullivan et

[1] Earlier studies of technical analysis and patterns in stock returns include Alexander (1961, 1964), Fama (1965, 1970), Fama and Blume (1966), Levy (1967), James (1968), Jensen and Benington (1970), and Sweeney (1980).

al. 1999; Lo et al. 2000; Kavajecz and Odders-White 2004; Qi and Wu 2006; Hsu et al. 2010; Neely and Weller 2011; Shynkevich 2012; Taylor 2014). Despite this, other papers report that technical trading strategies are unable to predict future prices, especially when transaction costs are considered (Bessembinder and Chan 1998; Allen and Karjalainen 1999; Ready 2002; Marshall et al. 2008; Bajgrowicz and Scaillet 2012; Yamamoto 2012).

Inevitably, *data snooping* effects arise in most of the studies mentioned above, particularly when a large number of trading strategies are implemented and tested. Amongst the pioneers in studying the data snooping effects are Jensen and Benninghton (1970), defining it as "*selection bias*", as well as Lo and MacKinlay (1990) who summarize that more likely patterns can emerge when data are severely exploited. This is apparently true if one considers that, by exploring a sizeable universe of different trading rules, it is highly likely one will find a rule that works well, even by chance. Many efforts have been made to minimize the undesirable consequences of data snooping, which are illustrated in the studies of White (2000), Romano and Wolf (2005), Hansen (2005), Romano and Wolf (2007), Romano et al. (2008), Hsu et al. (2010), Bajgrowicz and Scaillet (2012), and Hsu et al. (2014).

In this paper, we evaluate the performance of the whole universe of 7846 technical trading rules (TTRs) proposed by Sullivan et al. (1999) on the crude oil market. This universe of rules is the most popular and common, and creates a connection with the previous literature in this field of research (Brajgowicz and Scaillet 2012; Marshall et al. 2008). In particular, we apply five families of strategies (i.e., *filter rules*, *moving averages*, *support and resistance*, *channel breakout* and *on-balance volume*) to the daily prices of West Texas Intermediate (WTI) light, sweet crude oil futures, as well as the United States Oil (USO) fund, covering a period from April 2006 to January 2016.[2]

[2] We chose this specific period in order to examine the TTRs' performance on the same trading days for crude oil futures and the USO, given that the inception date of the USO was in April 2006.

Crude oil futures offer the opportunity to trade one of the world's most liquid oil commodities on the New York Mercantile Exchange (NYMEX) for up to 108 consecutive months.[3] In line with Wang and Yu (2004) and Marshall et al. (2008), we employ Datastream continuous price series of crude oil futures, which represent the price of the most actively traded contract. This guarantees that the underlying instrument should last longer than the observation period when analyzing the performance of the TTRs. Furthermore, futures markets are more attractive for pursuing active trading strategies than stock markets, since they involve much lower transaction costs (e.g. spreads and commissions), and short-selling is easily applied.

The USO is the largest and most liquid oil-related exchange traded fund (ETF) ($3.7 billion in assets), and is designed to track the daily price movements of WTI light, sweet crude oil.[4] It is exposed to crude oil prices by means of holding positions on front-month crude oil futures contracts. The USO is free of the substantial storage costs involved in other crude oil inventories, entailing low total costs of management, a feature that makes it very attractive to investors. As far as we are concerned, this is the first time that the effectiveness of TTRs will be tested specifically on the crude oil market. We believe that it is a rather interesting area of investigation, since crude oil prices have exhibited considerable fluctuations over the years in response to geopolitical and economic turmoil.[5] Today the oil industry is being shaped by one of its most dramatic price movements of recent times, having experienced almost a 70% fall since June 2014. These extreme fluctuations mark the crude oil market out as a *trending* market, potentially lucrative for applying the TTRs, since trend following is one of the key aspects of technical analysis. Furthermore, since previous empirical findings on the hedge fund industry and

[3] The final settlement date is the 4th U.S. business day prior to the 25th calendar day of the month preceding the contract month.

[4] The USO periodically "rolls over" its underlying futures contracts by selling those that are approaching expiration and buying those that expire farther into the future. The investment objective of the USO is publicized on its website (http://www.unitedstatesoilfund.com/).

[5] For instance, in 2008, crude oil reached its highest value, followed by a fall below $50 per barrel due to the Lehman Brothers crisis in 2009.

the Dow Jones Industrial Average (DJIA) index have shown that TTRs perform quite well when *strong* negative or positive returns occur in the market (Fung and Hsieh 1997; Bajgrowicz and Scaillet 2012), we have a strong motivation to explore them on the crude oil market as well.

In this paper, the first contribution is to revisit the historical success of TTRs in the oscillated market of crude oil.[6] This allow us to access the power of multiple hypothesis testing methods in an environment in which momentum trading rules tend to work well by definition, and so measure the level of significance of rules yielding positive performance. For this purpose, we divide and assess the rules' performance in four different subperiods, each one characterized mostly by having bearish or bullish trends. As crude oil futures and ETFs have small expense ratios compared to other assets, there is a strong potential for trading rules to achieve gains. On the other hand, low transaction costs may help to increase market liquidity, leading to market efficiency (Hedge and McDermott 2004). Thus, the second contribution is to give an answer to the question of which one of the above two cases holds in the case of the crude oil market. With this aim, for the in-sample analysis, we compare the performance when applying TTRs to the USO and the crude oil futures to explore potential differences due to contango or backwardation effects as a result of rolling over crude oil futures in the case of the USO. An in-sample analysis including the impact of transaction costs is also reported. The transaction costs are embodied endogenously in the trading rule selection process, each time a buy or sell signal is generated. This helps investors foresee which TTRs' performance can outweigh transaction costs ex ante (Bajgrowicz and Scaillet 2012). The reason for this is that strategies' predictability is occasionally neutralized when TTRs are selected before the implementation of transaction costs because of frequent signals. Another contribution is that, instead of only using regular evaluation criteria, such as the mean return and the Sharpe ratio, we also employ the Calmar ratio criterion as

[6] Marshal et al. (2008) evaluate the performance of the Sullivan et al. (1999) universe of TTRs in 15 major commodities futures series, while considering naïve methods of accounting for data snooping effects. One of these series refers to crude oil futures covering the period 1984-2005.

an essential performance measure for technical traders, especially when momentum strategies are employed. Indeed, the Calmar ratio is an important indicator for the hedge fund industry in general since it displays the average annual return on an investment, per unit of maximum drawdown (Schuhmacher and Eling 2011). In technical analysis its usage is particularly useful, especially when momentum strategies, which can suffer significant drawdown, are employed.

Finally, a comprehensive persistence analysis of TTRs is employed. For that analysis the *false discovery rate* (FDR) technique (Barras, Scaillet and Wermers 2010; Bajgrowicz and Scaillet 2012; Bajgrowicz et al., 2016) is used to minimize data snooping effects. The performance of the FDR technique is compared with the equally powerful *k-familywise error rate* (*k*-FWER) technique of Romano and Wolf (2007) and Romano et al. (2008), rather than more conservative methods such as the *bootstrap reality check* (BRC) of White (2000) or its stepwise extension proposed by Romano and Wolf (2005). The rationale behind this is our desire to investigate whether a generalized version of the conservative FWER measure could demonstrate the same performance level as the powerful FDR. In particular, we are the first to assess the out-of-sample performance of a portfolio of genuine TTRs, while applying the FDR and *k*-FWER methods respectively, and also including transaction costs. The portfolio is constructed and rebalanced on a semi-annual basis, each time using data from the previous six months and evaluating its profitability in the next half of the year. We report that the powerful nature of the FDR approach to identifying genuine TTRs is also verified in the case of crude oil. Furthermore, we observe that the less conservative *k*-FWER than the strict FWER of Romano and Wolf (2005) can achieve similar results to the FDR in terms of trading performance, while also allowing a certain number of false selections to occur. Moreover, the FDR succeeds in selecting a larger amount of genuine TTRs than the *k*-FWER portfolio.

For the in-sample simulation period, the findings indicate that more than half of the TTRs exhibit great predictive power, especially in periods of substantial crude oil price movements.

Additionally, the best TTRs are able to achieve high mean returns, as well as Sharpe and Calmar ratios, across the whole period considered. On the other hand, when it comes to persistence analysis, the TTRs selected by the data snooping methods show no persistent nature in the out-of-sample period. Although, the portfolios are able to generate positive performance in some periods, we observe that the superior returns are considerably small. However, there is only one period in which both portfolios achieve a Sharpe ratio slightly bigger than 1. Overall, we conclude that the best-performing TTRs mostly are accessible only to investors observing the returns ex post, in an in-sample period, and therefore it is not easy for them to foresee truly out-of-sample profitable rules ex ante, without hindsight.

The remainder of the paper is structured as follows. In Section 2, a detailed description of the universe of TTRs proposed by Sullivan et al. (1999), as well the performance criteria, are provided. Section 3 describes the time series and the descriptive statistics of the data considered. Section 4 presents a synopsis of the existing methods accounting for data snooping. A detailed description of the FDR and *k*-FWER approaches, as well as the portfolios' characteristics, is provided in the same section. Section 5 provides evidence of the TTRs performance in the in-sample period, with and without consideration of transaction costs. In Section 6, the persistence analysis is presented, while accounting for transaction costs at the same time. Finally, Section 7 presents the concluding remarks.

2. Technical trading rules universe and performance measures

In Section 2.1, we review the universe of TTRs proposed by Sullivan et al. (1999). We briefly present the performance measurement tools of mean return, Sharpe ratio, and Calmar ratio (Brock et al. 1992; Sullivan et al. 1999; Bajgrowicz and Scaillet 2012) in Section 2.2.

2.1. Technical trading rules universe

Technical analysis incorporates a large spectrum of approaches as a form of predictive modeling. Although these methods use mostly graphical rather than mathematical or statistical tools, they use time series of past prices, volumes, and other observables to define whether a buy (long), neutral (out of the market), or sell (short) strategy should be taken within the next time period. As stated earlier, we adopt the whole universe of 7846 TTRs for each subperiod for comparison purposes. The universe is separated into five categories of indicators, while different parameterizations can be employed for each rule.[7]

Filter rules: An investor *buys* if the 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*. An alternative definition of subsequent highs (lows) can be defined as the highest (lowest) closing price observed over a prespecified number of previous days, excluding the current day. Thus, the filter rule allows the initiation of an investor's position only in response to major price trends. We also consider the impact of two extra filters. The first one allows a neutral position when the price increases (decreases) from a previous low (high) by a smaller percentage than the percentage needed to initiate a buy (sell) position. The second one assumes a position is held for a fixed number of periods.

Moving averages: Assumes a crossover between short-long moving averages to generate a trade. Usually, an investor buys (sells) when the short-moving average moves above (below) the long-moving average. These upside (downside) penetrations of a moving average help an investor to discover new trends and maintain his position as long as the crossover remains. Three extra filters are applied. The first one demands that the short-moving average penetrates the long-moving average by a fixed percentage, otherwise no position is initiated. The second

[7] The interested reader can refer to the appendix of Sullivan et al. (1999) for a detailed description of each rule as well as the extra parameters used.

one applies a delay filter, which requires a signal to remain valid for a prespecified number of days before an action is taken. The third one considers a holding period similar to the one employed in filter rules.

Support and resistance: A trader buys (sell) when the price rises above (below) the local maximum (minimum) over the previous *n* days. The intuition behind this rule is that usually investors think that sooner or later the movement of the equity's price will tend to stop and return to a certain level (sell at the peak and buy at the bottom). However, if the price breaks through a certain resistance (support) level, it is more likely to continue drifting upward (downward) until it finds a new resistance (support) level. Thus, a buy (sell) signal is triggered. An alternative definition of extreme highs/lows is also used, similar to the one employed in filter rules. In addition to that, fixed-percentage-band, delay and holding-period filters are imposed.

Channel breakouts: An investor buys (sells) when the price moves above (below) the channel. A channel occurs when the high over the previous, prespecified, days is within a fixed percentage of the low over the previous prespecified days. The graphical representation of a price channel is equal to a pair of parallel trend lines. As soon as one of these trend lines is "broken", a buy or sell signal is generated. A fixed percentage band is also exercised around the channel, as well as a holding period for each position triggered.

On-balance volume averages: These operate in a similar way to the moving-average rules (crossover between short/long on-balance volumes). However, the indicator here is the volume. The economic meaning is that the volume is greater on days when the price movement is an extreme fall (bearish) or an extreme rise (bullish). A technical trader adds (subtracts) the daily volume to (from) that of the previous day, when the current closing price has increased (decreased), in order to construct the new on-balance volume indicator. Then, a moving average is applied. Furthermore, the same filters as in the case of the moving average are used.

All of the trading rules described above are considered *momentum* or *trend-following* TTRs except for the support and resistance rules that can be deemed *contrarian* (mean-reverting) trading strategies.

2.2. Measuring performance

The performance of the TTRs is assessed through the *mean return* and *Sharpe ratio* criteria. The mean return is the absolute criterion of each rule's returns, while the Sharpe ratio is a relative performance criterion since it represents the ratio of the average excess return to the total risk of the investment. Practically speaking, the TTRs earn the risk-free rate in periods when a neutral signal is triggered.[8] In our analysis, we are the first in the relevant literature to evaluate the performance of TTRs by also employing the Calmar ratio. The Calmar ratio[9] is an important indicator for investment banks as well as the hedge fund industry in general, since it displays the average annual return of an investment per unit of maximum drawdown. Furthermore, practitioners find it of great importance, especially when they are dealing with momentum strategies that can suffer a considerable drawdown. On the other hand, the Sharpe ratio is mostly applied for mean-reverting or contrarian strategies.[10] We believe that evaluating the performance of TTRs using only the Sharpe ratio has some drawbacks that can be rather misleading when setting an optimal portfolio (Jondeau and Rockinger 2006).

Specifically, let $s_{j,t-1}$ denote the trading signal for each trading rule $j, 1 \leq j \leq \ l$ (where $l = 7846$) at the end of each prediction period $t - 1$ $(\tau \leq t \leq T)$, where $s_{j,t-1} = 1, 0,\ or - 1$ represents a long, neutral or short position taken at time t. In addition to that let r_t designate the

[8] Actually, following the studies of Brock et al. (1992), Sullivan et al. (1999), and Bajgrowicz and Scaillet (2012) who implement the "double-or-out" trading strategy, a buy signal leads a trader to borrow money at the "risk-free" rate in order to double the investment in the commodity portfolio, a neutral signal leads to the trader simply holding the commodity, and when a sell signal occurs the trader liquidates and exits the market.

[9] Developed by Young (1991), the Calmar ratio stands for California Managed Account Reports. It is a performance measurement used to evaluate commodity trading advisors and hedge funds.

[10] We acknowledge Ernest P. Chan for pointing this out to us.

return of the price series exercised, and r_t^f be the “risk-free” rate.[11] The mean return criterion $\overline{f}_{j,t}$ for the trading rule j at time t is defined by

$$\overline{f}_{j,t} = \frac{1}{N}\sum_{\tau=R}^{T} \ln\left(1 + s_{j,t-1}r_t\right),\ j = 1, \dots, l,$$

where $N = T - \tau + 1$ is the number of days examined. We denote as τ the start date for each subperiod, and even for the first one, since lagged values up to 250 days are employed in the universe of rules. Then, the Sharpe ratio criterion expression SR_j for trading rule j at time t is defined by

$$SR_{j,t} = \frac{1}{N}\sum_{t=R}^{T} \frac{\ln(1+s_{j,t-1}r_t - r_t^f)}{\widehat{\sigma_j}},\ j = 1, \dots, l,$$

where $\frac{1}{N}\sum_{t=R}^{T} \ln\left(1 + s_{j,t-1}r_t - r_t^f\right)$ and $\widehat{\sigma_j}$ are the mean excess return and the estimated standard deviation of the mean excess return respectively. Finally, the Calmar ratio criterion $Calmar_j$ is obtained as the annualized mean return of each rule j over its maximum drawdown (MDD_j):

$$Calmar_{j,t} = \frac{\overline{f}_{j,t}*252}{MDD_j}, j = 1, \dots, l,$$

where $MDD_j = \min[r_t - \max\{\sum_{t=R}^{T} r_t\}]$.

3. Data description

In this section, the settlement prices and trading volumes for the USO and crude oil futures for four different subperiods are analyzed. Table 1 reports the intervals covered for each one of them, while Fig. 1 represents the time series dynamics for the two underlying instruments examined from April 2007 to January 2016.

[11] We use as a risk-free rate the daily effective federal funds rate, in accordance with all the previous literature.

[Insert Fig.1 somewhere here]

[Insert Table 1 somewhere here]

Subperiod 1 is characterized by a sharp increase and subsequent fall in crude oil prices for both CL futures and USO due to the Lehman Brothers collapse and so it is characterized by mixed trends. Subperiod 2 reveals an upward trend for the CL futures series, but this is not quite observable for USO ones. However, we define this period as a bullish due to the upward trend in crude oil spot prices. Subperiod 3 mainly dominated by mixed trends for both cases, while Subperiod 4 includes their recent extreme fall and for that reason is characterized as bearish in terms of trend. Moreover, we employ the nonparametric change point detection approach of Ross et al. (2011) for detecting location and scale changes respectively on the crude oil time series examined. The identified changes in both crude oil futures and USO time series lead to updated subperiods, which do not substantially differ from the ones we have initially set and defined above. The new subperiods along with the corresponding performance of TTRs are presented in Appendix A.

The summary of descriptive statistics of daily buy-and-hold returns for the USO and crude oil futures for the four subperiods is reported in Table 2. We follow the existing literature and calculate the daily returns as the natural logarithm of price relatives. The distribution characteristics are described by the *mean, standard deviation, skewness* and *kurtosis* statistics as well as the first-order autocorrelation under the Ljung-Box (1978) Q statistics at the 5% significance level.

[Insert Table 2 somewhere here]

The crude oil futures yield positive performance (with 10 basis points as the highest value) for the first three subperiods, with the only exception being the last one in which a highly negative mean return is reported, standing for the bearish market. The USO yields negative returns for all subperiods except the second. The standard deviation is also comparable across both USO and crude oil futures series. However, with regards to skewness, there is a split between negative and positive signs in the case of crude oil futures, while the USO exhibits mostly negative signs. Moreover, a considerable level of kurtosis is observed in both series for all subperiods. The Ljung-Box (1978) Q statistics test indicates that both crude oil futures and the USO have significant first-order autocorrelations in half of the subperiods.[12] Finally, for the last subperiod describing the extreme fall, the first-order autocorrelation is significant for both series.

4. Data snooping bias

Data snooping bias should always be adjusted for when examining the predictive ability of a large number of trading strategies (i.e., technical trading indicators). The issue emerges when the financial dataset is severely exploited by trading rules dependent to each other, such as in our case (i.e., weak dependence between same family rules). This may result in the identification of TTRs that generate profits purely out of luck, and for that reason multiple hypothesis testing is attempted to minimize data snooping bias. In addition, selecting one trading rule recognized as the best, without consideration of the entire universe of strategies that it is pooled from, when its statistical inference is tested can also lead to false discoveries. In our paper, we employ and compare the FDR and the *k*-FWER which are two of the most powerful data snooping methodologies in the relevant literature. In Section 4.1, we review the existing data snooping methods.

[12] Most of the trading rules employed in this study are designed to capture momentum. Their effectiveness is mainly based on the existence of significant autocorrelation of returns series.

The data snooping specifications employed in this study are outlined in Sections 4.2 and 4.3, respectively. The multiple hypothesis testing setup together with the construction of the portfolio of TTRs following the proposed methods are described in Sections 4.5 and 4.6.

4.1. Existing data snooping methods

The finance literature introduces several methods for mitigating data snooping bias. The majority of these focus on two main statistical approaches for testing multiple hypotheses: the FWER and the FDR. The difference between the two is mostly intuitive, rather than based on conscious reasoning. FWER is defined as the probability of making at least one false rejection (which is unacceptable), while FDR views “unacceptability” in terms of a proportion (Harvey and Liu 2014). For instance, a 10% false discovery rate denotes that more than 10 false discoveries in 100 tests would be unacceptable. Thus, the FWER is more conservative than the FDR, especially when the universe of rules is large.

In statistics, the most standard FWER method is the Bonferroni correction, in which individual null hypotheses (for each one of the total universe of rules) are rejected for each p-value less than a significance level of α/l in a single-step procedure. This structure is employed in the BRC of White (2000) and is carried out in such a way as to reassure that the significance level of the contemporaneous test of all l rules is less than α. In this way, the BRC evaluates whether the “best” performing strategy (drawn from l strategies) has significant predictive power with respect to the performance of the whole universe in a two-tailed-hypothesis framework. The null hypothesis tested is that the performance of the best trading rule is no better than the benchmark (e.g. “risk-free” rate):

$$H_{0j}\colon \max_{j=1,..l} \varphi_j \leq 0 \text{ , where } \varphi_j \text{ is the performance measure of the } j\text{th rule.}$$

Even though the BRC is used by Sullivan et al. (1999), we believe that it is a rather conservative measure that lacks power since it focuses only to the best strategy. Now, by also applying the Bonferroni correction, Hansen (2005) presents his *superior predictive ability* (SPA) test, which minimizes the influence of poor and inconsistent strategies by using studentized instead of non-studentized test statistics.[13] However, it also focuses on the rule that appears best for the observed financial series. Furthermore, the call to identify more outperforming strategies and not relying only on the best strategy when undertaking investment decisions led to the Holm (1979) method. The Holm method works in a stepwise structure, with individual *p*-values ordered from smallest (most significant) to largest (least significant), and each one compared with a less strict significance level moving “down” the list. Following the Holm procedure, Romano and Wolf (2005) introduce their *stepwise multiple testing* (StepM) method as an improvement to the *single-step* BRC testing method of White (2000), while Hsu et al. (2010) develop a stepwise extension of the SPA test of Hansen (2005). Although stepwise approaches are powerful tools, their main drawback is that they do not select further rules once they have detected a rule whose performance is due to luck.

In practice, investors do not search only for the best rule, but invest money in all possible outperforming strategies. Romano and Wolf (2007) develop a generalized methodology, controlling for the stringent FWER criterion. Their goal is to reject at least a specific number of false hypotheses to maximize diversification. In a similar way, Hsu et al. (2014) apply this generalization to the stepwise method of Hsu et al. (2010) to minimize the data snooping effects on the performance of the Commodity Trading Advisory fund.

Moreover, the FDR tolerates a certain proportion of false rejections so as to construct a well-diversified portfolio of trading rules, while accounting for the data snooping effect. Thus,

[13] A studentized test statistic refers to a simple test statistic divided by the consistent estimator of its standard deviation. This helps one to compare objects in the same units of standard deviation.

Bajgrowicz and Scaillet (2012) employ the modified FDR$^{+/-}$ version of Barras et al. (2010) in the context of identifying outperforming TTRs on the DJIA index. Their findings confirm the superiority of the FDR over the conservative FWER approach of Romano and Wolf (2005) in detecting and building a portfolio of genuine rules.

4.2. Multiple hypothesis testing framework

As data snooping techniques are actually multiple hypothesis testing procedures, in what follows we need first to define the test statistic. The Sharpe ratio criterion (as defined in Section 2.2) is chosen as the test statistic when performing the multiple hypothesis testing using the *k*-FWER and FDR methods for data snooping.[14] We selected this ratio not only for comparison with previous studies, but also for its undoubtable popularity across traders. The test statistic for each rule j defines the setup under the null hypothesis $(H_{0j}\colon \varphi_j = 0)$ that rule j does *not* outperform the benchmark, where $\varphi_j = SR_j$ in this case. On the contrary, the alternative hypothesis assumes the presence of abnormal performance, positive or negative $(H_{Aj}\colon \varphi_j > 0 \; or \; \varphi_j < 0)$ in a two-tailed test. However, since we are mainly interested in identifying significantly outperforming rules, we define a technical trading rule j as significantly positive, if it displays abnormal performance (i.e., reject H_{0j}) and its performance metric is positive (i.e., $\varphi_j > 0$). The "risk-free" rate is used as a benchmark, describing an investor being out of the market.

[14] We do not apply the Calmar ratio criterion since its formulation is based on at least a couple years of previous data, while in our persistence analysis we use a rebalancing period of six months (see Section 5).

4.3. The FDR$^{+/-}$ method

The *FDR*$^{+/-}$ has its foundations in the FDR statistical criterion introduced by Benjamini and Hochberg (1995), which assumes that, by tolerating a small proportion of false discoveries amongst all rejections (e.g., significant TTRs), one obtains a more powerful multiple hypothesis testing tool than via the conservative FWER method.[15]

The *FDR*$^{+/-}$ has some unique features that make it suitable for traders that are not just looking for the best rule, but also for a class of strategies with genuine predictive power that can help them diversify risk. In Bajgrowicz and Scaillet (2012), the FDR approach provides a sensible trade-off between significantly positive and false selections, making it less strict than the FWER method. Additionally, its comparative advantage is the ability to find the outperforming rules, even if the performance of the best rule in the sample is due to luck. In practice, it is not unusual for such a rule with no significant predictability to achieve the greatest performance in terms of profits. This feature is not available in the other methods, whose stepwise nature prevents them from detecting further outperforming rules once a "lucky" rule has been identified.

The FDR concentrates on estimating the expected value of the ratio of erroneous selections over the rules showing significant performance. Specifically, the *FDR*$^{+/-}$ is defined as the expected value of the proportion of false selections, *F*, among the significant rules, *R* (positive or negative). The latter are just the rules that perform either better or worse than the benchmark while at the same time their *p*-value rejects the null hypothesis of no abnormal performance under some threshold γ. Thus, the estimate is given by $\widehat{FDR}^{+/-} = \hat{F}^{+/-}/\hat{R}^{+/-}$, where $\hat{F}^{+/-}$ and $\hat{R}^{+/-}$ are the estimators of $F^{+/-}$and $R^{+/-}$, respectively. For instance, an *FDR*$^{+/-}$ 100%

[15] The initial FDR version of Benjamini and Hochberg (1995) adopted independence across multiple hypotheses. Later, studies by Benjamini and Yekuteli (2001), Storey (2002), and Storey et al. (2004) proved that the FDR holds under "weak dependence" conditions when the number of hypotheses is very large. Also, Bajgrowicz and Scaillet (2012) explain that the Sullivan et al. (1999) trading rules satisfy this feature, since the rules are dependent in small blocks (within the same family) and independent across different families.

conveys that, among both the outperforming and underperforming trading strategies, no rule generates genuine performance on average and vice versa.

The estimation of *FDR*$^{+/-}$ is not very tedious, especially when the *p*-value of each rule's corresponding test statistic has already been computed. In order to acquire the individual *p*-values, we follow the resampling procedure of Sullivan et al. (1999). Using the stationary bootstrap method of Politis and Romano (1994) to resample the returns of each strategy, the corresponding test statistic for each bootstrap series of returns is calculated.[16] The *p*-value is obtained by comparing the original test statistic (φ_j) to the quantiles of each bootstrapped test statistic vector. The estimate of $\widehat{FDR}$ is given by

$$\widehat{FDR}(\gamma) = \hat{F}/\hat{R} = \frac{\widehat{\pi_0} l\gamma}{\#\{p_j \leq \gamma;\ j = 1, \dots, l\}},$$

where l is the entire universe of TTRs, γ is the *p*-value cut-off and $\widehat{\pi_0} = \frac{\#\{p_j > \lambda;\ j=1,\dots,l\}}{l(1-\lambda)}$ is an estimator of the proportion of rules that show no abnormal (either positive or negative) performance in the entire universe and for a two-sided framework. The estimation of $\widehat{\pi_0}$ requires us to define the tuning parameter λ by visually examining the histogram of all *p*-values.[17] Thus, in our study, λ is chosen by employing the same method.

Following Barras et al. (2010), and after estimating $\widehat{\pi_0}$, we then focus on the right tail of the test statistic distribution (i.e. $\varphi_j > 0$), where the outperforming TTRs lie. Thus, we can compute a separate estimator for $\widehat{FDR}^{+}(\gamma)$.[18] This holds under the assumption that the false discoveries spread evenly between TTRs with positive and negative performance and with equal tail significance $\gamma/2$. Thus, the estimator is

[16] The block length used is equal to *q* = 0.1, and the number of bootstrap realizations is set to *B* = 1000, following previous studies.

[17] Bajgrowicz and Scaillet (2012) set the value of λ just by looking for the level above which the histogram of *p*-values becomes fairly flat, representing the region of null *p*-values. There is also an automated version of this process described by Storey (2002).

[18] The *FDR*$^{-}$ part can be calculated in a similar way.

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

Furthermore, the number of TTRs showing abnormal performance can be extrapolated as $\pi_A = 1 - \pi_0$. Now, defining the positive, π_A^+, and negative, π_A^-, proportions of rules in the population, we acquire $\pi_A^+ = \frac{T(\gamma)^+ + A(\gamma)^+}{l}$ and $\pi_A^- = \frac{T(\gamma)^- + A(\gamma)^-}{l}$, where $T(\gamma)^+$ and $T(\gamma)^-$ symbolize the number of strategies with positive and negative returns, respectively, and *p*-values less than *γ*. On the other hand, $A(\gamma)^+$and $A(\gamma)^-$ indicate the size of alternative models showing positive and negative performance without rejecting the null hypothesis (*p*-value greater than *γ*), respectively.

To conclude, $\hat{T}(\gamma)^+$ (likewise $T(\gamma)^-$) is defined as the estimator of the significantly positive rules minus the estimator of false selections:

$$\hat{T}(\gamma)^+ = \widehat{R^+(\gamma)} - \widehat{F^+(\gamma)} = \#\{p_j \le \gamma, \varphi_j > 0;\ j = 1, \dots, l\} - \frac{1}{2}\widehat{\pi_0}l\gamma.$$

However, the most crucial part of identifying the genuine TTRs is the method of controlling a predetermined level of *FDR*$^+$ (i.e. 10%) or, in other words, finding the right *p*-value cutoff *γ* above which lie the rules with no statistically significant performance. We achieve this by following Storey et al. (2004), while using point estimates of the FDR. In particular, the *p*-values of the TTRs with positive performance are placed in ascending order. Then, starting with the smallest one, while adding the next *p*-value corresponding to the second rule, the *FDR*$^+$ is recomputed. This procedure is repeated until the desired *FDR*$^+$ is attained.

4.4. The k-FWER method

The second method employed for data snooping bias in our study is the *k*-FWER approach developed by Romano and Wolf (2007) and Romano et al. (2008). The rationale for implementing it in our case is its flexibility in detecting a great number of genuine trading strategies once the strict FWER criterion is eased, making it more suitable for investors who want to identify as many outperforming strategies as possible. However, we also want to examine whether a generalized version of the conservative FWER measure, allowing some false rejections, would achieve the same results as the powerful FDR^{+}. Contrary to its predecessor, the *k*-FWER criterion is defined as

$$k - FWER_p = P\{Reject\ at\ least\ k\ of\ the\ H_{0j}\},$$

which is the probability of rejecting at least k true null hypotheses. In the multiple hypothesis testing setup, under a statistical significance level of α, the k-FWER is controlled if k-FWER$_p \leq \alpha$. The k-*FWER* framework has an analogous structure to that of the *StepM*-BRC technique of Romano and Wolf (2005). However, it allows for at least a small number of false selections to be retained. Moreover, a resampling mechanism also needs to be used. Thus, for comparison purposes, we also employ the stationary bootstrap of Politis and Romano (1994), while using the same procedure to calculate the bootstrapped test statistics and thus the critical values for the BRC, *StepM*-BRC, and FDR tests. Each bootstrap test statistic vector also needs to be centered on its original value.

After the computation of the empirical bootstrapped distribution and the critical values, since the setup of the k-*FWER* approach is similar to that of the *StepM*-BRC test, only a few other steps need to be modified. The more general k-FWER approach needs to satisfy the criterion that at least *k* hypotheses will be rejected, instead of just one. Specifically, the TTR test statistics, φ_j, of the strategies showing positive performance are relabeled in descending order, with the first referring to the largest. During the first stage, individual decisions are executed for each rule, the null hypothesis H_{0j} being rejected if the test statistic, φ_j, is greater than the critical value

$$\widehat{c_1} = c_{\{1,\ldots,l\}}(1-a, k, \widehat{P_T}) \text{ for } 1 \leq j \leq l,$$ [19]

where $\widehat{c_1}$ is the estimated smallest $(1-a)$ quantile of the re-centered sampling distribution of the k^{th} largest rule under the bootstrapped probability measure $\widehat{P_T}$. Then, denote by R_1 the number of statistically significant rules (hypotheses rejected) during the first stage. If $R_1 < k$ the procedure is terminated. This happens because it is feasible that all significant rules are true rejections. On the other hand, if $R_1 > k$ there is a strong possibility that some false rejections will have been included in the total number of rejections. Therefore, we need to move on to the second stage, excluding the test statistics of the rejected strategies. The remaining ones are tested in a new hypothesis testing setup. This time, each of the test statistics, φ_j, is compared with the critical value

$$\widehat{c_2} = \max\{c_K(1-a, k, \widehat{P_T})\} \text{ for } R_1 + 1 \leq j \leq l$$

while individual decisions are also made, $\widehat{c_2}$ depicts the maximum quantile of the set of quantiles, including the rejected $k-1$ hypotheses from the first step, as well as all the hypotheses that have not been rejected yet. The intuition is that we are not certain which of the rejected hypotheses might be true so both rejected and non-rejected hypotheses, together with the largest quantile, must be considered. Finally, if no further hypotheses are rejected in the second step, the procedure terminates. Otherwise, the stepwise setup is maintained, and new decisions are carried out involving new $\widehat{c_m}$ maximum critical values, until no other rejections occur.

The steps described above reflect the one-sided framework, which is meaningful when searching for genuine TTRs among the entire set of rules, which display positive performance and fall within the right tail of the distribution.

[19] The critical value $\widehat{c_1}$ asymptotically controls the k-FWER criterion. According to the theory $c_1 = c_K(1-a, k, P)$. However, the set K and the probability mechanism P are unknown. Therefore, K is replaced by the set of all rules $\{1, \ldots, l\}$ and the probability measure $\widehat{P_T}$ of the bootstrapped distribution is used instead of P.

4.5. Portfolio construction

We construct the portfolios of rules by selecting them in accordance with the *FDR*$^+$ and *k*-FWER. In particular, we set the $\widehat{FDR^+}$ and *k*-FWER equal to 10%, as a good trade-off between truly outperforming TTRs and wrongly chosen ones (Bajgrowicz and Scaillet 2012). Despite the fact that *k* is an integer in the case of *k*-FWER, we adjust it to a number that is equal to 10% of the rules showing positive performance for each interval examined. Thus, we acquire 10%-FDR$^+$ and 10%-FWER portfolios, which means that which means that 90% of the total number of the portfolio's rules, significantly outperform the benchmark. The signals of the chosen rules are pooled with equal weight, similarly to a forecast averaging technique. We do not attribute more weight to more effective rules since this would result in reducing the FDR$^+$ and *k*-FWER portfolios below the desired level. We finally 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.

5. In-sample performance

5.1 In-sample performance with no transaction costs

Table 3 reports the number of TTRs displaying a positive performance[20] under the *mean return, Sharpe* ratio, and *Calmar* ratio criteria, for crude oil futures and USO, Subperiods 1-4.

[Insert Table 3 somewhere here]

Concerning Subperiods 1 (18 April 2007 – 29 May 2009) and 4 (1 August 2013 – 1 January 2016), that contain strong trends or an unstable environment, it seems that a significant

[20] Positive performance means a mean return or Sharpe ratio above zero, or a Calmar ratio above one.

proportion of the TTRs considered are able to achieve a positive performance for both crude oil futures and USO. This outcome could almost have been anticipated since the majority of the strategies are momentum or trend-following rules capturing extreme movements. On the other hand, the number of outperforming rules is reduced for the mean return and Sharpe ratio for Subperiods 2 (1 June 2009 – 31 May 2011) and 3 (1 April 2011 – 31 July 2013), which show a more balanced evolution of prices. Regarding the Calmar ratio, the number of outperforming rules is even less (at most 25%), as a TTR needs to achieve a Calmar ratio above 1 to be considered a good strategy.[21] It is also worth mentioning that, during Subperiod 3, the outperformed TTRs on crude oil futures are almost half of those on the USO for the mean return, the Sharpe and the Calmar ratio. Moreover, the number of outperforming rules identified, based on the Sharpe ratio, is consistently less than or (almost) equal to those identified by the mean return criterion, for crude oil futures and USO. This feature stems from the subtraction of the risk-free rate, which results in the elimination of returns of a very small magnitude.

For the same sample periods, Table 4 shows the in-sample performance of the best rule under the mean return, Sharpe ratio and Calmar ratio criteria respectively, free of transaction costs, for crude oil futures (Panel A) and USO (Panel B). The corresponding *p*-value of the BRC test for the best rule and for each performance criterion is also displayed in parenthesis. The buy-and-hold strategy for the crude oil futures and the USO is also displayed in the columns on the right-hand side of the table.

[Insert Table 4 somewhere here]

For both Panels A and B, the best rule results seem very encouraging compared with the buy-and-hold strategies, for crude oil futures and USO. The best rule's performance indicates

[21] A Calmar ratio value of 1-2 is assumed a good strategy, a value between 2-5 very good, and a value greater than 5 recognized as excellent (Young 1991).

that outperformed trading strategies exist in all subperiods, according to all criteria. Specifically, the Sharpe and Calmar ratios are high enough that the best rules can be characterized as very good trading opportunities in the majority of the years covered. However, we should mention that the evidence provided above has no economic value, since transaction costs are not considered, and it is just a trivial experiment concerning predictability. Moreover, the corresponding BRC *p*-values are quite high in half the cases, indicating that the performance of some of the best rules is not significant. Also, the information reported in Table 4 relies only on findings that are discovered ex post, and there is no guarantee that a trader will have selected the potentially best rule in advance only by looking at its long-term historical behavior. In practice, investors rebalance their positions more frequently to capture any changes in the economic and financial milieu. Despite the above, Tables 3 and 4 still reveal the existence of technical indicators that are able to capture patterns in the daily prices for both crude oil futures and the USO.

5.2 In-sample performance including transaction costs

Since predictive power is not always synonymous with profitability, an investor should always check carefully whether the returns gained from trading strategies are sufficient to cancel out the transaction costs. Indeed, trading rules pooled before transaction costs are more likely to generate frequent signals, thus increasing the probability of their performance benefits being eliminated once the transaction costs are included.

The majority of the previous studies examine the performance of TTRs through a breakeven analysis, wherein the effect of transaction costs is computed ex post, once outperforming rules have been identified. However, this undoubtedly makes it more complicated for a trader to foresee profitable rules that will offset transaction costs a priori. Contrary to that, we again follow Bajgrowicz and Scaillet (2012), handling transaction costs "*endogenously*" and not "*exogenously*" to the selection process. In particular, we subtract the transaction costs every time a buy or sell signal is triggered. Following the study of Locke and Venkatesh (1997), who estimate that futures markets' one-way transaction costs range from 0.04 to 3.3 basis points, we consider the second, larger amount for the crude oil futures. Furthermore, we assume that an investor funds their position with 100% equity rather than using a margin, since we measure daily returns as the log of the difference in price relatives (Bessembinder, 1992; Miffre and Rallis, 2007; Marshall et al., 2008). For the case of the USO, we incorporate one-way transaction costs of 5 basis points on each trade. This level of transaction costs is justified based on the literature, as well as information from floor traders (Hsu et al. 2010), for the trading of ETFs.

Tables 5 and 6 display the number of outperforming rules as well as the in-sample performance for the crude oil futures and the USO when one-way transaction costs of 3.3 and 5 basis points are considered, respectively, ex ante. Comparing the results with Table 3, in Table 5 we can see that the number of corresponding outperforming rules has decreased considerably for all evaluation methods, especially when the Calmar ratio is employed.

[Insert Table 5 somewhere here]

[Insert Table 6 somewhere here]

In Table 6, the in-sample performance with transaction costs provides a similar picture. Thus, as expected, the values for all evaluation criteria are reduced. However, in Table 6, we also observe that the best trading rules are still able to achieve better performance than the buy-and-hold strategy, for all criteria and across all subperiods. Interestingly, compared with the values in Table 4, the Sharpe ratios for the best rules are lower. However, the trading strategies remain very promising, as the Sharpe ratios are still above 1.5. Similarly, in Table 6, the Calmar ratios are high enough that we can conclude that the generated returns are sufficient to outweigh the transaction costs. When it comes to the statistical significance of the best rules' performance, Table 6 demonstrates that none of the *p*-values from the BRC test are significant. However, this does not mean that TTRs with genuinely good performance do not exist among the most profitable rules.

Tables 7 and 8 demonstrate the impact of transaction costs on the historically best TTRs, selected with respect to the mean return, Sharpe ratio and Calmar ratio criteria, while accounting for zero and non-zero one-way transaction costs, for the crude oil futures and the USO, respectively, in each subperiod.

[Insert Table 7 somewhere here]

[Insert Table 8 somewhere here]

Generally speaking, the best trading strategy nominated remains within the same family of rules, before and after one-way transaction costs are considered, in most cases, and for both the crude oil futures and the USO. In addition to this, Table 7 demonstrates that, in the case of crude oil futures, the small magnitude of transaction costs does not have a strong impact on the chosen best rule, and this applies to all criteria and across all subperiods. However, Table 8 portrays a contradictory picture, especially under the mean return and Sharpe ratio criteria, for the case of the USO. TTRs selected without consideration of the transaction costs produce more frequent trading signals than those for which the transaction costs have been taken into account endogenously. For instance, under the Sharpe ratio measure and Subperiod 1, a 25-50 day on-balance volume rule is more likely to trigger frequent signals than a channel breakout rule with a 20-100 window of days, 0.075 channel width and a five-day holding period, which suffers more constraints. The best rules after the inclusion of transaction costs are *not* usually among the best ones before their inclusion. One explanation might be that trading the USO entails larger transaction costs than trading crude oil futures. Moreover, the successful rules do not trade on longer-term price movements once transaction costs are incorporated, in either case. While there are cases (under the Sharpe ratio and in Subperiod 2 for the USO) where the best rule, in-sample, uses a larger window of 200 days of data when transaction costs are included, compared to a window of 10 days used by the best rule under zero transaction costs, this is not true in most of the cases.

Another interesting finding emerges when employing the Calmar ratio criterion. The best rules derived before and after the inclusion of transaction costs are closely related, perhaps due to the maximum drawdown factor employed. Searching for the best strategy, while minimizing the maximum drawdown of its returns, increases an investor's probability of ending up with a rule that generates less frequent signals, even before the inclusion of transaction costs to avoid larger drawdowns. This might be the reason for TTRs selected under the Calmar ratio approach being almost the same before and after consideration of transaction costs in each sample period.

Finally, the most important evidence gleaned from observing both tables together, is that the rules selected as the best ones based on technical analysis for the crude oil futures belong to a different family from those selected when trading the USO, under all criteria and across all subperiods considered. This may be a potential justification for the different dynamics that characterize the crude oil futures compared to the USO, with contango or backwardation outcomes having a significant effect on the calculation and redemption procedures for ETFs. On the other hand, the above findings may just reveal the considerable effects of data snooping bias, in that the best rule's performance may be achieved merely through luck in most cases, leading to different rules being identified as the best for the two assets and for the different subperiods.

6. Persistence analysis

The economic evaluation of TTRs' performance in the crude oil market is covered in this section. One of the fundamental questions that technical traders must answer when evaluating TTRs' predictive power is whether the rules selected as superior *ex ante* during backtesting are also able to generate abnormal returns once the transaction costs are considered, for an out-of-sample period. We shed light on whether some of the outperforming rules would have been able to produce profits in practice due to the high volatility in the crude oil market, using only past price data.

A persistence (out-of-sample) analysis of TTRs' performance in the crude oil market is applied here for the very first time. With this aim, we build portfolios of outperforming rules, and re-evaluate the portfolios' performance on a semi-annual basis. In the first six months, when the total universe of rules' performance is tested, we construct equally weighted portfolios while accounting for data snooping bias using the FDR^+ and *k*-FWER methods as described in Sections 4.2 - 4.5. Specifically, every six months, two portfolios are constructed employing price data from the previous six months. Then, the out-of-sample performance of the chosen rules is evaluated over the following half of the year. As mentioned earlier, the "risk-free" rate is considered as the benchmark. In particular, we assess the performance of rules in-sample (IS), before constructing portfolios of the "genuine" (statistically significant) in-sample rules and measuring their performance out-of-sample (OOS). This structure matches how investors in practice set up their own strategies based only on a priori information.

Table 9 displays the results of the persistence analysis under the annualized Sharpe ratio and 3.3 and 5 basis points of transaction costs for the crude oil futures and the USO respectively, in accordance with the 10%-FDR^+ and 10%-FWER rules selection criteria, as well as the best rule's performance, observed across the different in-sample and out-of-sample periods. The table also includes each portfolio's median size as well as its percentage amount over the total universe of TTRs in brackets.[22]

[Insert Table 9 somewhere here]

The results clearly indicate that there is *no* persistence in the trading rules' performance, as both selection criteria verify. However, in periods when both portfolios are able to generate positive performance out-of-sample, the Sharpe ratio levels are considerably smaller than those of the in-sample performance. For instance, no investor will choose a portfolio whose Sharpe ratio level is below one. However, the only case of a Sharpe ratio exceeding one is in Subperiod 2 for the trading of crude oil futures contracts, for both portfolios. Overall, the picture is opposed to the evidence found regarding the in-sample performance in Section 5, which implies that the best-performing rules are accessible *only* to investors observing the returns ex post. Additionally, comparing the 10%-FWER and 10%-FDR$^+$ portfolios with the best rule's performance, we notice that, in most cases, both portfolios achieve better performance out-of-sample than just employing the best rule, verifying the benefits of employing the proposed data snooping methods as portfolio construction techniques. To summarize, interestingly, the *no hot hands* phenomenon that is confirmed in Bajgrowicz and Scaillet (2012) for the DJIA also appears in the crude oil market.

As presented in Table 9, one of the most important findings is the overall performance of the 10%-FWER and 10%-FDR$^+$ specifications for data snooping as portfolio compilers. We could say that the two approaches don't seem able to achieve attractive Sharpe ratios (i.e., above 1) for both the crude oil futures and USO in most cases apart from the Subperiod 2 when trading

[22] The relevant table and results of the persistence analysis considering the subperiods given by the nonparametric change point detection approach of Ross et al, (2011) are presented in Appendix A.

the crude oil futures. The generalization of the FWER measure (to allow for some false selections) improves its performance, which is demonstrated by the achievement of more or less similar Sharpe ratios to the FDR$^+$ portfolio. The *k*-FWER circumvents the lucky rules that do not produce significantly good performance, while keeping only the genuine ones. Furthermore, it does not suffer from preventing the selection of further rules once it has identified a lucky one, as its predecessor did. We also note that the number of the rules selected by the two methods (i.e., *Median size*) varies significantly. For example, for the 10%-FWER approach this ranges from 24 to 167 rules, while for the 10%-FDR$^+$ approach this ranges from 37 to 300 rules.

Finally, Table 10 presents both the 10%-FWER and 10%-FDR$^+$ portfolios' average decomposition according to each of the five families of TTRs for the crude oil futures and the USO respectively. In particular, we report the number of rules selected from each family divided by the total number of rules included in each portfolio, as a percentage.[23]

[Insert Table 10 somewhere here]

We observe that the 10%-FWER and 10%-FDR$^+$ portfolios show different selection preferences among the five families of rules. Furthermore, the preferences seem to differ depending on whether the crude oil futures or the USO are being traded. For instance, the 10%-FWER portfolio seems to mostly choose TTRs from the on-balance volume and channel breakout families, followed by the support and resistance rules, while the 10%-FDR$^+$ portfolio selects TTRs mostly from the support and resistance and channel breakout families, with the on-balance volume rules coming next in order of preference. The filter rules family displays the smallest percentages for both portfolios, crude oil futures and USO, as well as across all subperiods. In

[23] We should mention that the number of rules chosen varies substantially from one six-month period to the next. Sometimes, the portfolio consists of almost exclusively new rules, even after the first rebalancing.

general, the support and resistance, channel breakouts and on-balance volume rule families are the most significant in capturing the patterns of the crude oil market.

7. Conclusion

Evidence of the historical success of technical trading rules is revisited, this time in the trending crude oil market. Although, numerous efforts have been made in the field of evaluating the performance of technical trading rules (forex, stock, large/small cap markets, etc.), this is the first time it has been done for crude oil. Findings from previous studies are divided on whether technical analysis can achieve genuine abnormal performance. The motivation of this study was to examine whether technical trading indicators and oscillators could benefit from the severe fluctuations characterizing the crude oil market lately. The majority of these rules are designed to capture such patterns, being momentum and trend following strategies. We focus on the crude oil futures and the United States Oil fund (USO) as the largest and most liquid crude oil exchange traded fund (ETFs), developed to track the daily price movements of West Texas Intermediate ("WTI") light, sweet crude oil.

First, we reassess the predictive power of Sullivan et al. (1999)'s universe of trading rules in the overall crude oil market, to verify that patterns exist. Evidence of the rules' performance on crude oil futures as well as the USO in an in-sample simulation demonstrates that, during periods of dramatic crude oil price movements, more than half of the rules show great predictive power. However, the corresponding *p*-values of the best rules of the *bootstrap reality check* (BRC) test of White (2000) are not statistically significant most of the time. Popular performance measures used by fund managers and traders, such as the Sharpe and Calmar ratios, are employed to measure the profitability of rules. All of them show the best rule for each period to be a very good trading opportunity.

Second, we endogenously incorporate transaction costs when evaluating trading rules' performance in the case of crude oil futures contracts and the USO. Strategies employed by Brock et al. (1992) and Sullivan et al. (1999) can trigger very frequent signals, which might lead to the elimination of superior returns when transaction costs are considered. However, technical trading rules are still able to achieve profits, although their performance is decreased, because of the relatively small transaction costs applied when trading commodities futures and ETFs.

Third, we employ two of the most powerful techniques for accounting for data snooping, in order to identify significantly profitable trading strategies. The *false discovery rate* (FDR) approach, as described by Bajgrowicz and Scaillet (2012) when evaluating technical trading rules on the DJIA index, is used to control false discoveries. The *k-familywise error rate* (*k*-FWER) methodology developed by Romano and Wolf (2007) is also applied to check whether a generalized version of the conservative FWER criterion allowing for some false rejections performs equally well to the powerful FDR method. Both specifications are able to select more rules and better diversify against model uncertainty than the previous BRC and Romano and Wolf approaches that are prevented from searching for more rules once a "lucky" one has been detected.

Finally, a persistence analysis is carried out for the purpose of economically evaluating the rules' performance. The question that needs to be answered here is whether investors can foresee which rules will generate future returns – that will outweigh transaction costs – without prior knowledge. We respond to this argument by creating portfolios with the FDR and *k*-FWER approaches, using only past data in an in-sample period, and evaluating their performance out-of-sample. The findings show that there is no persistent nature to the rules' performance, contrary to the outstanding in-sample results, although tiny profits can be achieved in some periods. The results seem to be in favor of some short-term anomalies of efficient market hypothesis or an extreme tail risk episode of returns Moreover, the FDR$^+$ and *k*-FWER approaches show almost equal performance.

References

Alexander, S., 1961. Price movements in speculative markets: Trends or random walks: Industrial Management Review 2, 7-26.

Alexander, S., 1964. Price movements in speculative markets: Trends or random walks, no 2. The Random Character of Stock Market Prices (MIT Press, Cambridge, Mass.).

Allen, F., & Karjalainen, R., 1999. Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51(2), 245–271. doi: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(3), 473–491. doi: 10.1016/j.jfineco.2012.06.001

Bajgrowicz, P., Scaillet, O., & Treccani, A., 2016. Jumps in high-frequency data: spurious detections, dynamics and news, Management Science, 62 (8) 2198-2217. https://doi.org/10.1287/mnsc.2016.2568

Barras, L., Scaillet, O., Wermers, R., Wermers, R., & Stulz, R., 2015. False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. Journal of Finance, 65(1), 179–216. doi: 10.1111/j.1540-6261.2009.01527.x

Benjamini, Y., & Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57(1), 289–300.

Benjamini, Y., & Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29(4), 1165–1188. doi:10.1214/aos/1013699998

Bessembinder, H., & Chan, K., 1998. Market efficiency and the returns to technical analysis. Financial Management 27(2), 5–17.

Brock, W., Lakonishok, J., & LeBaron, B., 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. Journal of Finance, 47(5), 1731-1764. doi: 10.1111/j.1540-6261.1992.tb04681.x

Conrad, J., & Kaul, G., 1998. An anatomy of trading strategies. Review of Financial Studies, 11(3), 489–519. doi:10.1093/rfs/11.3.489

Fama E., & Marshall, B., 1966. Filter Rules and Stock-Market Trading. Journal of Business 39(1), 226–241.

Fama, E., 1965. The Behavior of Stock Market Prices. Journal of Business, 38, 34–105.

Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, 25(2), 383–417. doi: 10.1111/j.1540-6261.1970.tb00518.x

Hansen, P. R., 2005. A Test for Superior Predictive Ability. Journal of Business & Economic Statistics, 23(4), 365–380. doi:10.1198/073500105000000063

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The Model Confidence Set. Econometrica, 79(2), 453–497. doi: 10.3982/ECTA5771

Harvey, C. R., & Liu, Y., 2014. Evaluating Trading Strategies. Journal of Portfolio Management, 40(5), 108-118. doi: 10.3905/jpm.2014.40.5.108

Hedge, S. P., & McDermott, J. B., 2004. The market liquidity of DIAMONDS, Q's, and their underlying stocks. Journal of Banking and Finance, 28(5), 1043–1067. doi:10.1016/S0378-4266(03)00043-8

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6(2), 65-70.

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(3), 471–484. doi:10.1016/j.jempfin.2010.01.001

Hsu, Y., Kuan, C., & Yen, M., 2014. A Generalized Stepwise Procedure with Improved Power for Multiple Inequalities Testing. Journal of Financial Econometrics, 12(4), 730–755. doi:10.1093/jjfinec/nbu014

James, F., 1968. Monthly Moving Averages—An Effective Investment Tool. Journal of Financial and Quantitative Analysis, 3(3), 315–326. doi: 10.2307/2329816

Jensen, M., & Benington, G., 1970. Random Walks and Technical Theories: Some Additional Evidence. Journal of Finance, 25(2), 469-482. doi: 10.1111/j.1540-6261.1970.tb00671.x

Jondeau, E., & Rockinger, M., 2006. Optimal portfolio allocation under higher moments. European Financial Management 12(1), 29-95. doi: 10.1111/j.1354-7798.2006.00309.x

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

Keynes, J. M., 1936. The General Theory of Employment, Interest, and Money. Harcourt, Brace & Co., New York.

Levy, R., 1967. Relative Strength as a Criterion for Investment Selection. Journal of Finance, 22(4), 595–610. doi: 10.1111/j.1540-6261.1967.tb00295.x

Ljung, G.M. & Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303. doi: 10.1093/biomet/65.2.297

Lo, A. W., & MacKinley, 1990. Data snooping biases in tests of financial asset pricing models. Review of Financial Studies, 3(3), 431–467. doi:10.1093/rfs/3.3.431

Lo, A. W., Mamaysky, H., & Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. Journal of Finance 55(4), 1705-1765. doi:10.1111/0022-1082.00265

Malkiel, B., 1981. A Random Walk Down Wall Street. Norton New York, 2ed.

Marshall, B. R., Cahan, R. H., & Cahan, J. M., 2008. Does intraday technical analysis in the U.S. equity market have value? Journal of Empirical Finance, 15(2), 199–210. doi:10.1016/j.jempfin.2006.05.003

Marshall, B. R., Cahan, R. H., & Cahan, J. M., 2008. Can commodity futures be profitably traded with quantitative market timing strategies? Journal of Banking and Finance, 32(9), 1810-1819. doi:10.1016/j.jbankfin.2007.12.011

Neely, C. J., & Weller, P., 2011. Technical Analysis in the Foreign Exchange Market. Working paper 2011-001B, Federal Reserve Bank of St. Louis.

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(4), 405-426. doi: 10.2307/2331231

Neftci, S. N., 1991. Naive Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of “Technical Analysis.” Journal of Business, 64(4), 549-571. doi:10.1086/296551

Politis, D., & Romano, J., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Qi, M., & Yangru, W., 2006. Technical Trading-Rule Profitability, Data Snooping , and Reality Check : Evidence from the Foreign Exchange Market. Journal of Money, Credit, and Banking, 38(8), 2135–2158. Do: 10.1353/mcb.2007.0006

Ready, M. J., 2002. Profits from Technical Trading Rules. Financial Management 31, 43-61.

Romano, J. P., & Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica, 73(4), 1237–1282. doi:10.1111/j.1468-0262.2005.00615.x

Romano, J. P., & Wolf, M., 2007. Control of generalized error rates in multiple testing. Annals of Statistics, 35(4), 1378–1408. doi:10.1214/009053606000001622

Romano, J. P., Shaikh, A. M., & Wolf, M., 2008. Formalized Data Snooping Based on Generalized Error Rates. Econometric Theory, 24(2), 404–447. doi:10.1017/S0266466608080171

Ross G. J., Tasoulis D. K., Adams N.M., 2011. Nonparametric Monitoring of Data Streams for Changes in Location and Scale. Technometrics, 53(4), 379-389.

Schuhmacher, F. & Eling, M., 2011. Sufficient conditions for expected utility to imply drawdown-based performance rankings. Journal of Banking & Finance, 35(9), 2311-2318. doi:10.1016/j.jbankfin.2011.01.031

Shynkevich, A., 2012. Performance of technical analysis in growth and small cap segments of the US equity market. Journal of Banking and Finance, 36(1), 193–208. doi:10.1016/j.jbankfin.2011.07.001

Storey, J., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B 64(3), 479–498. doi 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(1), 187–205. doi: 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(5), 1647–1691. doi:10.1111/0022-1082.00163

Sweeney, R., 1988. Some New Filter Rule Tests: Methods and Results. Journal of Financial and Quantitative Analysis, 23(3), 285–300. doi: 10.2307/2331068

Taylor, N., 2014. The rise and fall of technical trading rule success. Journal of Banking and Finance, 40, 286-302. doi: 10.1016/j.jbankfin.2013.12.004

Wang, C., & Yu, M., 2004. Trading activity and price reversals in futures markets. Journal of Banking and Finance, 28(6), 1337-1361. doi:10.1016/S0378-4266(03)00120-1

White, H., 2000. A Reality Check for Data Snooping. Econometrica, 68(5), 1097–1126. doi:10.1111/1468-0262.00152

Yamamoto, R., 2012. Intraday technical analysis of individual stocks on the Tokyo Stock Exchange. Journal of Banking and Finance, 36(11), 3033–3047. doi: 10.1016/j.jbankfin.2012.07.006

Table 1. Sample periods for crude oil futures and USO

Sample period	Dates	Trading days	Market trend
Subperiod 1:	18 April 2007- 29 May 2009	534	Mixed
Subperiod 2:	1 June 2009- 31 March 2011	464	Bullish
Subperiod 3:	1 April 2011- 31 July 2013	586	Mixed
Subperiod 4:	1 August 2013- 1 January 2016	618	Bearish

Table 2. Descriptive statistics of daily returns on crude oil futures and USO

Instrument	Subperiod	Mean (%)	St.dev. (%)	Skewness	Kurtosis	First AC
Crude oil futures	**Subperiod 1**	0.01	3.03	0.44	11.4	0.01
	Subperiod 2	0.10	1.59	-0.01	4.55	0.18**
	Subperiod 3	0.01	1.42	-0.35	4.88	0.08
	Subperiod 4	-0.19	1.77	0.66	5.97	0.12**
USO	**Subperiod 1**	-0.06	2.91	-0.19	4.28	-0.11**
	Subperiod 2	0.04	1.92	-0.07	3.23	0.00
	Subperiod 3	-0.02	1.75	-0.57	6.24	-0.06
	Subperiod 4	-0.22	1.99	0.00	5.36	-0.10**

This table reports the descriptive statistics of daily returns on the crude oil futures and the USO. Means and standard deviations are reported in percentage points (%). “First AC” stands for first-order autocorrelation. Asterisks (**) denote significant first-order autocorrelation for returns at the 5% level according to the Ljung-Box (1978) Q statistics.

Table 3. Numbers of outperforming rules for in-sample performance with no transaction costs

Instrument	**Sample period**	**Outperforming rules**		
		Mean return	**Sharpe ratio**	**Calmar ratio**
Crude oil futures	**Subperiod 1**	4049 [51%]	3979 [50%]	574 [7%]
	Subperiod 2	2226 [28%]	2222 [28%]	406 [5%]
	Subperiod 3	840 [10%]	838 [10%]	82 [1%]
	Subperiod 4	5148 [65%]	5147 [65%]	485 [6%]
USO	**Subperiod 1**	4114 [52%]	4096 [52%]	2025 [25%]
	Subperiod 2	1774 [22%]	1774 [22%]	341 [4%]
	Subperiod 3	1920 [24%]	1919 [24%]	166 [2%]
	Subperiod 4	5000 [63%]	5000 [63%]	1445 [18%]

This table presents the outperforming rules in levels, and as percentages over the total universe in brackets, identified according to the positive daily mean return, annualized Sharpe ratio, and Calmar ratio criteria respectively, for the crude oil futures and the USO, across the different subperiods.

Table 4. In-sample performance with no transaction costs

Sample period	**Best rule**			**Buy-and-hold strategy**		
Panel A Crude oil futures	**Mean return (%)**	**Sharpe ratio**	**Calmar ratio**	**Mean return (%)**	**Sharpe ratio**	**Calmar ratio**
Subperiod 1	0.34	1.82	8.21	0.01	0.30	0.25
	(0.09)*	(0.05)**	(0.03)**			
Subperiod 2	0.14	2.30	7.70	0.10	0.95	1.45
	(0.99)	(0.39)	(0.83)			
Subperiod 3	0.12	1.71	8.83	0.01	0.09	0.07
	(0.97)	(0.80)	(0.43)			
Subperiod 4	0.26	2.10	9.80	-0.19	-1.3	-0.62
	(0.08)*	(0.06)*	(0.10)*			
	Best rule			**Buy-and-hold strategy**		
Panel B USO	**Mean return (%)**	**Sharpe ratio**	**Calmar ratio**	**Mean return (%)**	**Sharpe ratio**	**Calmar ratio**
Subperiod 1	0.38	2.05	6.95	-0.06	-0.13	-0.04
	(0.09)*	(0.07)*	(0.11)			
Subperiod 2	0.17	2.07	8.92	0.04	0.37	0.54
	(0.96)	(0.71)	(0.69)			
Subperiod 3	0.13	1.78	7.31	-0.02	-0.09	-0.05
	(0.98)	(0.74)	(0.23)			
Subperiod 4	0.23	2.19	9.35	-0.22	-1.69	-0.69
	(0.08)*	(0.07)*	(0.06)*			

This table reports the performance results of the best rule, and its corresponding BRC *p*-value in-sample in parenthesis, as well as the buy-and-hold strategy for the crude oil futures and the USO respectively, under the daily mean return, annualized Sharpe ratio, and Calmar ratio criteria, and across the different subperiods. * denotes a rejection of the null hypothesis at the 10% level of significance, ** denotes rejection at the 5% level.

Table 5. Numbers of outperforming rules for in-sample performance with transaction costs

Instrument/ Transaction costs	**Sample period**	**Outperforming rules**		
Crude oil futures (3.3 bps)		**Mean return**	**Sharpe ratio**	**Calmar ratio**
	Subperiod 1	3623 [46%]	3558 [45%]	320 [4%]
	Subperiod 2	1555 [19%]	1559 [19%]	274 [3%]
	Subperiod 3	525 [6%]	525 [6%]	69 [0.8%]
	Subperiod 4	4569 [58%]	4565 [58%]	250 [3%]
USO (5 bps)				
	Subperiod 1	3777 [48%]	3739 [47%]	1341 [17%]
	Subperiod 2	1066 [13%]	1065 [13%]	168 [2%]
	Subperiod 3	931 [11%]	929 [11%]	79 [1%]
	Subperiod 4	4346 [55%]	4346 [55%]	363 [4%]

This table presents the outperforming rules in levels, and in percentages over the total universe in brackets, identified according to the positive daily mean return, annualized Sharpe ratio, and Calmar ratio criteria, for the crude oil futures and the USO, assuming 3.3 and 5 basis points (bps) one-way transaction costs respectively, across the different subperiods.

Table 6. In-sample performance with transaction costs

Sample period	**Best rule**			**Buy-and-hold strategy**		
Panel A	**Mean**	**Sharpe**	**Calmar**	**Mean**	**Sharpe**	**Calmar**
Crude oil futures	**return (%)**	**ratio**	**ratio**	**return (%)**	**ratio**	**ratio**
Subperiod 1	0.30	1.73	7.76	0.01	0.30	0.25
	(0.25)	(0.15)	(0.29)			
Subperiod 2	0.11	2.21	7.13	0.10	0.95	1.45
	(1.00)	(0.49)	(0.82)			
Subperiod 3	0.08	1.69	7.80	0.01	0.09	0.07
	(0.99)	(0.81)	(0.42)			
Subperiod 4	0.23	1.89	8.53	-0.19	-1.3	-0.62
	(0.31)	(0.27)	(0.35)			
	Best rule			**Buy-and-hold strategy**		
Panel B	**Mean**	**Sharpe**	**Calmar**	**Mean**	**Sharpe**	**Calmar**
USO	**return (%)**	**ratio**	**ratio**	**return (%)**	**ratio**	**ratio**
Subperiod 1	0.33	1.87	6.00	-0.06	-0.13	-0.04
	(0.44)	(0.68)	(0.39)			
Subperiod 2	0.13	1.89	6.05	0.04	0.37	0.54
	(1.00)	(0.88)	(0.74)			
Subperiod 3	0.08	1.60	5.11	-0.02	-0.09	-0.05
	(1.00)	(0.91)	(0.53)			
Subperiod 4	0.18	2.09	8.67	-0.22	-1.69	-0.69
	(0.33)	(0.22)	(0.28)			

This table reports the performance results of the best rule, and its corresponding BRC p-value in-sample in parenthesis, including one-way transaction costs for the crude oil futures and the USO, as well as the buy-and-hold strategies for the crude oil futures and the USO respectively, under the daily mean return, annualized Sharpe ratio, and Calmar ratio criteria, across the different subperiods.

Table 7. Best in-sample technical trading rules for crude oil futures

Sample period	Costs	Best rule
Mean return		
Subperiod 1	Zero	Sup.&Res. (10-day alt. extrema, 0.05 band)
	3 bps	Sup.&Res. (20-day alt. extrema, 0.05 band)
Subperiod 2	Zero	Filter (0.01 pos. initiation, 5-day hold. per.)
	3bps	Filter (0.01 pos. initiation, 5-day hold. per.)
Subperiod 3	Zero	Moving Avg. (25-30 day, 25-day hold. per.)
	3 bps	Moving Avg. (25-30 day, 25-day hold. per.)
Subperiod 4	Zero	On-Bal.-Vol. (40-50 day, 10-day hold. per.)
	3 bps	On-Bal.-Vol. (40-50 day, 10-day hold. per.)
Sharpe ratio		
Subperiod 1	Zero	Chan.Br. (20-day, 0.075 width, 10-day hold. per.)
	3 bps	Chan.Br. (20-day, 0.075 width, 0.01 band, 10-day hold. per.)
Subperiod 2	Zero	Sup.&Res. (10-day, 0.1 band)
	3 bps	Sup.&Res. (20-day, 0.03 band)
Subperiod 3	Zero	Chan.Br. (15-day, 0.05 width, 0.03 band 10-day hold. per.)
	3 bps	Chan.Br. (15-day, 0.05 width, 0.03 band 10-day hold. per.)
Subperiod 4	Zero	Chan. Br. (10-day, 0.05 width, 0.01 band 25-day hold. per.)
	3 bps	Chan. Br. (10-day, 0.03 width, 0.001 band 25-day hold. per.)
Calmar ratio		
Subperiod 1	Zero	Sup.&Res. (4-day al. extrema, 4-day delay, 5-day hold. per.)
	3 bps	Sup.&Res. (4-day alt. extrema 4-day delay, 25-day hold. per.)
Subperiod 2	Zero	Sup.&Res. (200-day, 0.03 band, 5-day hold. per.)
	3 bps	Sup.&Res. (200-day, 0.03 band, 5-day hold. per.)
Subperiod 3	Zero	Chan.Br. (100-day, 0.1 width, 0.001 band, 50-day hold. per.)
	3 bps	Chan.Br. (100-day, 0.15 width, 0.01 band, 5-day hold. per.)
Subperiod 4	Zero	Sup.&Res. (150-day, 2-day delay, 25-day hold. per.)
	3 bps	Sup.&Res. (200-day, 2-day delay, 50-day hold. per.)

This table presents the historically best-performing trading rule, chosen under the daily mean, annualized Sharpe ratio, and Calmar ratio criteria, for the crude oil futures, in each sample period and for either zero or 3.3 basis points (bps) one-way transaction costs.

Table 8. Best in-sample technical trading rules for USO

Sample period	Costs	Best rule
Mean return		
Subperiod 1	Zero	Moving Avg. (5-25 day, 50-day hold. per.)
	5 bps	Moving Avg. (1-250 day, 50-day hold. per.)
Subperiod 2	Zero	On-Bal.-Vol. (10-50 day, 0.001 band)
	5 bps	On-Bal.-Vol. (5-50 day, 0.003 band)
Subperiod 3	Zero	On-Bal.-Vol. (15-200 day, 0.005 band)
	5 bps	On-Bal.-Vol. (50-125 day, 0.005 band)
Subperiod 4	Zero	Moving Avg. (1-25 day, 10-day hold. per.)
	5 bps	Moving Avg. (1-25 day, 10-day hold. per.)
Sharpe ratio		
Subperiod 1	Zero	On-Bal.-Vol. (25-50 day)
	5 bps	Chan.Br. (20-100 day, 0.075 width, 5-day hold. per.)
Subperiod 2	Zero	Sup.&Res. (10-day, 0.01 band)
	5 bps	Sup.&Res. (200-day, 0.04 band, 10-day hold. per.)
Subperiod 3	Zero	On-Bal.-Vol. (75-100 day, 0.04 band)
	5 bps	Chan.Br. (25-day, 0.03 width, 0.05 band 10-day hold. per.)
Subperiod 4	Zero	Chan.Br. (10-day, 0.05 width, 0.1 band 25-day hold. per.)
	5 bps	Chan.Br. (10-day, 0.05 width, 0.1 band 25-day hold. per.)
Calmar ratio		
Subperiod 1	Zero	Chan.Br. (15-day, 0.075 width, 5-day hold. per.)
	5 bps	Chan.Br. (50-day, 0.1 width, 0.005 band 5-day hold. per.)
Subperiod 2	Zero	Sup.&Res. (5-day, 3-day delay, 50-day hold. per.)
	5 bps	Sup.&Res. (5-day, 3-day delay, 50-day hold. per.)
Subperiod 3	Zero	Chan.Br. (25-day, 0.03 width, 0.05 band 10-day hold. per.)
	5 bps	Sup.&Res. (20-day, 4-day delay, 5-day hold. per.)
Subperiod 4	Zero	Chan.Br. (10-day, 0.01 width, 5-day hold. per.)
	5 bps	Chan.Br. (10-day, 0.01 width, 5-day hold. per.)

This table presents the historically best-performing trading rule, chosen under the daily mean, annualized Sharpe ratio, and Calmar ratio criteria, for the USO, in each sample period and for either zero or 5 basis points (bps) one-way transaction costs.

Table 9. Performance persistence analysis under the Sharpe ratio criterion with transaction costs and risk-free rate as benchmark

Sample period	**10%-FWER portfolio**			**10%-FDR$^+$ portfolio**			**Best rule**	
Crude oil futures	***Median size***	**IS**	**OOS**	***Median size***	**IS**	**OOS**	**IS**	**OOS**
Subperiod 1	47 [0.6%]	3.37	0.19	50 [0.6%]	2.71	-0.32	3.64	-0.64
Subperiod 2	26 [0.3%]	4.32	1.06	37 [0.4%]	4.44	1.01	3.46	0.68
Subperiod 3	24 [0.3%]	3.27	-1.13	40 [0.4%]	3.44	-0.62	3.05	-0.72
Subperiod 4	25 [0.3%]	3.20	-0.24	112 [1.4%]	3.98	-0.48	3.72	-1.20
USO	***Median size***	**IS**	**OOS**	***Median size***	**IS**	**OOS**	**IS**	**OOS**
Subperiod 1	123 [1.5%]	2.82	0.28	170 [2.1%]	2.71	0.37	3.64	-0.64
Subperiod 2	28 [0.3%]	4.19	-0.15	73 [0.9%]	3.98	-0.29	3.46	0.68
Subperiod 3	42 [0.5%]	3.30	-1.27	70 [0.9%]	3.00	-0.42	3.05	-0.72
Subperiod 4	167 [2.1%]	3.37	-0.26	300 [3.8%]	3.66	0.23	3.72	-1.20

This table indicates the in-sample (IS) and out-of-sample (OOS) annualized Sharpe ratio of trading rules chosen according to the 10%-FWER portfolio of Romano and Wolf (2007), and the 10%-FDR$^+$ portfolio of Bajrowicz and Scaillet (2012), with a semi-annual rebalancing and a risk-free rate as benchmark, as well as the best rule in-sample. The table also displays the portfolios' median sizes in levels, and in percentages of the total rules universe in brackets, across different subperiods.

Table 10. Portfolio decomposition into families of rules

Sample period	F	MA	SR	CB	OBV
Crude oil futures					
Subperiod 1	2% (1%)	18% (28%)	5% (22%)	30% (8%)	45% (39%)
Subperiod 2	1% (2%)	0% (3%)	56% (64%)	40% (31%)	3% (0%)
Subperiod 3	6% (0%)	8% (3%)	2% (66%)	26% (25%)	58% (6%)
Subperiod 4	4% (1%)	16% (16%)	15% (49%)	43% (16%)	21% (18%)
USO					
Subperiod 1	13% (1%)	34% (43%)	10% (24%)	21% (14%)	22% (17%)
Subperiod 2	2% (2%)	1% (3%)	64% (81%)	9% (11%)	22% (2%)
Subperiod 3	11% (0%)	17% (2%)	5% (95%)	25% (0%)	42% (2%)
Subperiod 4	1% (1%)	15% (16%)	11% (26%)	29% (31%)	42% (25%)

This table reports the average percentage of rules belonging to each family, in each portfolio constructed using the 10%-FWER (10%-FDR) methods, for the crude oil futures and the USO respectively, across each subperiod. F: filter rules, MA: moving averages, SR: support and resistance rules, CB: channel breakouts, and OBV: on-balance volume averages.

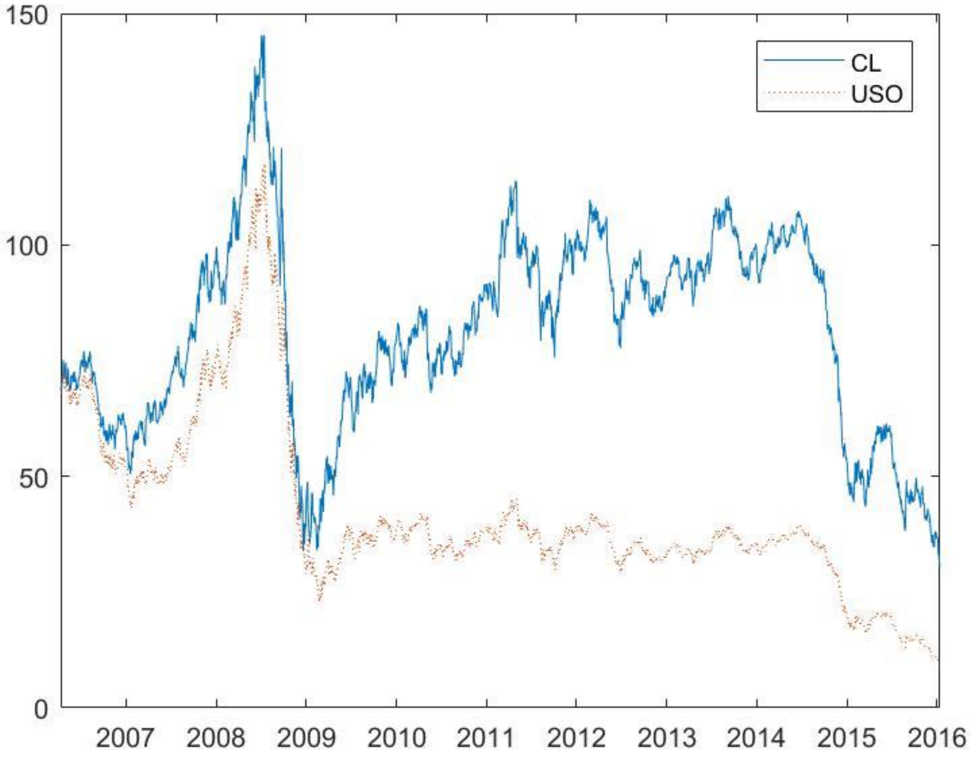

Fig.1. The time series dynamics of crude oil futures (CL) and United States oil fund (USO)

Appendix A

In the Appendix, we report the subperiod intervals selection and the corresponding OOS persistence analysis as a result of utilizing the nonparametric change point detection approach of Ross et al, (2011), for robustness purposes. Their method introduces a streaming change detection algorithm in time series (location and scale changes), especially when the distributional form of the stream variables is unknown, under hypothesis tests involving ranking data points, in which observations are received and processed sequentially over time. Table A1 displays the range of each subperiod defined by the above method, while Fig. A1 presents the graphical representation of the time series dynamics for the two underlying instruments along with the location and scale changing points (i.e., vertical dotted lines) examined from April 2007 to January 2016.

[Insert Figure A1 somewhere here]

[Insert Table A1 somewhere here]

In general, the full sample dataset is separated again in four subperiods. Subperiod 1 is characterized by a sharp increase (bullish), while Subperiod 2 depicts the subsequent fall in crude oil prices due to the global financial crisis (bearish). In fact, Subperiod 1 defined in Section 3 has now divided in in two subperiods with different trends. Subperiod 3 has now significantly increased almost to double but no clear trends appear during that period, which is characterized by mixed movements. Finally, even though Subperiod 4 might be of a smaller size now, it still represents the extreme recent fall of the crude oil market. We could say that the change detection method of Ross et al, (2011) makes no substantial changes to the periods initially defined based on historical events but it slightly adjusts them in order to reveal only major movements (bullish/bearish), leaving only one bigger period characterized by mixed trends.

Similar to Section 6, Table A.2 reports the corresponding evidence of the persistence analysis under the annualized Sharpe ratio criterion and based on the 10%-FDR$^+$ and 10%-FWER portfolios of significant rules, realized across the new in-sample and out-of-sample periods as explained above.

[Insert Table A2 somewhere]

Likewise, Section 6, the results obtained reveal once again limited outperformance in terms of Sharpe ratios. We observe almost equal trends with those of Table 9 in both periods of negative and positive performance, sometimes of a different magnitude though. For example, Subperiod 2 still remains a promising period for exploiting technical analysis in crude oil market, this time not only when trading the crude oil futures but also the USO. A major exception appears in Subperiod 1 when investing on the futures contracts, in which the 10%-FWER portfolio achieves a very healthy Sharpe ratio of 1.69, while the 10%-FDR$^+$ portfolio faces loses, which provide a Sharpe ratio of -1.77. Again, the power of both methods to allow for some false selections doesn't lead to "empty" portfolios. However, median size of portfolios constructed under the two approaches has decreased on average across all subperiods compared to those of Table 9. We should also mention that the median size between the two selection criteria varies considerably here as well. For example, for the 10%-FWER method this spans from 1 to 92 rules, while for the 10%-FDR$^+$ method this spans from 3 to 85 rules.

Table A1. Sample periods for crude oil futures and USO under the nonparametric change point detection model of Ross et al, (2011)

Sample period	Dates	Trading days	Market trend
Subperiod 1:	18 April 2007 - 11 July 2008	311	Bullish
Subperiod 2:	14 July 2008 - 12 February 2009	151	Bearish
Subperiod 3:	13 February 2009 - 26 September 2014	1415	Mixed
Subperiod 4:	27 September 2014 - 1 January 2016	325	Bearish

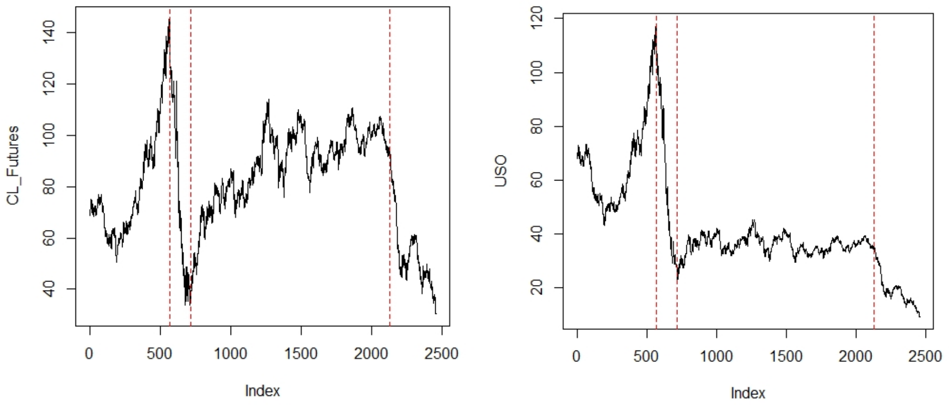

Fig.1. Time series nonparametric change detection points of crude oil futures (CL) and United States oil fund (USO) under the Ross et al, (2011) method.

Table A2. Performance persistence analysis under the Sharpe ratio criterion based on sequential change detection periods.

Sample period	**10%-FWER portfolio**			**10%-FDR$^+$ portfolio**			**Best rule**	
Crude oil futures	***Median size***	**IS**	**OOS**	***Median size***	**IS**	**OOS**	**IS**	**OOS**
Subperiod 1	12 [0.1%]	3.89	1.69	6 [0.05%]	2.79	-1.77	2.06	-0.64
Subperiod 2	75 [2.5%]	3.36	0.42	68 [0.8%]	3.75	1.24	3.46	0.68
Subperiod 3	1 [0.0%]	1.74	0.13	3 [0.04%]	3.87	-0.52	3.35	-0.09
Subperiod 4	92 [1.1%]	3.79	-0.63	57 [0.7%]	3.21	-1.42	2.99	-2.57
USO	***Median size***	**IS**	**OOS**	***Median size***	**IS**	**OOS**	**IS**	**OOS**
Subperiod 1	4 [0.05%]	3.94	-0.29	21 [0.3%]	3.71	-2.61	3.20	0.00
Subperiod 2	45 [0.5%]	3.61	0.98	5 [0.05%]	3.52	1.12	2.96	0.00
Subperiod 3	1 [0.0%]	3.63	-0.27	9 [0.1%]	2.74	-0.66	2.66	0.15
Subperiod 4	65 [0.8%]	3.69	-1.20	85 [1.0%]	3.74	-1.22	3.50	0.00

This table indicates the in-sample (IS) and out-of-sample (OOS) annualized Sharpe ratio of trading rules chosen according to the 10%-FWER portfolio of Romano and Wolf (2007), and the 10%-FDR$^+$ portfolio of Bajrowicz and Scaillet (2012) based on subperiods chosen with the nonparametric change point detection approach of Ross et al. (2011). Semi-annual rebalancing is employed using the risk-free rate as benchmark, as well as the best rule in-sample. The table also displays the portfolios' median sizes in levels, and in percentages of the total rules universe in brackets, across different subperiods.