Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Statistical Testing of DeMark Indicators in Commodity Futures Markets

Department of Management, Technology, and Economics (D-MTEC)
Chair of Entrepreneurial Risks (ER)

Master Thesis for the Degree of
Master of Science in Management, Technology and Economics

Examiner: Didier Sornette, Prof. Dr.
Supervisors: Donnacha Daly, Dr.
James Isilay

Marco Lissandrin

August 2014

STATISTICAL TESTING OF DEMARK INDICATORS

IN

COMMODITY FUTURES MARKETS

Disclaimer:

The DeMark Indicators® are a collection of registered trademarks and are protected by U.S. trademark law. They all belong to Market Studies, LCC. This thesis makes no commercial use or makes any claim on the industrial application. This is an accademic research and any implementation of what might refer to DeMark Indicators (i.e. TD Setup™, TDST™, TD Countdown™, TD Sequential™, TD Combo™, TD Aggressive Sequential™, TD Aggressive Combo™, TD Camouflage™, TD Clop™, TD Clopwin™) is a personal interpretation from the book *DeMark Indicators* by Perl J.[45].

Abstract

The profitability of trading strategies based on Technical Analysis (TA) is still a debatable topic. On one hand there is little academic research that supports the usefulness of TA despite a long debate in academia that goes back to the first half of the past century. On the the other hand its use has been widespread in the commodity futures markets since then. This work examines one corner of this vast topic by studying the performance of DeMark indicators over 21 commodity futures markets and 10 years of data. These indicators are a reference for practitioners and, from a commercial perspective, it's possible have the DeMark Indicators® as an upgrade in leading financial platforms such as Bloomberg Professional® and Thomson One®. The goal is to backtest the predictive power of these indicators and, therefore, to discorver examples of statistical significance. The method used to test significance consists of two phases. Initially, market entry signals are tested by comparing conditional returns (i.e. conditioned on the entry signals) to unconditional returns. For the analysis of trades, which also comprise of market-exit signals, randomization tests have been performed for benchmarking. By doing so it is then possible to generate distributions of performance metrics (i.e. cumulative profits or profit factors) that can be used for hypothesis testing. A further step has been taken in the analysis by checking the impact of the rolling strategy of future contracts on the performance of this class of indicators. All tests have been done in a Matlab® based simulation environment where the performance of trading signals is evaluated.

Simulation results suggest statistically significant predictive power on a wide range of commodity futures, but before using the suggestions of this work to make trades, results should be cross-validated on the most recent out-of-sample data available [3]. There are multiple reasons that suggest this further step: the first is related to the data mining bias that is always a potential threat in long, iterative selection processes where a lot of data is analysed and optimized based on results. The second reason is that indicators may loose their predictive power over time due to changes in market conditions.

Contents

Introduction

Technical Analysis (TA) refers to a set of methods that use past market activity such as price, volume and open interest to identify patterns that can predict future price movements. Traders using TA believe that some of these methods have predictive power. This approach to price forecasting has been popular among merchants already in ancient times at the time of Babylonians, Greeks and Romans. In 1961, 76% of amateur speculators examined price charts according to a survey by Smidt [52] and in 1983 a survey by the Chicago Board of Trade(CBOT) concluded that 50% of all speculators consulted charting services [32]. A more recent study from 2001 with a focus on foreign exchange market partecipants [41] showed that only a very small percentage of traders did not us TA. Furthermore, TA usage seemed to be increasing in the last decade.

Despite of its popularity among practitioners, TA has always been facing difficulties to be accepted in academia. In modern times, the weak form efficient market hypothesis (EMH), deeply discussed in the 1960s, suggested that past returns had no effect on future returns, therefore no TA based trading strategy could be profitable according to the hypothesis. This economic way of thinking was supported at the time by Samuelson and Fama [20, 48], the latter was also suggesting that gross profits could have been captured by trading strategies based on TA, but in those cases profits went to zero as soon as transaction costs were taken into consideration. Weak EMH was the starting point for the development of Disequilibrium Theory which provided a theoretical reason why trend following systems might work. It suggested that markets are in a short run disequilibrium. In 1980, Beja and Goldman [7] were claiming that it is intuitive to believe that a man made institution such as the market could not be mechanically perfect. The practical message given by their studies on the dynamics of the markets is that there is a difference between equilibrium and non-equilibrium speculation. The first assesses future equilibrium based on fundamental financial data while the second forecasts non-equilibrium adjustments by using TA. Therefore, it's not a fight between which approach is correct since both are complementary and useful for speculation purposes. This is in line with Oberlechner's survey on the approach used by market participants [41].

It has to be said that most academic studies until 1998 on TA's backtesting[1] have given contradictory results. There has been in the past a failure to establish a direction of research in this area due, for example, to the highly subjective nature of TA.

One of the goals of this work is to test rigorously the predictive power of entry signals by using as a reference Lo's systematic approach to evaluating the efficacy of TA [29] which compares unconditional empirical distributions of daily returns to the conditional distribution. The first element of originality of the work that will be explained in the next chapters is that the test is performed on a selection of DeMark Indicators. This family of indicators is currently one of the most popular of TA and it is possible to make use of them as an upgrade in leading financial platforms such as Bloomberg Professional® and Thomson One®. While the first platform has approximately 300'000 subscribers, the second is part of Thomson Reuters and can reach another 300'000 users; this means a combined market share of 60%. It is worth pointing out that these indicators are not trading systems, but they are indicators of turning points therefore they only provide entry signals. To convert them into trading signals a possible solution is to parametrize the number of holding days so that no exit strategy can influence the performance of DeMark's entry signals. In this way it is also possible to compute randomization tests which, in comparison to conditional and unconditional returns, can provide not only information regarding the predictive power of these indicators, but also more direct information for practitioners such as gross profits per trade. Journals for practitioners such as *Technical Analysis of Stocks & Commodities* are continuously providing articles describing new indicators, but it's highly uncommon to find statistical test of existing indicators as this work tries to do.

As already mentioned, Technical Analysis goes back to the Babylonians and it was used to study prices of several commodities such as staples, barley, dates, sesame, wool, mustard and cress [30]. Moreover, the first known backtesting of trading signals using historical data comes from Kurz, a commodity trader in the Antwerp area who tried in 1540 to distinguish himself from the other traders since commodity trading in that area was becoming increasingly hectic and risky. In the same spirit, we have constructed signal backtesting on a total of 21 commodity futures which can be categorized as grains, softs, energy, industrial metals and precious metals. A commodity market is a market that trades raw materials. These markets are physical when the product is delivered, otherwise they are financial (for hedging and speculative purposes). Besides an opposite approach towards market risk which also influences the trading style, financial instruments are a key element in both physical and speculative trading. Commodities use futures contracts as the basic type of exchange based financial instrument. Globally, there are dozen of exchanges dedicated to commodities.

[1]Spyros Skouras provides a detailed bibliography of academic studies on TA and it's performance until 1998 [51].

Some of the most famous are the *Chicago Board of Trade* (CBOT) for grains, the *IntercontinentalExchange* (ICE) for softs and energy, the *New York Board of Trade* (NYMEX) for energy, the *Commodity Exchange, Inc* (COMEX) for precious and industrial metals, but metals are also traded on the *London Metal Exchange* (LME) for Europe and on the *Shanghai Futures Exchange* (SHFE) for Asia. Yet, there are relatively few studies of technical analysis on commodity futures markets. Lukak et al. [32, 31] and Roberts [47] cover what is currently available from the academic side. A possible explanation is that constructing a continuous price series using futures data is not straightforward since future prices are represented in contract months and for each trading day there are multiple prices available each one coming from a different contract month. Once a specific contract is used to determine the price of the trading days, there is no rule that tells when to "roll over" to the next contract. This work goes beyond what is now available because for commodities where DeMark Indicators have suggested positive and statistically significant performance, different strategies for the roll over are tested in order to understand how sensitive these indicators are to the rolling strategy applied to these markets.

The chapters are organized as follows. Chapter 1 describes the Indicators used to generate entry and trading signals. In Chapter 2 different ways for evaluating the performance of the entry/trading signals are discussed. Chapter 3 explains the problem of rolling futures contracts and shows how to create continuous daily returns. Chapter 4 shows the performance of signals described in Chapter 1 and backtested on the continuous returns time series of Chapter 3. The work ends with some concluding remarks.

CHAPTER *1*

DeMark Indicators

The chapter is structured as follows: §1.1 provides a high level overview on different approaches to Technical Analysis (TA). §1.2 justifies the choice of the indicators selected for backtesting. The remaining sections describe in details each selected indicator. §1.6 concludes the chapter with the set of parameters that will be used for backtesting. DeMark Indicators carry the name of the technical analyst who introduced these indicators. This person is Tom DeMark. He started in the investment business in the 1970s and, since then, he developed his career as a chartist. His clients included George Soros, Goldman Sachs, Union Carbide and IBM. He has also advised Paul Tudor Jones and Leon Cooperman. Currently, Market Studies, LCC (which he founded) provides his indicators to Bloomberg, CQG and Thomson Reuters. He is also a consultant to Steven Cohen, founder of SAC Capital Advisors LP, which manages $14 billion.

1.1 Price-based and Time-based Forecasting [2, 13]

In TA prices have always been the primary reference of past market activity with which technicians could predict future market sentiment. In ancient Babylon price records for commodities were kept for centuries. Ancient Greeks used price levels as a market sentiment analysis to decide how much of a given commodity to buy, hold or sell, while in Ancient Rome prices were used to identify seasonality patterns [30]. Like physical objects, prices have inertia and, when at rest, they often stay at rest. On the other side, when in motion, they often stay in motion along the trend. [2] provides a framework for TA based on a probabilistic mechanical view of market movements. In this setup, technical indicators can be seen as combinations of measurements of price velocities and price accelerations. A price velocity is the rate of change of the price and a price acceleration is the rate of change of the price velocity. Price levels and price ratio relationships between highs and lows are price-based forecasting techniques that try to identify conditions in which prices are in motion along the trend. A very general example is given by support and resistance levels. These levels are price areas where abundance of trading has taken place and there is buying or selling pressure. The crossing of these levels means that prices are in motion and are likely to continue along the trend. This concept can be applied in several different ways, for example with trendlines, moving averages, or Elliott waves. Kosar and Widner provide an overview on the topic [27, 58]. There is another way to predict future market movements other than price-based forecasting, and this is time-based forecasting [13, 38]. This class of methods tries to identify patterns in time series that should repeat over time. Bar counting techniques are a wide subset of the class and a couple of famous examples are the Fibonacci counting method and the Lucas number series.

1.2 A Selection of Tom DeMark (TD) Indicators

The book from Perl [45] provides a detailed description of 39 indicators. Many of them are revised versions of traditional indicators such as Elliott waves, trendlines, price ratios and moving averages. Despite this, DeMark is mainly renowned for his TD Sequential. It is a time-based indicator that identifies potential turning points from trends (TD Setup) and then forecasts the beginning of price reversals (TD Countdown). TD Combo is the main variant to TD Sequential: while it uses different rules to forecast the timing of price reversals, turning point identification stays the same via the TD Setup phase. For the following reasons Sequential and Combo should be the first DeMark indicators to be tested. The 3rd and last indicator is TD Setup Trend (TDST). It comes from the same family of the other 2 indicators because is uses TD Setup as well. Anyhow, it doesn't use the Setup phase to identify price patterns, but to generate support and resistance lines. As discussed in §1.1, when support and resistance curves are crossed then prices are in motion and should continue along the trend. Conversely, Sequential and Combo try to identify patterns for price reversals.

To summarize, 3 indicators will be part of the backtesting:

- ▷ TD Sequential
- ▷ TD Combo
- ▷ TD Setup Trend (TDST)

They all have a common starting point, but still each one has its own features. TD Sequential is DeMark's most famous indicator, TD Combo is its natural variant and TDST is complementary to both because it doesn't search for price reversal patterns, but it tries to capture sustainable trends. It is possible to find additional information on DeMark indicators directly from DeMark's book [16].

1.3 TD Sequential

This time-based indicator tries to identify areas of trend exhaustion that will lead to price reversals. Coles [14] provides a short description of the indicator. It is made up of two sequential parts: the first one is the Setup, which tries to capture momentum, and it is followed by the Countdown with looks for trend exhaustion. When all the conditions in both phases are satisfied, then an entry signal is generated according to one of the entry strategies described by DeMark. To generate a long entering signal, the Setup has to identify a bearish momentum in the market. In the Countdown, prices can continue to be bearish or they can go sideways while in the meanwhile the trend exhaustion pattern is building up. As soon as the Countdown is completed, prices should start rising up within the following 12 price bars (according to Perl). A symmetrical algorithm generates a short entry signal. In reality instead, there is no symmetry between uptrends and downtrends in the markets (Benyamini [8]). In fact, downtrend pullbacks are deeper, pauses in selling are common while buying pressure is relatively even and tops have higher volatility than bottoms. Given that the signal is built symmetrically, it will be interesting to analyse its performance in markets that don't behave symmetrically. Such analysis is provided in Chatper 4, while the rest of this chapter will give further details about the signals and their setup.

Before proceeding with a description of the single phases it should be stressed that Sequential is not a trading system, but it only provides and entry signal which can be long or short. No method is given on how to handle the trade. Despite this, nothing is missing for a complete mechanization of the indicator. More on the implementation side, additional algorithms are available which, according to specific price developments, can:

- ▷ cancel or restart ("recycle") a TD Sequential before it can reach completion and, therefore, generate a new entry signal
- ▷ exit positions when specific price levels are reached just after a position has been entered

1.3.1 TD Setup

The following description of the Setup phase focuses only on the generation of a long entry signal because TD Sequential is symmetrical for long and short positions.

This phase identifies momenta in price series and in case of long positions it looks for a bearish momentum. To do so closes are compared to the close n bars earlier. The idea of using n-days momentum to avoid noise is not unique and can be found in Chan & Lin [11]. A Long Setup is completed when there are m consecutive closes, each one less than the corresponding close n bars earlier. Here is the conditions that completes a long Setup, t represents the t-th bar:

$$\forall\, t \in \text{Setup}, \quad PC(t) < PC(t-n) \tag{1.1}$$

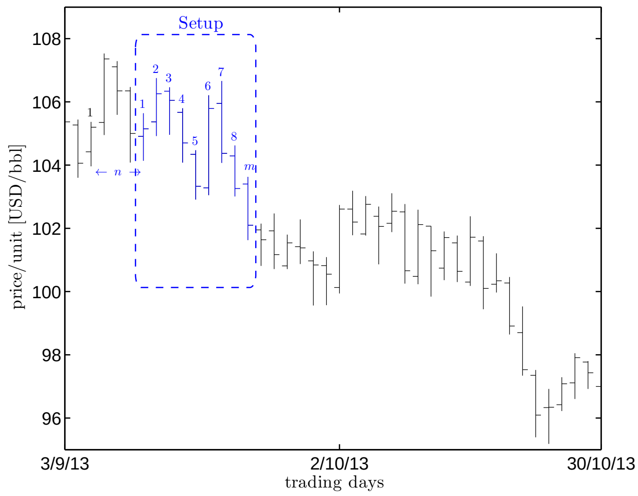


Figure 1.1: *To generate a long entry signal with TD Sequential, TD Setup has first to identify a bearish momentum in the market. A completed Long TD Setup occurs when there are m consecutive closes, each one less than the corresponding close n bars earlier. In the figure m is set to 9 and n to 4. The underlying is the continuous future price for Light Crude Oil which has been constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M.*

Fig.1.1 shows an example of a completed Setup. In this intial phase the goal is to identify a negative price velocity, i.e. a negative trend. There can still be price oscillations, but an n-day momentum guarantees that the amplitude is small enough and the period is short enough so that there is no interruption of the Setup. A Setup ends only when the number of consecutive closes cannot be increased. Just at that point a new non-overlapping Setup can build up. It is also possible to generate entry signals directly from the Setup, but to do so several other conditions have to be fulfilled. This signal is not covered in these chapters, but an extensive description can be found in Perl's book [45].

1.3.2 TD Countdown

Due to the symmetry of TD Sequential the description focuses on the generation of long entry signals only, as in §1.3.1.

The Countdown is the phase in TD Sequential that follows the Setup. The goal is to identify patterns of trend exhaustion which will lead to price reversals. Starting from the last m-th price bar of the Setup, a long Countdown is completed when there's a total of p closes each one less than or equal to the low q bars earlier. Here is the condition that makes the current bar increase the total of the Countdown, t represents the t-th bar:

$$\text{Given } t, \qquad PC(t) \leq PL(t-q) \tag{1.2}$$

In the aggressive countdown version instead, bar lows are tested against previous bar lows:

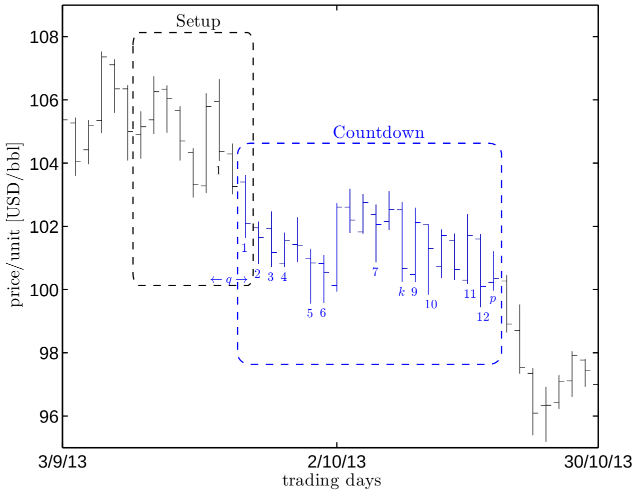


Figure 1.2: *The Countdown phase starts on the last m-th price bar of a completed long Setup. For a Long Countdown to be completed there must be a total of p closes each one less than, or equal, to the low q bars earlier. In the aggressive countdown version (the one in the chart) there must be total of p lows each one less than, or equal, to the low q bars earlier. Furthermore, the p-th bar where the prior condition holds must have a low less than, or equal, to the close of the k-th bar. The underlying is the continuous future price for Light Crude Oil which has been constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M.*

$$\text{Given } t, \qquad PL(t) \leq PL(t-q) \tag{1.3}$$

This case is shown in Fig.1.2. In addition, for both cases the p-th bar where the prior condition holds must have a low less than, or equal, to the close of the k-th bar. If it is not verified immediately, then the completion of the countdown is postponed until this condition is met in one of the later bars. Just at that point the p-th bar will be identified and the countdown completed. Like for the Setup phase, the test tries to identify another negative trend that can include oscillations which are limited in terms of amplitude and period. Anyhow, oscillations are different compared to the Setup because q and n are two separate parameters and in the countdown phase larger oscillations are tollerated; this will only cause a delay and not a cancellation of the entry signal. In addition, the PC are compared with PL which means that the trend should accelerate compared to intraday volatility.

Price bars do not always translate into a TD Sequential. For example, it could happen that a price reversal already started before the completion of a long Countdown. In this case, an interruption of the Countdown will easily occur. There are additional checks in the Countdown that can restart ("recycle") this phase or, in the worst case, end up the current Sequential before its completion. All of these conditions are active when entry and trading signals are generated in the backtesting phase. Here is the list of them:

1. Opposite Setup: when turned on, the Countdown is restarted as soon

as a Setup in the opposite direction completes. For example, a long TD Sequential signal is building up and before its completion a short Setup is completed. The long Sequential will be ended and a short Countdown will start.

2. Price extremes cross support and resistance levels: when turned on, in case of a buy, if PL is above the resistance level determined by the Setup, then the Countdown is stopped and cancelled (refer to §1.5 read how completed Setups can generate support and resistance curves).

3. New Setup, but a close price is still in the range of the current Setup: if the new Setup has a PC extreme in the range of the current Setup, then there is no switch to the new Setup. The old Countdown is kept going. The range of a Setup is the difference between the highest PH and the lowest PL both within the same Setup.

4. New Setup, but of a wider range: when turned on, if for the current Setup the associated Countdown phase has still to be finished and there is a new completed Setup in the same direction as the last Setup, then the Countdown is being reset only if the range of the new Setup is wider than the current one.

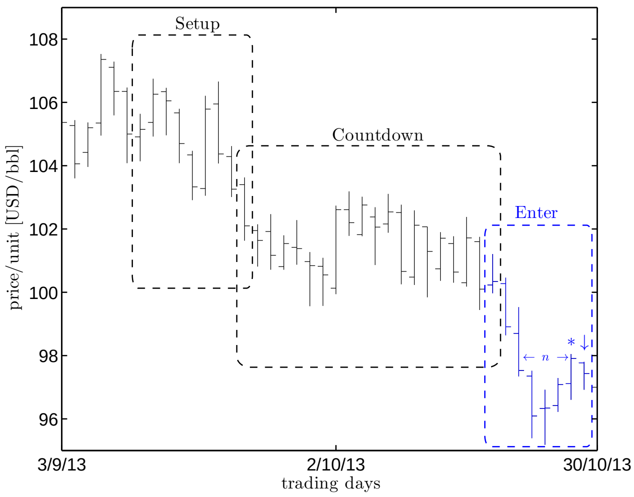


Figure 1.3: *A conservative strategy enters a long postion as soon as the price close is greater than the close n price bars earlier. This does not happen on the same day t in which the test is true (check for *), but at day t+1. An arrow points to the entry bar of day t+1. The underlying is the continuous future price for Light Crude Oil which has been constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M.*

Once the Countdown is completed, an entry strategy determines when to start a long entry signal. In [45] there are several ways described to do so . The *aggressive* strategy enters a long position as soon as the Countdown is completed. On the other hand, the *conservative* strategy enters at time t as soon as the price close is greater than the close n price bars earlier:

$$\text{Given } t, \quad PC(t) > PC(t-n) \tag{1.4}$$

This last method has been chosen for backtesting in Chapter 4. It uses a similar check as the one used for the Setup phase (1.1), but here the direction of the test is flipped because the price reversal should start now, with a successful check of (1.4). Fig.1.3 provides an example. For the interest, other entry strategies are *camouflage* and *clopwin*. Refer to Perl's book for the description. Once (1.4) is satisfied at day t, a long position will be entered. This does not happen on the same day t, but at day *t+1*. The reason is the following: at the end of day t the final price bar of that day is finally available. Therefore once the entry condition is also satisfied, a position can be entered, but only from the following day. It is a common mistake in backtesting to enter a position in the same day in which all the conditions are met, because it doesn't represent reality. By entering at day t the entry signal looks-ahead into the future [1].

1.4 TD Combo

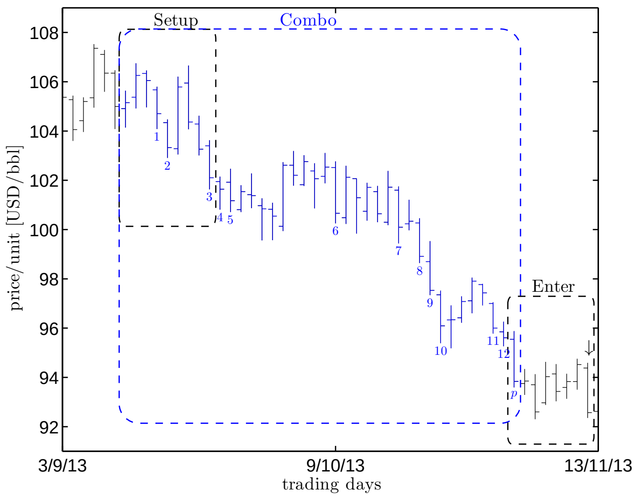


Figure 1.4: *Combo is an alternative to the Countdown phase. The Setup and the entry strategy remain unchanged. It uses the same condition (1.2) as in Countdown to increase the total number of bars but, differently from the Countdown, the check starts from bar one of the Setup. The underlying is the continuous future price for Light Crude Oil which has been constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M.*

[1] some traders use the price 30 minutes before the market close as a proxy for the real close price and execute the trade in the remaining time before the closure. Yet, it is not examined in this work.

It is possible to use Combo as an alternative to Countdown in the TD Sequential multi-phase signal generation. The other phases (Setup and entry) are not influenced by this change. Like the Countdown, there is an internal bar counter that has to reach bar p before this phase is completed. (1.2) must be fulfilled to qualify bar t as part of the Combo. As it is also pointed out in fig.1.4, a major difference with the Countdown is that the bar check starts at the first bar of the Setup instead of the last. This would guarantee that it takes less bars from the first Setup bar to the bar at which the signal is entered (only if p is the same for Countdown and Combo). Unfortunately this is not true because additional conditions (1.5), (1.6) and (1.7) have to be met for bar t to be part of the Combo:

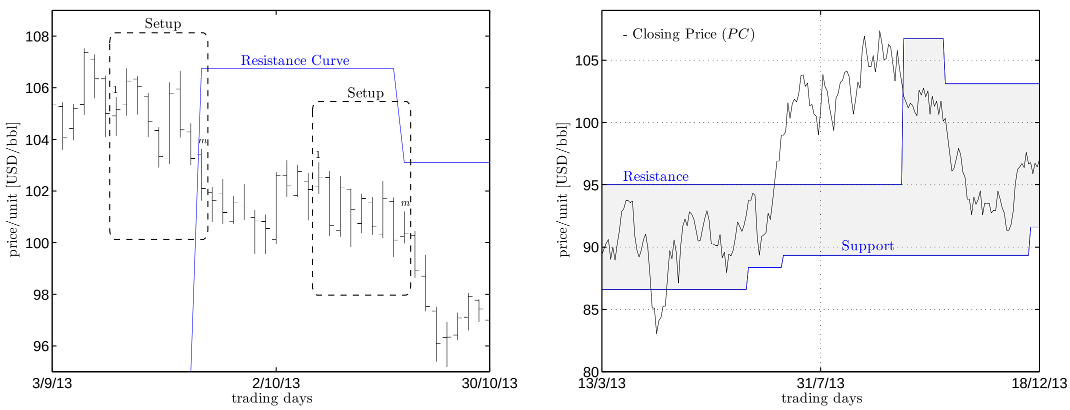
Given t,

$$PL(t) \leq PL(t-1) \tag{1.5}$$

$$PC(t) < PC(\text{prev. Combo bar}) \tag{1.6}$$

$$PC(t) < PC(t-1) \tag{1.7}$$

1.5 TD Setup Trend



(a) (b)

Figure 1.5: *It shows how to build a resistance line (Fig. a) and, more in general, how to use support and resistance lines to generate entry signals (Fig. b). In Fig. a), the resistance line needs long Setups to change its values. Each completed long Setup has a bar containing the maximum price high with the bars belonging to the Setup. This value will update the resistance line, but only at the last bar of the Setup to avoid look-ahead. In Fig. b), When support and resistance lines are crossed and the closing price is outside the shaded area, then prices are in motion and should continue along the trend. The underlying is the continuous future price for Light Crude Oil which has been constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M.*

As the name suggests, this indicator uses TD Setup just like TD Sequential and TD Combo. The Setup phase is used to generate support and resistance levels. As discussed in §1.1, when support and resistance levels are crossed, then prices are in motion and should continue along the trend. Vice versa, Sequential and Combo try to identify patterns for price reversals. Looking at Fig.1.5b the concept of support and resistance can be better understand. Whenever the closing price escapes the shaded area, then the price trend has the strength to continue. The figure shows a interesting example for a long trend. Furthermore, this indicator could also provide a good exit signal. Discontinuities in support and resistance lines could help to exit in a time period in which the price delta is in favour of the trade, just before a price reversal. Unfortunately, only the entry signal will be backtested for this indicator. Since for Sequential and Combo there is just an entry signal, a fair comparison should test all the entry signals for all the indicators. This choice leaves out the exit signal of the Setup Trend indicator from the statistical tests although this exit signal could have predictive power.

The last thing missing in the description is how to use the Setup to generate the lines. A support line gets updated each time a short Setup is completed. On the other hand, the resistance line needs long Setups to change its values. Let's focus on resistance lines. Each completed long Setup has a bar containing the maximum price high with the bars belonging to the Setup. This value will update the resistance level. See the example in Fig.1.5a. To avoid the look-ahead problem, also if the bar containing the highest price high is within the first bars of the Setup, the resistance line gets updated only when the Setup is complete. Setups can go beyond the minimum number of bars required for completion, but waiting until the last Setup bar is not desired. By doing so there would be no benefits since the resistance levels would just be updated later and the time delay with the closing price would be wider.

1.6 Parameters for Signal Backtesting

Before proceeding to the next chapters, it is important to set the parameters that will be used in Chapter 4 (Tab. 1.1). All numerical parameters are set according to Perl's description. He is not inclined to optimize them although the study was done more than thirty years ago. The reason for him is that those parameters have proven to be robust irrespective of the market, its volatility, or the time frame. Studies that go into the direction of parameter optimization would tend to fail when the behaviour of the market changes. The only remark is that TD Sequential uses the aggressive version (1.3) instead of the standard condition (1.2).

1.7 Summary

This chapter examines the three indicators that will be part of the performance backtesting. TD Sequential is DeMark's most famous indicator and TD Combo is its natural variant: they both are time-based forecasting methods that try to identify patterns for price reversals. TDST is complementary to both because it doesn't search for price reversal patterns, but it tries to capture

Table 1.1: *List of parameter settings.*

General parameters	
m	9
n	4
Sequential & Combo	
p	13
q	2
k	8
recycle 1	on
recycle 2	on
recycle 3	on
recycle 4	on
aggressive Sequential	on
Entry Strategy	
conservative	on

sustainable trends. For all of the indicators, long and short entry signals are always generated by symmetrical algorithms. In reality instead, there is no symmetry between uptrends and downtrends in the markets. The parameters are set according to default values, many of which were already set by DeMark more than thirty years ago. Parameter optimization is not part of this work and it could a suggestion for its further development.

CHAPTER *2*

Performance Measurement

The ultimate goal of this chapter is to describe a quantitative process to measure and estimate the performance of trading signals and to test their predictive power. §2.1 explains how the test period was chosen. The next step in §2.2 is to decide what has to be measured in terms of trading statistics and performance metrics. it is not sufficient to measure performance without any comparison to reference values. In the case of trading signals, this is done in §2.3 with a comparison to market performance. Initially, conditional performance is compared to unconditional performance and then, the null hypothesis that long, short and neutral positions are paired randomly with daily market returns is tested by means of Monte-Carlo permutations.

2.1 Test Period

The time frame in which a trading signal is tested plays a critical role. In fact, an indicator usually gives a performance which is dependant on the time period during which it is tested. This might be due to changes in the market behaviour. De facto, the choice of a suitable time frame is problem dependent and by no means trivial. The period length depends on the signal's sparsity: for example, one single trade in the whole time frame doesn't provide enough information to build up statistics and, more importantly, it would be difficult to take future trade decisions based on an interpretation of backtest results. As a rule of thumb, it is desirable to have a minimum of 10 long and 10 short entry signals for each DeMark indicator across all 21 commodity futures. For this reason and given the DeMark indicators, at least 10 years of data for each commodity are needed. The end of the time frame should be as near as possible to the present because there are maximum probabilities that past genuine predictive power can still be predictive in the short term future. Nevertheless, there is the need for unused data following the time frame, because the results within the time frame should be retested on out-of-sample data to avoid data mining errors and also because market conditions might change during the tested period. For these reasons, the last trading day of the time frame is set to 1/1/2014. Ideally, the duration and the ending date of the time frame should also fix the backtesting period, but data constraints could still have an impact on the starting date. As discussed in Chapter 1, DeMark indicators require daily price bars. In Reuters Datastream Professional® this data is available for all commodities only from mid 2003 with Brent Crude and Gas Oil being the bottlenecks. Therefore the choice of the backtesting period is the following:

Table 2.1: *Backtesting Period.*

start:	1/1/2004
end:	1/1/2014
duration:	10 years
trading days:	2610

2.2 Measures for Trading Strategies

When an indicator gets combined with historical data, the output is a trading signal. It is possible to match signals with returns once daily returns for the same backtesting period have been generated (see §3.2.3). Each position is made up of an entering day and an exiting day, therefore all daily returns in between that time frame fully characterize the position in itself. This is shown in fig. 2.1. When each position is characterized by a sequence of returns, it is then possible to decide which performance measures are more suitable to evaluate the predictive power of the DeMark indicators.

Sequential and Combo provide entry signals, but the framework just described is still valid because each entry signal is matched with an exit signal once the number of holding days is fixed. TDST, instead, provides both entry and exit signals but, to compare all the indicators in the same way, entry signals will be considered for each indicator. Before discussing about performance it is important to understand the characteristics of the signal. Since all the information from the signal is limited to entry points, the analysis of the signal is limited as well, but it is still important.

Table 2.2: *Each commodity has a total sum of long and short positions per year. Across all commodities each indicator has a range of frequencies which is shown in the table below. The first results from the backtest is that these DeMark Indicators have sparse entry signals and this effects how performance is measured.*

Pos/Year	**Sequential**	**Combo**	**TDST**
min:	3.0	1.2	3.4
mean:	3.9	1.7	4.5
max:	4.6	2.2	6.1

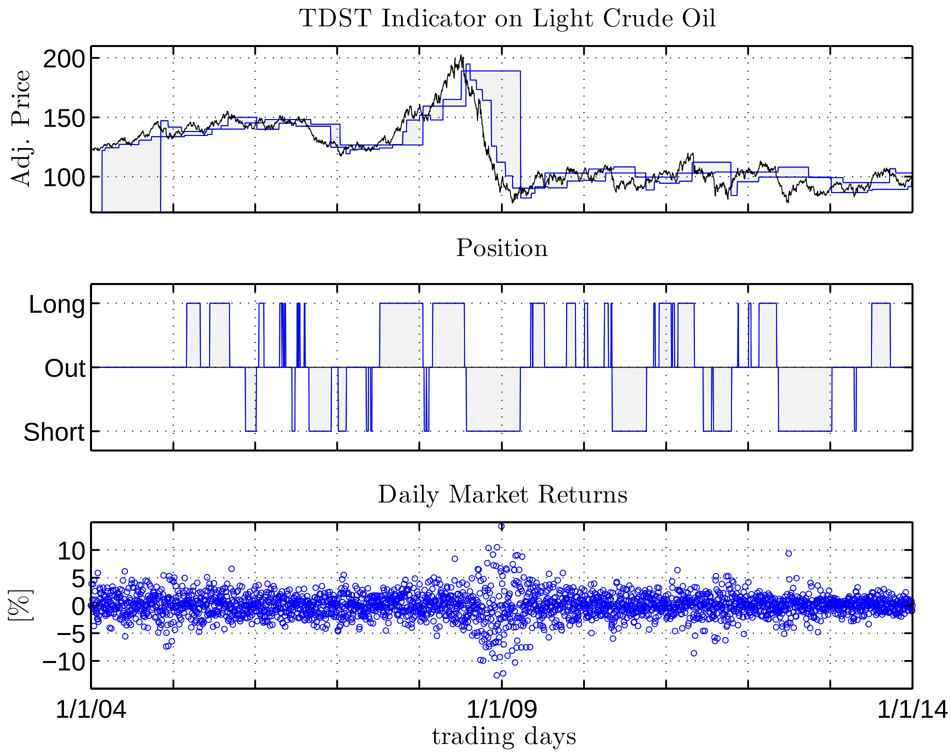


Figure 2.1: *To measure the performance of trading signals there has to be an indicator which can generate a signal by using historical data. The upper graph shows how it is possible to create a trading signal (in the middle graph) by using resistance and support lines from TDST. When the closing price crosses support and resistance lines and goes outside of the shaded area, then prices are in motion and should continue along the trend. The trading signal is then matched with market returns (bottom graph). To do this, each position has an entry and an exit date. All daily market returns in between those dates belong to the corresponding position. How to generate rolled and adjusted price time series for DeMark indicators is a problem that will be addressed in Chapter 3. How to retrieve continuous market returns from multiple futures contracts is also discussed in Chapter 3.*

For instance, tab. 2.2 shows the number of trades for each indicator across all 21 commodities. This result suggests that entry signals for the tested indicators are sparse and this puts some constraints on the performance measurement of §2.2.2.

2.2.1 Performance of Single Positions

A sequence of returns characterizes an entry position as soon as the number of holding days is given. If this number is swept instead of being fixed, then it is possible to see what happens to compounded returns during the days after each trade has been entered. The l-th-day continuously compounded returns of an entered position is the compounded return that starts being computed from the entering day for a duration on l holding days. These returns are called *conditional* because they are computed each time that an indicator generates a new entry signal. It is possible to represent all conditional returns of a signal of i trades in the following form where T is the maximum number of holding days:

$$\mathbf{R_c}= \overbrace{\begin{bmatrix} (r_c)_{1,1} & \dots & (r_c)_{1,l} & \dots & (r_c)_{1,T} \\ \vdots & & \vdots & & \vdots \\ (r_c)_{i,1} & \dots & (r_c)_{i,l} & \dots & (r_c)_{i,T} \end{bmatrix}}^{\text{holding days}} \tag{2.1}$$

Fig. 2.2 is the graphical translation of what has just been described.

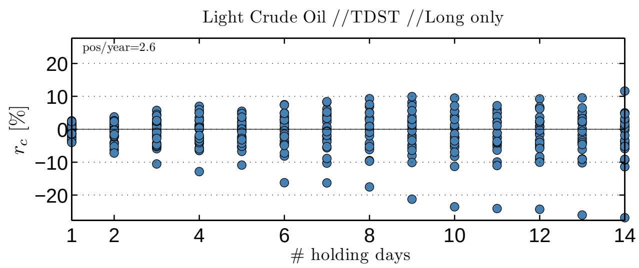


Figure 2.2: *Conditional returns are conditional to entry signals. For both long and short entry signals it is possible, by sweeping the number of holding days, to see how these returns evolve over time. This scatter plot shows conditional returns r_c for long Sequential positions on Light Crude.*

A further step in the analysis is to substitute conditional returns with a conditional return-to-risk ratio. A measure that can be applied to single trades is the *Risk Return Ratio* (*RRR*) [26]. In this context it can be defined as:

$$(RRR)_{i,l} = \frac{(r_c)_{i,l}}{(DD_{\max})_{i,l}} \tag{2.2}$$

where,

r_c is the conditional return for the given trade;
$DD_{\max}$ is the maximum drawdown within the trade's time period.

Maximum drawdown $DD_{\max}$ is defined as the largest (compounded, but it can be also found as uncompounded) cumulative return within a defined time period. Drawdown-based measures are widely used in practice. Commodity trading advisors impose drawdown regulations in their trading strategies to avoid worst-case scenarios [12]. These performance measures were mainly developed by practitioners and most of the literature on maximum drawdowns can be found in finance journals dedicated to the investment community [18]. $DD_{\max}$'s advantage is that it focuses on worst-case scenarios which capture characteristics of returns that cannot be captured by standard deviation. Furthermore, it can also be computed on a single trades.

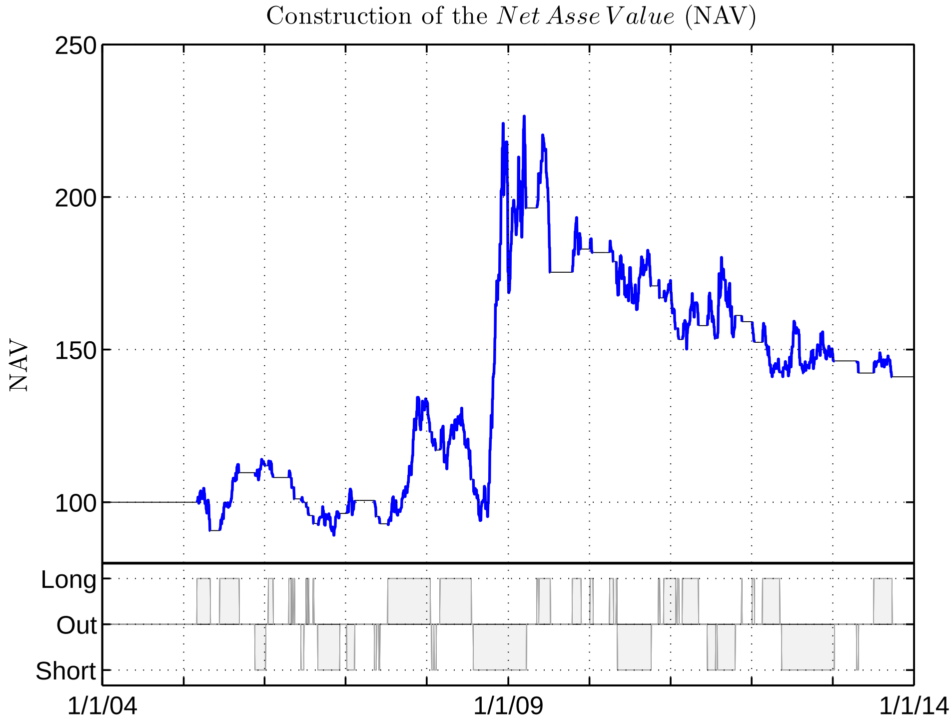


Figure 2.3: *Given the a signal with entry and exit points, the Net Asset Value (NAV) shows the evolution of an initial nominal wealth when positions are opened and closed according to the signal. In particular, the initial bidding value of each position is the total accumulated so far by the previous positions. Each postion contributes to the NAV by compounding daily returns to the initial bidding values. This example the NAV computation for TDST on the continuous future price for Light Crude Oil. The time series rolls between the front M and the next M+1 contract as soon as M+1 has an open interest higher that M in the last trading month of M, just before expiry. A reason why NAV is not used here to evaluate trading signals is that it mixes profits related to signals with profits related to money allocation rules. In addition, signals are sparse and they never cover more than 30% of the total trading days.*

2.2.2 Aggregate Performance of Multiple Positions

A signal can be evaluated on single positions as in fig.2.2, or with one parameter as a measure for the whole sequence of positions. A common way to evaluate all positions is by using the *Net Asset Value* (NAV). Given an initial nominal wealth, i.e. 100, the NAV shows how that value changes over time when positions are entered and exited according to the signal. The value from the last completed trade is the starting bid for the trade to come and, in such a way, all trades cumulate their daily returns. Unfortunately, this measure cannot be used for backtesting on the DeMark indicators. There is a difference between testing the performance of trading signals and testing the performance of systems that trade those signals with specific money allocations [50]. Profits can be due to trading signals, but also to money allocation strategies. Signal evaluation should eliminate the allocation component and this doesn't happen with the

NAV for which all previous profits are reinvested in the upcoming trade. If the cumulative effect of previous trades has generated profits, then the NAV allocates a higher wealth (compared to the starting trade) to the upcoming trade. Vice versa, with cumulative losses, money allocation for the trade to come will be lower than what was allocated to the initial trade. This can be seen in a NAV example in fig.2.3, here each trade starts from a different money level. Also if this problem was ignored, it wouldn't still be possible to show NAVs because, given the sparsity of the entry signals, the days of exposure would be always lower or equal to 30% of the total trading days also with 15 trading days per trade. Therefore, for most of the trading days the curve would show no NAV change.

Given i trades in a signal and the corresponding conditional gross returns r_1, r_2, ..., r_i, it is possible to use measures that give equal money allocation to each trade, for example by computing the arithmetic mean of conditional trade returns:

$$\text{Profit}_{\text{trade}} = \overline{r_i}. \tag{2.3}$$

Besides testing the predictive power of signals, it is also important to include a value that quantifies the performance of a signal on a given commodity. In fact, from a practical point of view it is not sufficient to have predictive power if the traded signal is not profitable. $\text{Profit}_{\text{trade}}$ provides an average gross performance of the signal, but this performance becomes net once average commission costs are subtracted. This step is left to the reader because these costs depend on the market and, furthermore, it is not the trader's first priority to update total commission costs in his trading book. This task is usually done by the back office.

Conditional gross returns can be also evaluated by the Profit Factor. This is a profit-to-loss ratio which is defined as:

$$P_f = \frac{\sum \text{Profits}}{\sum \text{Losses}}. \tag{2.4}$$

Since all trades have the same money allocation, P_f can be rewritten as the sum of returns from winning trades divided by the sum of returns from losing trades:

$$P_f = \frac{\sum r^+}{\sum r^-}. \tag{2.5}$$

It is possible to rewrite (2.5) in the following way:

$$P_f = \frac{N^+ \cdot \overline{r^+}}{N^- \cdot \overline{r^-}}, \tag{2.6}$$

where $N^+$and $\overline{r^+}$ are respectively the number of winning trades and their average return. On the other side, $N^-$and $\overline{r^-}$ are respectively the number of loosing trades and their average return. A trade is considered a winner when its gross return is strictly positive. Next, let's give two definitions: the Profit/Loss or Payoff Ratio is

$$PL = \frac{\overline{r^+}}{\overline{r^-}}, \tag{2.7}$$

and the Win Ratio is

$$W = \frac{N^+}{N^+ + N^-}. \tag{2.8}$$

The profit factor can be then written using eq. (2.7) and (2.8) to obtain

$$P_f = \frac{W}{1-W} \cdot PL. \tag{2.9}$$

By definition, there break-even is reached when $P_f = 1$. For vales above 1 the trading signal generates a profit and it is desirable to have values higher than 2. Eq. (2.9) highlights the two factors that make a trading signal profitable. Signals with low PL must have a high win ratio; this is how intraday traders can be profitable. On the other hand, high PL values can be coupled with a low win ratio (i.e <50%) and still generate profits. For example, there are profitable trend following signals which have only a few large winners and many small loosers. PL values for short term and intraday signals are between .25 and 2, while for other signals PL values should be much greater than 3 [23]. P_f includes the factors that describe a signal's profitability, therefore it is a performance measure that will be used to describe backtesting results in Chapter 4.

What is still missing is a return-to-risk measure also for aggregated trades. The RRR was introduced in §2.2.1 to compute return-to-risk performances of single trades. Now, given a set of n trades, the overall RRR value will be computed as the mean value of the RRR values of each trade:

$$RRR_{\text{trade}} = \overline{RRR_i}. \tag{2.10}$$

2.3 Comparisons to Market Performance

A positive trade record requires profit generation, but trades should also beat the market to be successful. This means that their performance should, on average, be better than what the market can do. If indicators generate signals that outperform the market, then those indicators are informative; in other words, they have predictive power. This is not sufficient to guarantee the profitability of an indicator, but yet it is the first step towards it.

Predictive power can be evaluated by comparing the conditional (conditioned on entry signals) and unconditional distributions of returns. If entry signals have been generated by informative indicators then this information should not be included into market returns and, therefore, the quantiles of conditional returns should be different to those of unconditional returns.

As already described in §2.2.1, an l-th-day continuously compounded return of an entered position is the compounded return that starts being computed from the entering day, for a duration on l holding days. This is the conditional return of an entered position and it is computed every time that a new position is entered. If the computation of l-th-day continuously compounded returns is extended from the i entering days of a signal to each day of the backtested period (i.e 2610 days), then these are called *unconditional* returns. This new sample represents the market's performance since it contains all possible l-th-day continuously compounded returns regardless of entry signals. Now it is possible to compare conditional distributions of returns (linked to the entry signals) with unconditional distributions of returns representing the market. Lastly, it is also possible, within the same framework, to compare conditional RRR values to unconditional RRR values. Fig. 2.4 compares conditional and unconditional distributions for TDST long entry signals on Light Crude Oil. Conditional quantiles are never above their respective unconditional quantiles meaning that TDST doesn't over perform the market. Vice versa, it might under perform the market since for a few holding days all conditional quantiles are below their respective unconditional quantiles. This comparison gives already an idea on what to expect from the permutation test (also called randomization test). It this last test, which will be described in the next section, it is possible to quantify over and under performance and to evaluate statistical significance of all the indicators.

2.3.1 Monte-Carlo Permutation Test

Each signal generated from DeMark indicators can be either long, short or out of the market during the backtested period. As discussed in §2.2.2, the entire signal (comprehensive of all entry positions and a fixed number of holding days) can be evaluated by an aggregated measure of quality. In this context those measures are $\text{Profit}_{\text{trade}}$, P_f and RRR_{trade}. The *null hypothesis* H_0 assumes that in the signal long, short and neutral positions are paired randomly with daily market returns (for pairing refer to fig. 2.1). In other words, the assumption is that the DeMark indicators don't have the capability to match long positions with positive daily returns and short positions with negative daily returns. The *alternative hypothesis* H_A (which is the one that the test would like prove by rejecting the null hypothesis) supports the idea that the current pairing improves performance beyond what could be expected from randomness.

To test the null hypothesis, the trading signal needs to be permuted [37]. For each randomized signal all the measures listed at the beginning of the section are computed. The total number of arrangements of positions within the signal grows very quickly given the number of trading days, the number of entered trades and the number of holding days for each trade. For example, let's assume that there are a total of 30 trading days and that the signal can only be long or neutral. If there is only 1 trade with 3 holding days, then there are only 28 permutations. If, instead, there are 5 trades of the same duration, then the number of permutations goes quickly up to $\sim 10^{4.2}$. In general, the higher is the number of trading days and the number of trades(true for sparse signals) the more permutations there are. On the other hand, an increase in the

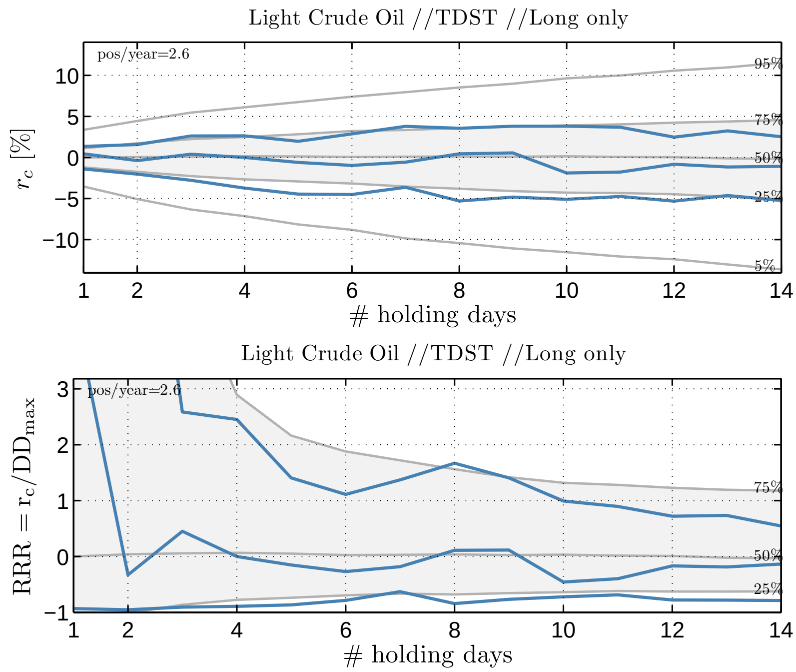


Figure 2.4: *It is possible to compare the performance of entry signals to market performance by using Lo's method [29] . In fig. a) this is done by comparing the conditional distribution of retuns (plotted using coloured quantiles at 25%, 50%, 75% levels) with the unconditional distribution of returns. Fig. b) uses the same framework, but replaces returns with RRR values to include risk. In both cases, the indicator has predictive power if the quantiles of conditional perfroamance are different from those of unconditional performance. These examples show how TDST long entry signals perform on Light Crude Oil. It never happens that all conditional quantiles are above their respective unconditional quantiles meaning that TDST donesn't overperform the market. On the other hand, for a few values of the holding days conditional quantiles are below their respective unconditional quantiles. This suggests that TDST tends to underperform compared to the market. The next step is to test statistical significance, but this is done using a different test which uses signal randomization (see §2.3.1).*

number of holding days reduces the number of permutations. In a nutshell, it is not possible to generate all possible permutations, therefore the test will be approximated by a fixed number of arrangements. The number of permutations is problem dependent and it also depends on the computational resources. 400 randomized trading signals were simulated from each trading signal on each test. It is a trade-off between quantile smoothness, size of confidence intervals and computational power.

Randomizations have some constraints. All possible permutations, including the original trading signal, must have both equal chances of appearing in real life and in the randomization process; furthermore, to avoid possible trade interactions, trades should not overlap in time. Randomizations have slightly different rules depending on whether only long, only short or combined long and short positions are tested. In case of only long or only short positions the goal is to check the superior pairing capability with daily returns. Therefore, the number of positions and the number of trading days has to be the same as in the original signal. For combined long and short positions the goal is to check both the paring with returns and the capability of the signal to go either more long or more short, in relation to the market bias. For this reason it is now the total of long and short positions that has to be the same as in the original signal. Fig.2.5 shows how permutations are done for long tests, short tests and combined long & short tests.

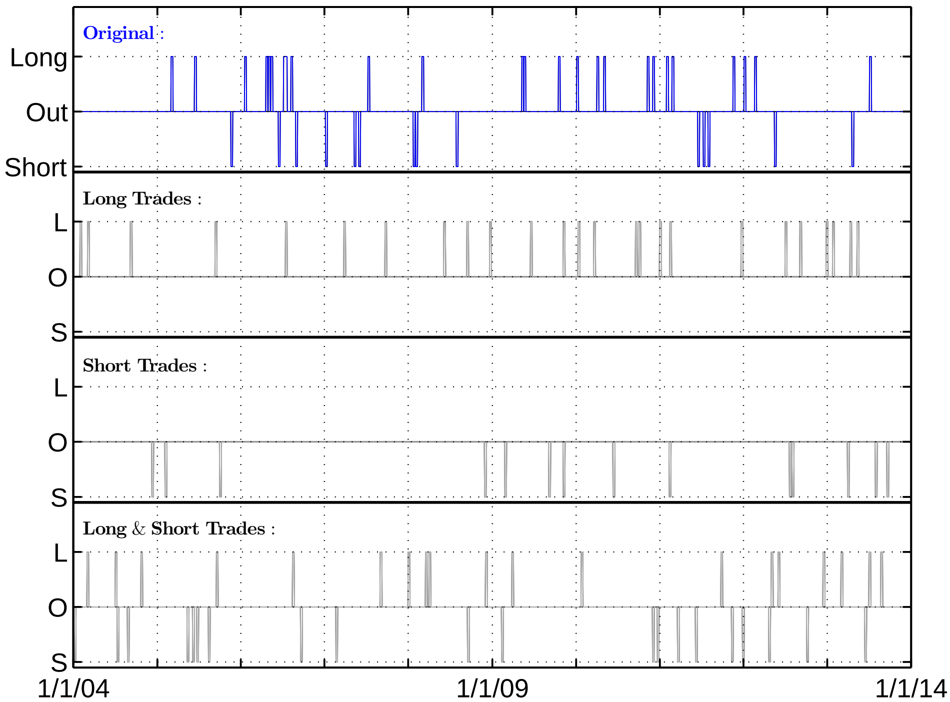


Figure 2.5: *The original TDST signal with 5 holding days per trade on Light Crude Oil needs to be permuted to run the Monte-Carlo permutation test. There are 2 main types of permutations. The first is the one that focuses only on long trades. In this case, there has to be the same number of long trades and the same total number of trading days as in the original signal. The goal is is to check the superior pairing capability of long trades with daily returns. The same flipped argumentation is valid for short trades. The second type of permutation occours when the whole signal is tested with long and short trades at the same time. In this case, while the total number of trading days has to stay the same, the difference is that the sum of long and short postions has to stay the same. In addition to superior pairing the latter test includes also the capabililty of the signal to go either more long or more short, in realtion to the market bias.*

Now, if only a small fraction of the random results exceeds the performance of the original signal, then the indicator (which originated the signal) gives significantly better performance compared to what sheer luck could do. The threshold that separates luck from significance is the p-value . In the current backtesting, the following p-values have been used [3]:

Table 2.3: *Quantile transformations in the Monte-Carlo permutation test. Permutation distributions are approximated, therefore observed quantiles need to be adjusted according to confidence intervals before being used for hypothesis testing. q and $\hat{q}$ are respectively the true and the observed quantile. These transformations use an approximated binomial method in accordance with Conover [15]. The other inputs are the number of randomized signals $n_0 = 400$ and the confidence interval $\alpha = 95\%$ required to the true q quantile . To avoid approximation it is generally recommended to use the Clopper-Pearson (1934) "exact" confidence interval for $\hat{q}$. Both methods provide the same rounded results.*

Left Tail	**Right Tail**
$q = 10.0\% \rightarrow \hat{q} = 7\%$	$q = 90.0\% \rightarrow \hat{q} = 93\%$
$q = 5.0\% \rightarrow \hat{q} = 3\%$	$q = 95.0\% \rightarrow \hat{q} = 97\%$
$q = 2.5\% \rightarrow \hat{q} = 1\%$	$q = 97.5\% \rightarrow \hat{q} = 99\%$

▷ $p = 0.1$, meaning that 10% of the random signals exceed the original signal's performance. This is called *possibly significant*.

▷ $p = 0.05$, meaning that 5% of the random signals exceed the original signal's performance. This is typically called *statistically significant*.

Looking at the original signal, the fraction of the random results that exceeds the performance of the original signal is called *observed* p-value and it is usually denoted as $\hat{p}$. The null hypothesis is rejected when $\hat{p} < p$.

The way in which p is used in hypothesis testing depends also on the type of test. While a single-sided test focuses its attention only on one side of the distribution of the permuted sample, a double-sided test considers both sides of the distribution as potential sources of statistical significance. As a consequence, in double-sided tests the threshold value is split into two parts, $p/2$ for each side of the distribution. In the case of the backtest, $\text{Profit}_{\text{trade}}$ and P_f use double sided tests. $\text{Profit}_{\text{trade}}$ measures the expected return per trade. If this value is i.e. -5% and the random distribution does significantly better $((1-\hat{p}) < (1-p))$, then, whenever the indicator suggests to enter a trade in one direction (long, for example), the trade will be entered short. The same is valid for P_f when the observed value is below the break-even ($P_f = 1$) and the random distribution is performing significantly better, then the direction of the trade should be flipped. RRR_{trade} uses a single-sided test instead. The definition uses $DD_{\max}$ at the denominator. As a consequence, it is not possible to assign to this measure a symmetrical interpretation when returns per trade are, for example, all negative.

It has already been mentioned that the permutation distribution can be only approximated by n_0 resampled signals due to computational limits. P-values are the thresholds that make a difference between the acceptance of the null hypothesis or its rejection. They are based on quantiles of the approximated distribution and, therefore, they carry uncertainty. Yet it is possible to assign a confidence interval to these p-value related quantiles by using a normally approximated binomial method in accordance with Conover [15]. The limitation of this method is that it requires large samples, but this is not a limitation for our permutation tests. Its strength lies in the fact that it can be applied to any quantile within a sample.

Given,

▷ n_o as the number of randomized signals (400 in the test);

▷ $\hat{q}$ as the observed quantile for which the confidence interval is needed;

▷ α as the confidence interval (95%) required to the true q quantile;

▷ Z_α as the Z-statistic (when α=95%, then $Z_{1-\alpha} \sim 1.65$);

then the true q quantile has the following confidence interval in terms of observed quantiles:

$$\hat{q} - \frac{Z_{1-\alpha} \cdot \sqrt{n_o \cdot \hat{q} \cdot (1-\hat{q})}}{n_o} \leq q \leq \hat{q} + \frac{Z_{1-\alpha} \cdot \sqrt{n_o \cdot \hat{q} \cdot (1-\hat{q})}}{n_o}. \tag{2.11}$$

Let's provide an example. If are 400 randomized signals and a 95% confidence interval is required to the true 95% quantile, then $93.2\% \leq q \leq 96.8\%$. In practice, this means that if a p-value refers to $q = 95\%$ then the quantile of the approximated distribution will be $\hat{q} = 97\%$. To avoid approximation it is generally recommended to use the Clopper-Pearson (1934) [1] "exact" confidence interval for $\hat{q}$. This method was used to check the normal approximation and rounded quantiles resulted the same in both cases. The In other words the significance level becomes more conservative. All the quantile adjustments adopted in the permutation tests are listed in tab. 2.3.

2.4 Summary

This part of the work focuses on how to evaluate an indicator's performance given the market returns, but first, since an indicator usually gives a performance which is time dependent, it is important to define a suitable evaluation period. The end of the time period is close to the present, but there should be still a time gap between the last trading day and the current date. This gap determines the out-of-sample data which is precious because the results of the backtest can be retested before making trading decisions on them. A smaller-scale test on the most recent out-of sample data is beneficial for the decision making process because it helps to identify recent changes in market behaviour and also to identify data mining biases which overestimate the profitability of an indicator in a long selection/optimization of data. The duration of the test period should provide plenty of trades for the statistical analysis, but the starting date could be constrained by specific data requirements as it happened in this work. Conditional returns on entry signals are the building block of performance measurement on single positions. With them it is also possible to evaluate the performance of all aggregate positions of a signal when the number of holding days is being swept. The current measures are $\text{Profit}_{\text{trade}}$, P_f and RRR_{trade} and all of them provide an estimation of the indicator's profit potential on the tested market, but before considering this value as such, the first step towards profitability is a comparison to market performance. An overperformance of the original trades would suggest that the indicator has predictive power. This is first tested by comparing conditional returns to exact unconditional return distributions. The result is then checked with an approximated permutation test.

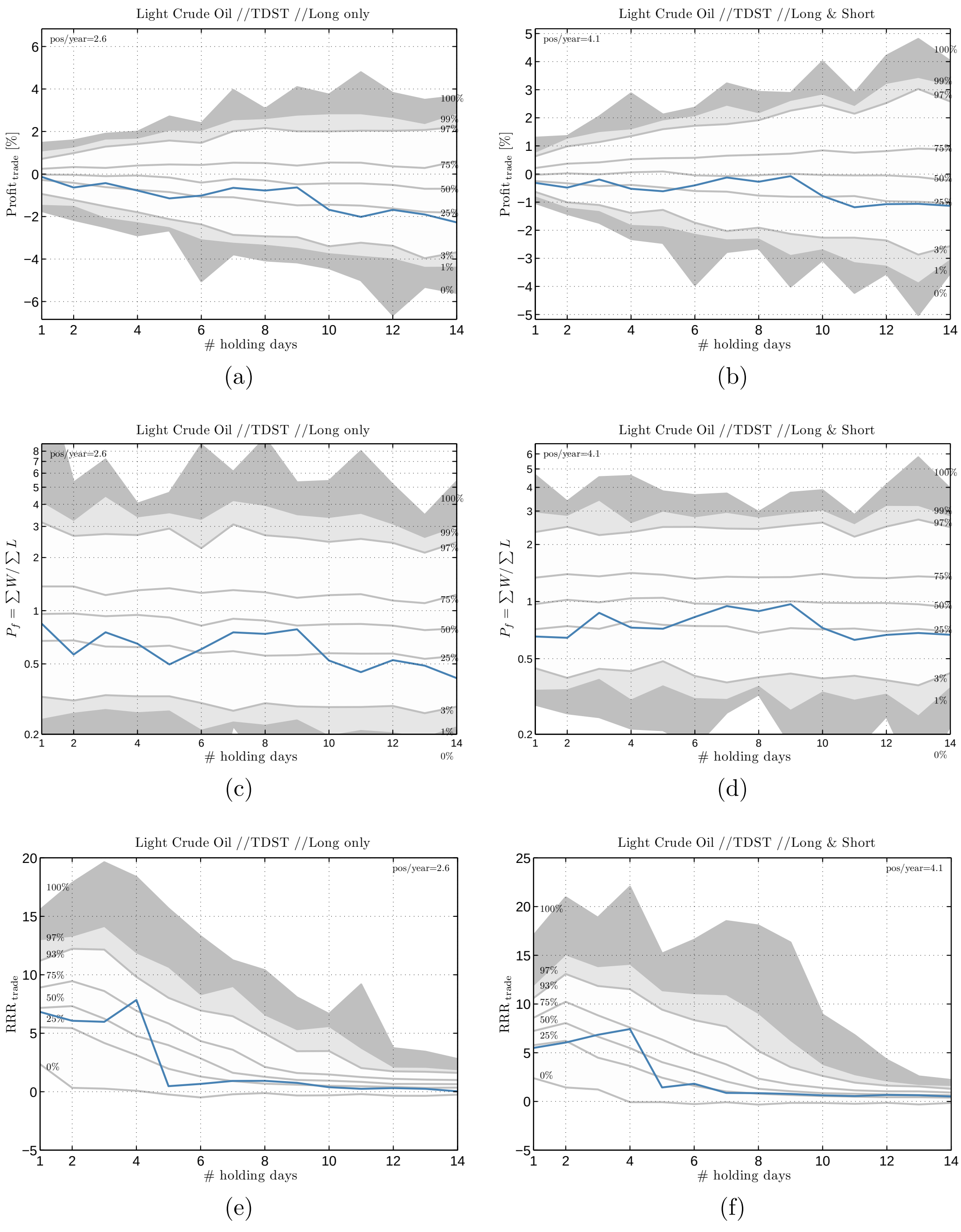


Figure 2.6: *All figures show Monte-Carlo permutation tests of the TDST signal on Light Crude Oil. For the null hypothesis, long, short and neutral positions are paired randomly with daily market returns. When the observed performance (in blue) reaches the light gray area, then there is a possible significance that the the null hypothesis H_0 should be rejected. Instead, if the observed performance reaches or passes the dark gray area, then there is statistical significance for H_0's rejection. As the previous conditional vs. unconditional performance comparision suggested (fig.2.4), the original signal underperforms compared to the profit possibilities offered by the market, but still not enough to reject the null hypothesis on the left tail. Profit possibilities are represented by the distribution sample that has been generated from permutations of the original signal. a) and b) show signal performance on* $\text{Profit}_{\text{trade}}$ *by means of a double-sided test. c) and d) show signal performance on P_f by means of a double-sided test. Lastly, e) and f) show signal performance on RRR_{trade} by means of a single-sided test.*

CHAPTER *3*

Financial Data

The first section of this chapter is an overview on commodity markets, followed by a more detailed section on futures markets. DeMark indicators are backtested on commodity futures and §3.1.2 provides a complete list of these financial instruments. The trading style chosen for the backtest requires continuous price curves. §3.2 explains how to generate and adjust these curves to then run the DeMark indicators.

3.1 Commodity Futures Markets

A commodity market is a market that trades raw materials. The market is physical when the product is delivered to the buyer, otherwise it is purely financial, for hedging, investment or speculative purposes. The market participants in these two subgroups can differ substantially. Physical traders place themselves in between producers and consumers. They have the logistical and physical infrastructure that, combined with a knowledge of the market and its participants, enables them to buy commodities with a discount from the producers to then deliver them to the customers with a premium. In parallel, financial instruments are used as a support to hedge market exposures. Financial settlement refers to commodities when financial instruments are secured by physical raw materials. Investors make profits by betting on the direction of the markets. Contrary to physical trading, financial investments need exposure to market prices to unlock profit opportunities. Besides an opposite approach towards market risk, financial instruments are a key element in both physical and financial trading. If it was not possible to hedge physical products on the financial markets, then every buyer should always need to find a producer to hedge the market price. Investors are willing to accept exposures to price fluctuations that physical producers, traders and buyers don't want and their participation in financially based commodity markets increases the liquidity of the related financial instruments, with benefits also on the physical side. For example, basis risk is the risk that the spot price for a physical product and the price of the financial instrument used for hedging may go in entirely opposite directions. A liquid financial instrument should limit this risk, at least from a price movement perspective.

Commodities use futures contracts as the basic type of exchange based financial instrument. Modern futures markets have their origin in Osaka where, in the 18$^{\text{th}}$ century, rice futures were being traded [21]. The main classes of raw materials are soft agriculturals, grains, energy products, precious metals and industrial metals. Exchanges are located in all continents, but the most active in terms of volume and history are in Europe and North America. Nevertheless, Asia has leading exchanges in China which plan to expand: as the world's top oil importer, China's *Shanghai Futures Exchange* (SHFE) is waiting Beijing's final approval to launch its crude oil futures contract with the hope that it will become a benchmark in Asia. As mentioned in the introduction, currently the world's main exchanges are the *Chicago Board of Trade* (CBOT) for grains, the *IntercontinentalExchange* (ICE) for softs and energy, the *New York Board of Trade* (NYMEX) for energy, the *Commodity Exchange, Inc* (COMEX) for precious metals and industrial metals, but metals are also traded on the *London Metal Exchange* (LME) for Europe and on the SHFE for Asia. Most of the contracts are denominated in US dollars. Each exchange sets its own rules for the quantity of material, the frequency and the expiry date of the contracts. Currently the most liquid contract is Soybean Meal, traded on the *Dalian Commodity Exchange* (DCE) with around 4 million open positions, followed by Light Crude Oil NYMEX, Natural Gas NYMEX, Brent ICE Europe and Corn CBOT, all with approximately 1.5-2 million open positions each.

Commodity markets are much broader than futures contracts, but these instruments still play a crucial role. Compared to the total volume of tradable commodities, liquid exchange based futures contracts cover a very small share. Gasoil, for example, is traded as a future on ICE Europe and covers the physical market for Low Sulphur Gasoil (10ppm Diesel) delivered in barges in the Amsterdam, Rotterdam, Antwerp (ARA) region. According to the contract the density is 0.845kg/litre in vacuum. A similar product with a different level of sulphur, or a more specific application, or a different density, or a different geographical scope, (e.g. the Mediterranean) will not have a dedicated liquid futures contract and will be traded over-the-counter (OTC). Despite these differences, the trade would typically use the Gasoil ICE futures contract both as a reference for pricing and as an instrument for hedging exposure. Regarding speculation, also exchange based options and swaps can be traded, but there is a difference with equity markets. Derivatives such as options and swaps use another derivative, a futures contract, as the reference price whereas equity markets use spot prices from stocks.

Switzerland plays a central role in commodity trading, in fact it handles more than 20% of global commodity trade according to the Swiss Bankers Association [4]. Despite the country's importance, only a small share of traded materials also crosses the country. In 2010, 3.6% of Switzerland's GDP came from commodity trading. The main hubs are Geneva, Zug and Lugano with the first being the world's biggest oil trading site. In 2013, according to Bloomberg, the world's largest oil trader (located in Geneva) traded 276 million metric tons of petroleum related products. More in general, the analysis published by the Swiss Federal Council [55] in March 2013 says that Switzerland has 35% of global crude oil trading and 60% of global metal trading. A stable legal and political framework, the historical expertise in the business and easy financing from the banks lead to a continuous increase in Swiss commodity trading activities during the last fifty years.

3.1.1 Different Trading Styles for Commodity Futures

A commodities futures contract is an agreement to buy or sell a specified quantity of a raw material at a future date and at a price agreed upon entering into the contract. We speak about *futures* when the contract is standardized, accessed through an exchange and with no physical counterparty. Otherwise, when the same agreement is done over-the-counter with a private counterparty the same agreement is called a *forward*. These financial instruments do not represent direct exposures to actual commodities, but they are a bet on future spot price and, by entering a futures contracts, an investor faces the risk of unexpected movements from the expected future spot price. These price deviations are generally unpredictable although good traders can avoid losses by timing the market. Contracts can start being traded at least one year before their expiry, but there is no common rule. All contracts from DCE and SHFE which are part of the backtest can be traded starting from 1 year before expiry. The other extreme comes from NYMEX contracts for Light Crude and Natural Gas which are active respectively for 7 and 6 years. All expiries are distributed along the timeline such that there is at most one closed contract per month. Tab. 3.1 gives a visual interpretation of the two dimensions of futures contracts: the *term structure* or *forward curve* (horizontal) and the *day-to-day prices* (vertical).

Table 3.1: *For each Close of Business (COB) there are multiple closing prices each one related to a futures contract with a different expiry date. This example shows closing prices ($/bbl) for NYMEX Light Crude on COB 07/02/2014. Along the same row the table shows the price values of the front contract M (March 2014) and of M+1 (April 2014), M+2 (May 2014), M+n (n=3, June 2014) and M+N (December 2022). N is the last of all n consecutive contracts that can be traded after the front one (the first to expire). Row d-1 shows closing prices for the prior day and d-2 show prices for day before d-1. The two dimensions of futures contracts are the term structure or forward curve (horizontal) and time* (vertical).

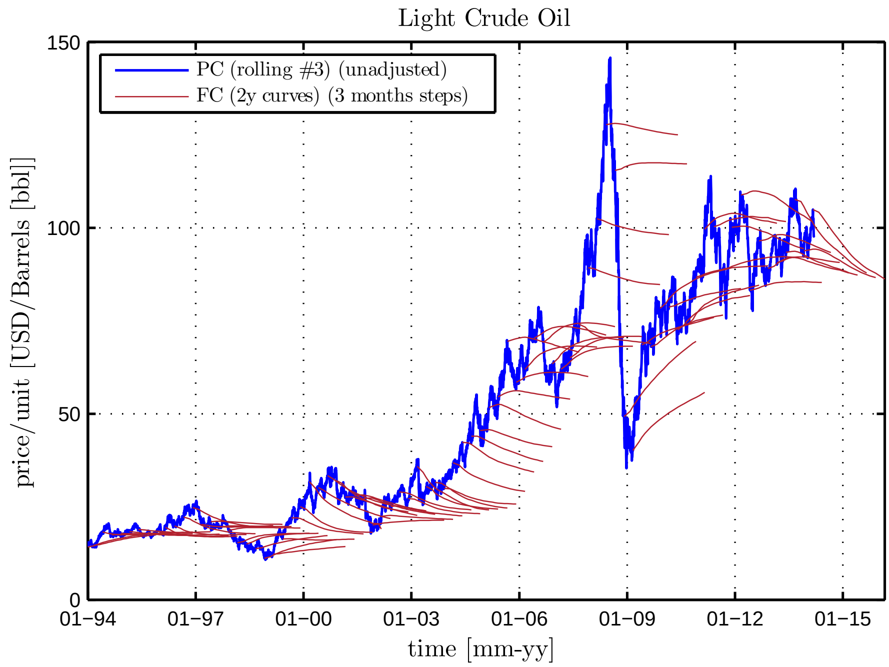
	M	**M+1**	**M+2**	**M+n**	...	**M+N**
⋮	⋮	⋮	⋮	⋮		⋮
d-2	97.38	96.76	96.02	95.22	...	76.02
d-1	97.84	97.32	96.66	95.94	...	76.17
07/02/2014	99.88	99.35	98.62	97.84	...	76.54

Forward curves show, on a specific time of date, market prices for futures contracts sorted by nearest expiry date. On 07/02/2014 Light Crude was in *backwardation*. This means that, on that date, contracts with an earlier exipry date had a higher market price compared to contracts expiring later in time. In other words, the forward curve had a negative gradient. The term structure is in *contango* when the gradient is positive. There could be multiple reasons behind Light Crude's backwardation. For example, lower global demand compared to the past years has caused an oil oversupply which is hedged by short futures. There is selling pressure on the futures markets which is not equally balanced by long risk taking investors.

The most straightforward way for investors to get exposure to market prices is with outright exposure. This means that long or short positions are entered just on one column of prices in Tab. 3.1, for example on the front contract M. This trading style makes profits when positions are properly matched with daily price movements of a contract. Given the upward price movements in Tab. 3.1, a long position would generate profits for two consecutive days. The blue curves in Fig. 3.1 are the continuous prices constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M. The backtest in Chapter 4 makes use of these blue curves both to compute daily market returns and to generate trades. This approach is the most common for technical traders who trade in commodity futures markets. Traditional commodity trading instead, makes a limited use of outright exposures. In fact, typical commodity traders speculate also without outright, which means that their daily profits do not depend on price movements of the blue curve. For completion, the end of this paragraph explains how speculation is carried on by commodity traders. Merchants who are completely dedicated to the trade of raw materials have a deeper understanding of how these markets behave when shocks of different types occur (supply, demand, geopolitical, macroeconomic, natural disasters). This knowledge helps them to understand how forward curves will evolve. As a consequence they get time-spread exposures by betting on relative movements of the curves. A long position on the Heating Oil front contract is a time-spread when it is coupled with a short position on the same market, but on the M+n contract. Instead, an understanding of market couplings is useful to enter long on market A and short on market B with equal contract expiries. This means that in Fig. 3.1 the bid is made on the spread between the two blue curves. This is called an inter-product exposure. In general outright, time and product exposure can be combined at the same time and the share also depends on the traded class of commodities.

3.1.2 List of Commodities

The performance of DeMark indicators is examined over 21 commodity futures markets and 10 years of data. Tab. 3.2 lists all the futures that were chosen for the backtest and provides additional information for each contract. The main driver in the selection was to have a balanced group of commodities that could include grains, softs, energy, industrial and precious metals.



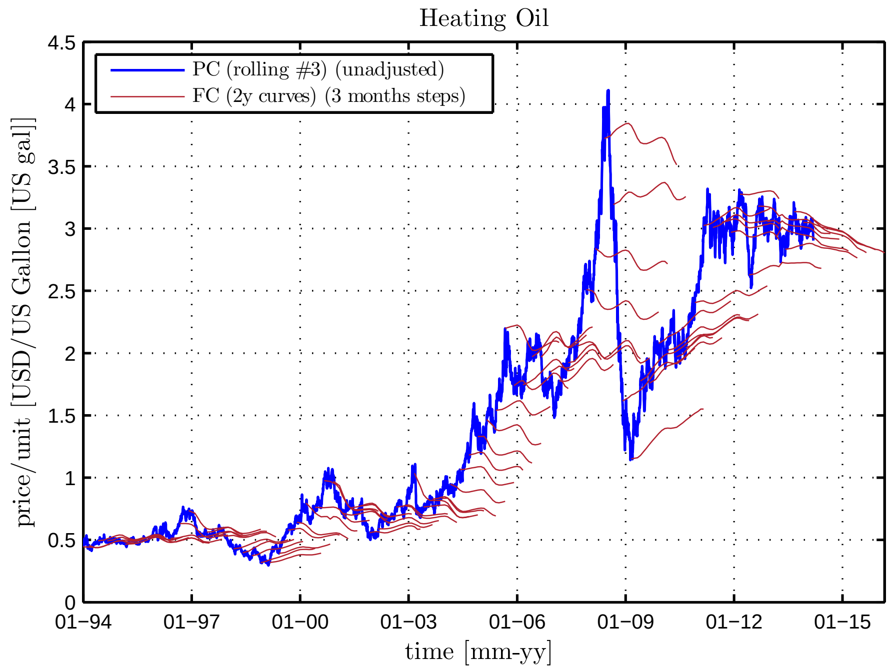


Figure 3.1: *Price evolution for Light Crude and Heating Oil on NYMEX. The blue curves are the continuous closing prices (PC) constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M. The red lines are forward curves. For each date there is a forward curve (FC) linked to it. One curve every three months was plotted. Furthermore, each curve was built using only the contracts that had expiry dates within 2 years from the date to which each forward curve is related. The most straightforward way for speculators to get exposure to market prices is with outright exposure. This means that long or short positions are entered on the blue curve. It is also possible to be exposed to market prices without outright risk (profits and losses don't depend on the behaviour of the single blue lines). It can be done by trading product differentials (inter-product risk, i.e. long Light Crude and short Heating Oil on front contracts), by betting on changes of the forward curves (time-spread risk, i.e long M+1 and short M+2) or by doing both at the same time. In this work the predictive power of DeMark Indicators is tested just on outright exposure.*

Changes had to be made within the single classes. Soybean Oil and Soybean Meal traded in CBOT had to be excluded from grains because the data was not available for the whole time period. Gasoline NYMEX had to be excluded from energy products for the same reason. It is advisable to add this futures contract in more energy related backtests because the list in Tab. 3.2 includes crude, natural gas and middle distillates, but still misses Light Ends (i.e. Gasoline) and Heavy Distillates. Industrial metals are missing Copper, Zinc and Aluminium from the London Metal Exchange (LME). DeMark indicators can generate signals only if daily bar charts have complete prices, but the data source did not provide the highest and lowest price of the day. For softs and precious metals no futures contract was excluded.

3.2 Rolling Futures Contracts

Previously, in §3.1.1, the blue curves in Fig. 3.1 were described as the continuous prices constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 was higher than that of M. The backtest in Chapter 4 makes use of these blue curves both to compute daily market returns and to generate trades. The goal of this section is to provide an overview on how to generate these curves.

The construction of a continuous price series using futures data is not straightforward since future prices are represented in contract months and for each trading day there are multiple prices available, each one coming from a different contract month (Tab. 3.1). Once a specific contract M+n is used to determine the prices of the day, there is no rule that tells when to "roll over" to the next contract M+(n+1) and how to merge the prices of each contract without having price jumps on the rolling date. Therefore, the problem of rolling futures contracts can be split in a timing and a price adjustment problem. Both aspects will be discussed in the next subsections.

3.2.1 Timing

A continuous price series of a futures contract must have one contract selected for each trading day, as in Fig. 3.2. In general it uses prices from contracts which are close to expiry (i.e. M, M+1, M+2) to capture both spot price and short term expectations. It is not possible to stick to one contract for the whole backtest period (i.e Light Crude, Jan. 2014). The first reason is that each futures contract can be traded only for a few years, not for the whole 10 years of the test. In addition, when a contract starts being traded it represents the most long term future price expectation of the forward curve and there is a weak dependency from price changes on the front contract (the best approximation to spot price). Lastly, the most liquid contracts are not the newest, but the oldest which are also the closest to expiry.

Data companies provide continuous price curves with the roll following the last traded day of the front contract. A more sophisticated method should anticipate the roll because contracts in their last weeks of life show abnormal volatility (Samuelson, [48]). Academic papers [10, 33] suggest to roll the front

Table 3.2: *List of commodity futures used for backtesting.*

	Name	**Type**	**Exchange**	**Ticker**	**Unit**	**Quantity**	**Currency**	**Expiry Month**[a]
1:	Wheat	Grain	CBOT	CW.	Bushels [bu.]	5’000	.01 $	{H, K, N, U, Z}
2:	Corn	Grain	CBOT	CC.	Bushels [bu.]	5’000	.01 $	{H, K, N, U, Z}
3:	Oats	Grain	CBOT	CO.	Bushels [bu.]	5’000	.01 $	{H, K, N, U, Z}
4:	Soybean Meal	Grain	DCE	DM.	Metric Tonnes [MT]	10	¥	All \ {G, J, M,V}
5:	Cocoa	Soft	ICE US	NCC	Metric Tonnes [MT]	10	$	{H, K, N, U, Z}
6:	Coffee C	Soft	ICE US	NKC	Pounds [lb]	37’500	.01 $	{H, K, N, U, Z}
7:	Sugar #11	Soft	ICE US	NSB	Pounds [lb]	112’000	.01 $	{H, K, N ,V}
8:	Cotton #2	Soft	ICE US	NCT	Pounds [lb]	5’0000	.01 $	{H, K, N, V, Z}
9:	Light Crude Oil	Energy	NYMEX	NCL	Barrels [bbl]	1’000	$	All
10:	Nat. Gas NYMEX	Energy	NYMEX	NNG	Million BTU [MMBtu]	10’000	$	All
11:	Heating Oil	Energy	NYMEX	NHO	US Gallon [US gal]	42’000	$	All
12:	Brent Crude	Energy	ICE EU	LLC	Barrels [bbl]	1’000	$	All
13:	Nat. Gas ICE	Energy	ICE EU	LNG	Thousand BTU [MBtu]	1’000	.01 £	All
14:	Gas Oil	Energy	ICE EU	LLE	Metric Tonnes [MT]	100	$	All
15:	Aluminium	Ind. metals	SHFE	SHA	Metric Tonnes [MT]	5	¥	All
16:	Copper COMEX	Ind. metals	COMEX	NHG	Pounds [lb]	25’000	$	All
17:	Copper SHFE	Ind. metals	SHFE	SCU	Metric Tonnes [MT]	5	¥	All
18:	Gold	Prec. metals	COMEX	NGC	Ounces [oz]	100	$	All
19:	Silver	Prec. metals	COMEX	NSL	Ounces [oz]	5’000	$	All
20:	Platinum	Prec. metals	NYMEX	NPL	Ounces [oz]	50	$	All
21:	Palladium	Prec. metals	NYMEX	NPA	Ounces [oz]	100	$	All

[a] Conventional letter codes used in tickers to specify delivery months. The exact expiry date can vary considerably depending on the type of commodity and the exchange. For example, the July 2014 contract expired in June for Energy products and in mid July for Agriculturals.

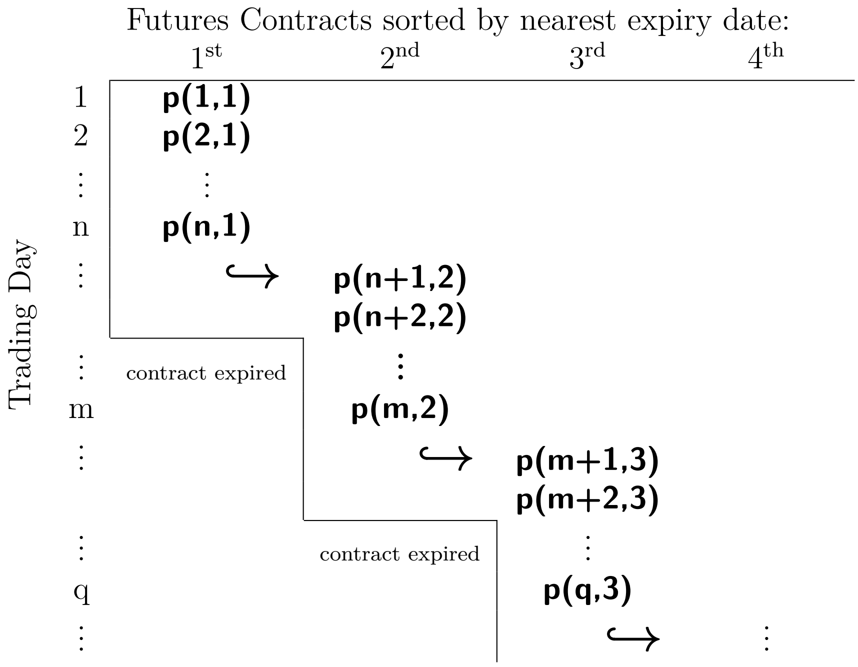


Figure 3.2: *Roll over of the price time series to contracts with later expiries. This creates a continuous path along the matrix of trading days vs. futures contracts (matrix also shown in Tab. 3.1) where exactly one contract is selected for each day. The next step is to decide the days on which the roll to the next contract should be done.*

contract 1 or 2 weeks before maturity, or on the first day of the delivery month, or, in alternative, to stay on the most liquid contract, for example by rolling from M+n to M+(n+1) as soon as the open interest (OI) on M+(n+1) is higher than in M+n. The daily OI is the total number of options and/or futures contracts that are not closed (or delivered) on a particular trading day. Similar solutions are also proposed by Thomson Reuters Datastream™ [46] with the addition of rolling methods that switch from M to M+1 based on weighted volumes. Backtests on Demark indicators use the rolling strategies in Tab. 3.3. It is not allowed for all strategies to roll back to the previous contract, i.e. from M+1 to M.

3.2.2 Price Adjustments

When each trading day has a contract linked to it, then continuous closing price series shows discontinuities on the days in which contracts are rolled. Data providers make local adjustments by weighting price values across multiple trading days, but there are other solutions, more computationally intensive, which fit better with backtests on DeMark indicators. These indicators use day-to-day price movements to generate entry signals and a price movement in the wrong direction could stop a signal before its completion. This is shown in Fig. 3.3 where the market is in a strong contango and the new contract has

Table 3.3: *List of rolling strategies available for backtesting. Rolling #1 is the way in which data providers build their continuous closing price curves. From #1 to #4 there is a decreasing dependence on the last traded day of the front contract M. In fact, #4 is not constrained to M, but it can always roll to a further contract if that has a higher open interest (OI).*

rolling #1:	from M to M+1, following the last traded day of M.
rolling #2:	from M to M+1, 10 days before M's expiry.
rolling #3:	from M to M+1, when first $OI(M) < OI(M+1)$.
rolling #4:	always on the contract with maximum OI.

always a higher value than the previous one. By looking at the time series this would seem a gain while in reality this is a loss because additional cash needs to be added to keep the same physical security (13$ and 15$ are paid to roll the contract). The correct performance measurement can be computed on the red curve, which should also change a loss into a profit when the forward curve is in backwardation. For the DeMark indicators, new long signals should not be completed when the price time series requires to add money to the contract. Therefore, entry signals are being generated on the blue curve, which is neutral to discontinuities from the rolling process.

There are multiple ways to adjust discontinuities on the continuous price time series [35, 36, 40, 44]. There is general consensus that prices in the old contract should be adjusted to prices in the new contract. In this way, continuous prices which are later in time should be less affected by adjustments. On the other hand, this backward-adjustment is more computational intensive than a forward adjustment because a change on a single roll over requires additional changes on all previous trading days. Fig. 3.3 uses backward-adjustments, this is why trading days have negative signs and absolute values increase from the right to the left of the graph. △ is the price adjustment on the roll such that:

$$\triangle = PO_d - PC_{d-1}. \tag{3.1}$$

To generate a similar curve to the blue curve, which is neutral to rolling, a quantity △ needs to be added to all the prices (PC, PO, PH, PL) of all the previous trading days. For the red performance curve this quantity needs to be added two times.

Instead of adding a fix quantity △, the proportionally adjusted method uses the ratio ρ defined as:

$$\rho = \frac{PO_d}{PC_{d-1}}. \tag{3.2}$$

where d is the day following the roll.

To generate a similar curve to the blue curve, which is neutral to rolling, prices (PC, PO, PH, PL) of all the trading days prior to the roll need to be multiplied by the quantity ρ. Ratios have the advantage that negative prices are not possible by construction. On the other side, there are bigger data fluctuations which discourage the use of ρ in favour of the △ approach.

For the backtests the data has been backward-adjusted with $\triangle$ quantities. §3.2.3 explains some of the advantages of this approach over the proportionally adjusted method.

Lastly, the Perpetual Method uses progressively smaller percentages of the current contract and larger percentages of the new contract. Anyhow, this method is more suitable for statistical analysis rather than trading activities where the real values of the market are necessary.

3.2.3 From Rolled Prices to Returns

The backward-adjusted method using fixed adjustments generates negative prices on the trading time series and the performance time series (respectively blue and red curve in Fig. 3.3) for Soybean Meal, Copper COMEX and Copper SHFE. In other cases, adjusted time series are close to the null price. This is not a problem for the trading time series because DeMark indicators only use relative prices to build up entry signals. The ultimate goal of the performance time series is to generate daily market returns. The daily return at day t is defined as

$$r_t = \frac{PC_t - PC_{t-1}}{PC_{t-1}}. \tag{3.3}$$

The red curve cannot be used directly to compute daily returns because closing prices (PC) can be negative or near the zero value and this would distort the returns. A simple solution is to use the red curve to compute the nominator while the denominator uses closing prices from the unadjusted continuous price curve (black curve in Fig. 3.3). In fact, the red curve shows correct changes in relative prices, while the black curve refers to the absolute price levels of the market.

3.3 Summary

This chapter has explained how continuous daily returns are computed from discontinuous commodity futures markets. With these returns it is possible to measure the performance of the DeMark indicators. The first section gave a general introduction to commodity markets and explained the fundamental link between futures contracts and commodity markets. These financial instruments can be traded based on continuous price curves or on forward curves, but only continuous prices curves are used in this work to create market exposure. Given this choice, the chapter ends with a detailed explanation on how to create continuous price curves and how to adjust them to compute daily market returns. The task can be divided into 3 sequential steps, like in Tab. 3.4.

Table 3.4: *From discontinuous commodity futures market prices to daily returns.*

Timing	each trading day is linked to a specific contract;
Adjustment	creation of *trading* and *performance* price-series;
Returns	from *unadjusted* and *trading* price-series to daily returns.

A further development of the backtest, based on this chapter, would be to incorporate time-spreads on forward curves and interproduct-spreads on continuous price curves as different sources of market exposure.

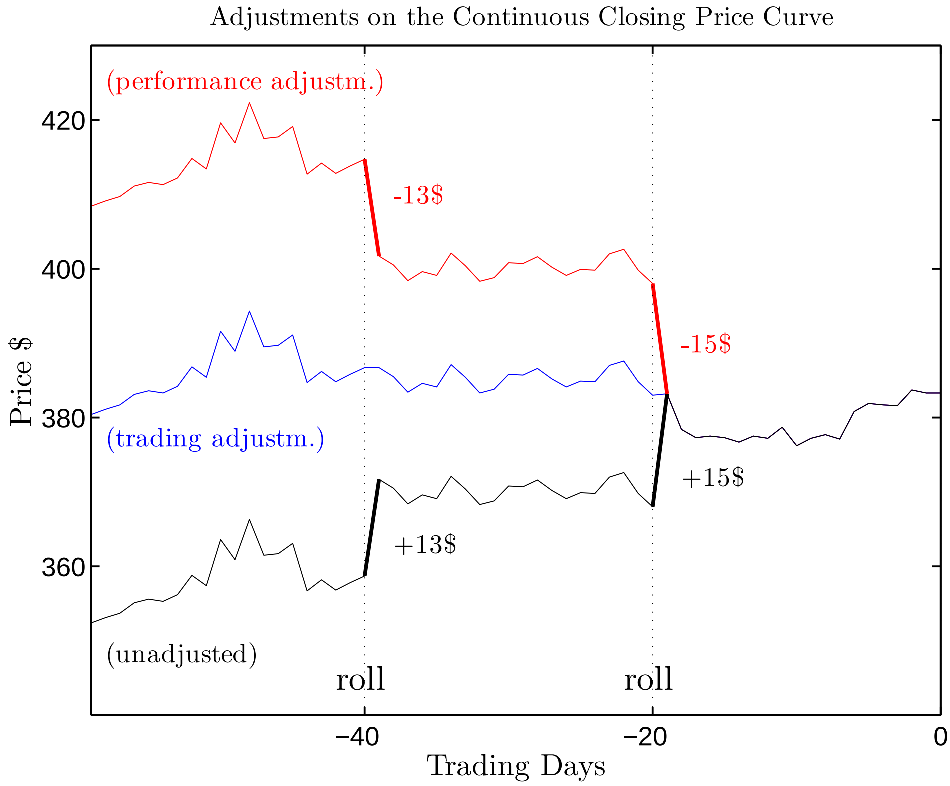


Figure 3.3: *The unadjusted (black) curve is the continuous closing price curve that has been generated by one of the rolling strategies in Tab. 3.3. Trading Day 0 is the last trading day of the time period and the plot goes backwards in time, from right to left. This curve shows discontinuities on the days in which contracts are rolled. In this example, this market is in contango. This means that each time that a contract is rolled to the next there is a price increase. The crucial point is that this is a loss for those who own the current contract and have to switch to the next contract because a contract with less value is sold and a contract with a higher value has to be bought, but the underlying physical asset stays the same. The time series on which performance is measured should show these rolls as losses of 13$ and 15$ and this is what the red curve is showing. In signal generation, a position that requires money should not be preferred. The curve used for generating signals is the blue one, which is neutral to discontinuities from the rolling.*

CHAPTER *4*

Backtesting

Each of the previous chapters focuses on a different aspect of the backtest. Chapter 1 examines the three DeMark indicators that have been chosen for the backtest. In Chapter 2 different ways for evaluating the performance of DeMark entry/trading signals are discussed. If indicators generate signals that outperform the market, then those indicators are informative. In other words, they have predictive power, and this is tested by means of Monte-Carlo Permutation Tests. Financial data is discussed in Chapter 3. Commodity futures can be traded based on continuous price curves or on forward curves, but only continuous prices curves are used in this work to create market exposure. As soon as daily market returns are computed from continuous price curves it is possible to run the tests. A sequence of returns characterizes an entry position when the number of holding days is given. If this number is swept instead of being fixed, then conditional returns can be analysed.

This Chapter focuses on the results. §4.1 serves as a summary on the predictive power of each DeMark indicator. A more holistic approach is provided in §4.2. For only long, only short and a combination of long and short positions DeMark indicators can give statistically significant predictive power on some of the 21 commodity futures markets analysed in the backtests. When this happens, the next step is to quantify the profit potential. This is done by plotting a measure of the stability of predictive power over the holding days versus the expected profit per trade. This framework provides examples where indicators match well with specific markets. The next section (§4.3) focuses on three of these examples: one for long only signals, one with short only signals and one with both long and short signals. For each of them the idea is to check if other performance measures (Profit Factor and Risk Return Ratio per trade) provide conclusions which are coherent with the ones based on Profit per trade. Lastly, §4.3.1 studies for the same examples the impact of different rolling strategies on the results.

4.1 Predictive Power, An Overview

Table 4.1: *A summary of the backtest setup.*

Backtesting Period	
start:	1/1/2004
end:	1/1/2014
duration:	10 years
trading days:	2610

DeMark parameters	
m	9
n	4
Sequential & Combo:	
p	13
q	4
k	8
recycle 1	on
recycle 2	on
recycle 3	on
recycle 4	on
aggressive version	on
Entry Strategy:	
conservative	on

Rolling Strategy	
Main:	#3
Additional:	#1
	#2
	#4

Quality Measures	
Main:	$\text{Profit}_{\text{trade}}$
Additional:	P_f
	RRR_{trade}

List of Commodities	
1:	Wheat
2:	Corn
3:	Oats
4:	Soybean Meal
5:	Cocoa
6:	Coffee C
7:	Sugar #11
8:	Cotton #2
9:	Light Crude Oil
10:	Natural Gas NYMEX
11:	Heating Oil
12:	Brent Crude
13:	Natural Gas ICE
14:	Gas Oil
15:	Aluminium
16:	Copper COMEX
17:	Copper SHFE
18:	Gold
19:	Silver
20:	Platinum
21:	Palladium

Significance Test	
Permutation Test	
Parameters:	
# of holding days	1-14
# of permuted signals	400
p-values	5%, 10%
Confidence intervals:	
Clopper-Pearson	
Conover (1999)	

Each signal generated from DeMark indicators can be either long, short or out of the market during the backtested period. As discussed in §2.2.2, the entire signal (comprising all entry positions and a fixed number of holding days on a specific commodity futures market) can be evaluated by an aggregated measure of quality. Here the measure is $\text{Profit}_{\text{trade}}$. The *null hypothesis* H_0 assumes that the signal's long, short and neutral positions are paired randomly with daily market returns. In other words, the assumption is that the DeMark indicators don't have the capability to match long positions with positive daily returns and short positions with negative daily returns. The *alternative hypothesis* H_A

(which is the one that the test would like prove by rejecting the null hypothesis) supports the idea that the current pairing improves performance beyond what could be expected from randomness.

To test the null hypothesis, the trading signal needs to be permuted. If only a small fraction of the permuted results exceeds the $\text{Profit}_{\text{trade}}$ performance of the original signal, then the indicator (which originated the signal) gives significantly better performance compared to what sheer luck could do. In other words, the indicator has predictive power on the tested commodity futures market. It might also happen that only a small fraction of permuted results is inferior to the performance of the original signal. In this case the indicator is still informative. For example, let's assume that the original signal has a negative mean $\text{Profit}_{\text{trade}}$ for a fixed number of holding days, inferior to the performance of all the permuted signals. The sequence of daily returns determined by the signal's timing is not driven by luck, but, to make it a profitable signal, the directions of the trades need to be reversed (from long to short or/and from short to long). In a nutshell, there is predictive power when $\text{Profit}_{\text{trade}}$ performance of the original signal is either on the left or on the right tail of the performance distribution generated by permuted signals, but the direction of the trades needs to be changed when the signal is on the left tail fo the distribution. The predictive power of each indicator is discussed in the following subsections. The summarized results of the permutation tests are available in Tab. 4.2, 4.3 and 4.4. For each indicator the results are split on only long, only short and the combination of long and short entry signals. The reported values are: the number of trades, the % of significant holding days out of the 14 days that follow the trade entry (it will be described from §4.2 as the *stability* of predictive power), and $\text{Profit}_{\text{trade}}$. Profit per trade is computed only when there is at least one holding day value that provides statistical significance in the permutation test. The reason is that $\text{Profit}_{\text{trade}}$ values when there is no statistical significance can be misleading because their use is not supported by overperformance compared to the market. Supposing that there are 3 out of 14 holding days values for which an entry signal shows predictive power, then the $\text{Profit}_{\text{trade}}$ value will be the average of the three $\text{Profit}_{\text{trade}}$ values on the three significant holding days. For example, ff an indicator shows overperformance with a time delay, its average $\text{Profit}_{\text{trade}}$ should not be penalised by lower $\text{Profit}_{\text{trade}}$ values referring to lower holding days for which the indicator is also not providing statistically significant overperformance. For signals that have less than 1 trade per year, $\text{Profit}_{\text{trade}}$ shows only the direction of the profit, but not the value. An expected value based on a low number of trades could be far from the value measured on the complete population of past, but also future trades (which cannot be observed at present).

4.1.1 Sequential

For all the commodity futures apart from Platinum the indicator has shown statistically significant predictive power for either long or short entry positions. Out of the 3 tested indicators Sequential is the one that exhibits predictive power more often. The majority of significant long Sequential entry positions

Table 4.2: *Sequential's Performance on each Commodity Futures Market. The Market ID refers to Tab. 3.2.*

	Long			**Short**			**Long & Short**		
ID	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]
1:	2.0			1.6	21%	+4.9	3.6	14%	+2.7
2:	2.2			1.7	14%	-2.2	3.9		
3:	1.4	21%	-3.2	1.6			3.0		
4:	1.5			2.5	14%	+1.2	4.0		
5:	1.7			2.1	64%	+3.1	3.8	36%	+2.6
6:	2.1			1.5	21%	-1.6	3.6		
7:	2.4			1.9	14%	-2.8	4.3		
8:	2.2	7%	+1.0	1.7			3.9		
9:	1.6	7%	-3.3	2.8	7%	+1.4	4.4	14%	+0.8
10:	2.8	7%	+0.9	1.6			4.4	7%	-3.0
11:	1.4	50%	-4.5	2.6	7%	+1.3	4.0		
12:	1.2	14%	-4.3	2.9			4.1		
13:	3.1	36%	-12.0	1.1			4.2	29%	-6.8
14:	1.1	50%	-4.4	2.5	7%	+0.9	3.6	7%	+0.8
15:	2.0			2.0	86%	-1.5	4.0	86%	-1.0
16:	1.7	29%	-2.9	2.9			4.6	21%	-1.9
17:	1.5			2.6	29%	-4.1	4.1	21%	-2.5
18:	0.8			2.7	7%	+0.7	3.5		
19:	1.0			2.4	14%	+1.1	3.4	7%	+0.8
20:	1.2			2.2			3.4		
21:	1.3			2.4	21%	-1.7	3.7	21%	-1.4

Table 4.3: *Combo's Performance on each Commodity Futures Market. The Market ID refers to Tab. 3.2*

ID	**Long** Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	**Short** Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	**Long & Short** Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$
1:	1.1	14%	+2.0	0.6			1.7	7%	+1.7
2:	1.0	100%	+5.0	0.8			1.8	14%	+3.0
3:	0.9	7%	(+)	0.6	14%	(+)	1.5	7%	+1.0
4:	0.8			1.3	14%	-1.6	2.1	7%	-1.1
5:	0.5	7%	(-)	0.8	7%	(+)	1.3		
6:	0.9			0.7	50%	(+)	1.6	29%	+3.5
7:	1.2			1.0	86%	-4.6	2.2	93%	-2.3
8:	0.9			0.8			1.7		
9:	0.4	36%	(+)	1.3			1.7	7%	+1.0
10:	1.3	7%	+2.3	0.5			1.8	7%	+2.7
11:	0.6	29%	(-)	1.1	21%	+2.1	1.7		
12:	0.6	50%	(-)	1.2			1.8		
13:	1.3	43%	-11.6	0.3			1.6	50%	-7.6
14:	0.6			1.4	7%	-1.0	2.0		
15:	0.6			0.6			1.2		
16:	0.5	7%	(-)	1.1			1.6	21%	-2.4
17:	0.4	14%	(-)	1.5			1.9	14%	-3.2
18:	0.1	29%	(+)	1.1	14%	+1.2	1.2	14%	+1.6
19:	0.3	7%	(-)	1.3	7%	+1.4	1.6	14%	+1.3
20:	0.4			1.2			1.6		
21:	0.4			1.1	7%	+3.2	1.5		

Table 4.4: *TDST's Performance on each Commodity Futures Market. The Market ID refers to Tab. 3.2*

ID	**Long**			**Short**			**Long** & **Short**		
	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$ [%]	Pos/Year	Significant holding days	$\text{Profit}_{\text{trade}}$
1:	2.1	14%	+3.8	2.8			4.9	29%	+3.1
2:	1.6	7%	-1.2	3.2			4.8		
3:	3.2			2.8			6.0		
4:	2.4	79%	+3.1	1.2			3.6	50%	+1.4
5:	2.4			2.3			4.7		
6:	2.2			2.8			5.0		
7:	1.6			2.3	7%	+2.9	3.9		
8:	1.8			3.0			4.8		
9:	2.6			1.5			4.1		
10:	1.1			2.3			3.4	7%	+4.3
11:	2.7			1.6	7%	-1.4	4.3		
12:	3.0			1.2			4.2	14%	+1.6
13:	3.1	43%	+2.1	1.1	50%	-2.7	4.2		
14:	2.2	7%	-0.6	1.3			3.5		
15:	1.8	21%	+1.0	2.3			4.1		
16:	2.0	7%	+3.2	1.6			3.6	21%	+2.1
17:	1.6			1.8			3.4		
18:	3.4			1.4	29%	+1.2	4.8	7%	+0.8
19:	3.0			2.1			5.1		
20:	3.1			3.0			6.1		
21:	3.4			2.4			5.8		

is focused on energy. Yet, it is interesting to notice that long Sequential on energy is more a trend detector than turning point detector and the market is more likely to continue the trends rather than reverse them. On the other side, short Sequential entry positions show most of their predictive power for all the remaining commodity classes. Also in this case Sequential is more a trend detector. A conclusion is that Sequential identifies trends which, for some classes like energy, are more likely to continue rather than to reverse. The indicator shows significant predictive power for both long and short trades on Light Crude, Heating Oil and Gas Oil. In addition, these three products have a similar number of trades per year. This should not be a surprise because these markets are strongly linked. Furthermore, Heating Oil and Gas Oil, which represent respectively US and EU Diesel contracts, have interestingly the same number of holding days for which predictive power is significant (50% stability for long signals and 7% stability for short signals).

4.1.2 Combo

The first observation is on the number of trades, which are on average 30-40% fewer compared to Sequential. This is a limitation when it comes to deciding which signal to trade on which markets. Long Combo entry signals show predictive power not just for energy, but also for grains, Copper, Silver and Gold. Like for Sequential, also this indicator can be a trend follower. Before using it on real-time markets it should be clarified which are the markets for which the trend is expected to continue. Short positions are informative for softs and, again, for precious metals. Unlike for Sequential there are more cases in which the signal is predictive for both long and short signals, for example with Gold and Silver.

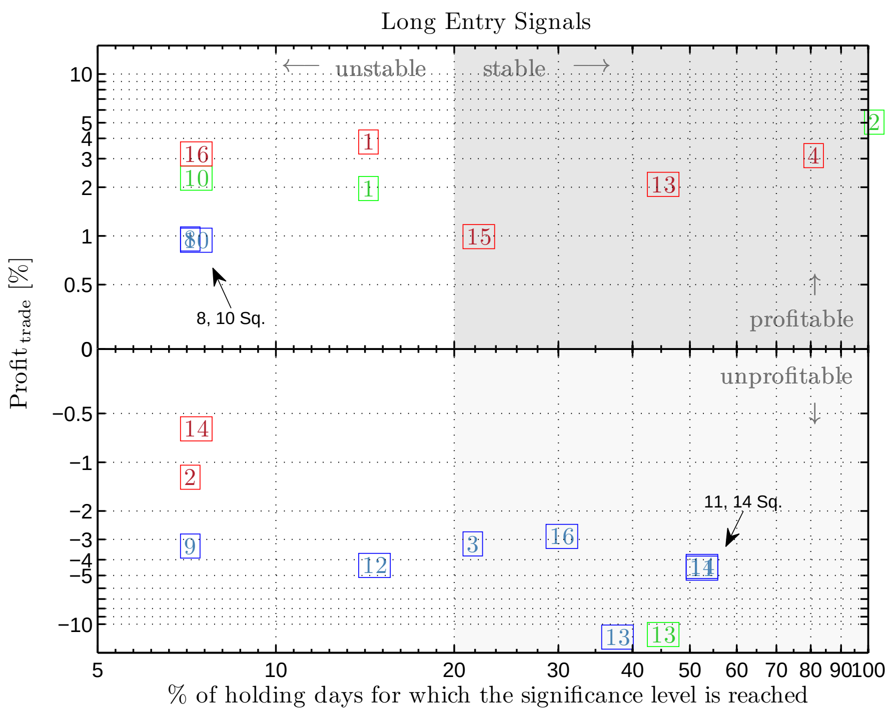
4.1.3 TDST

For most of the commodity futures there is no predictive power. Yet the indicator provides the highest number of trades and is predictive on a few markets. Long TDST signals are informative in particular for grains while for combinations of long and short positions there is statistical significance for Wheat, Soybean Meal, Natural Gas NYMEX, Brent, Gold and Copper NYMEX. Compared to the other signals, the direction of the TDST (trend following) entry signal captures correctly the direction of the market. This is always true when both long and short positions are possible in the signal.

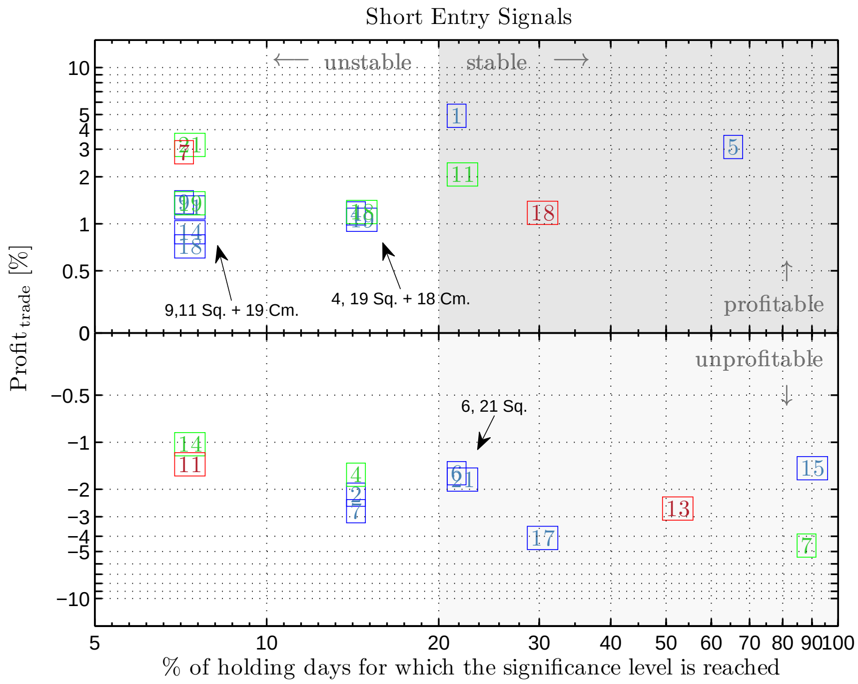
4.2 Predictive Power versus Profit Potential

Stability of predictive power over the holding days is the main criteria for choosing an indicator on a specific futures market. *Stability* in this context means that there should be predictive power not just for one fixed number of holding days, but ideally for all the swept holding days that follow the entry signal. Stability is measured by the % of holding days that show statistically significant performance of the pairings between the trades and the daily market returns. It goes from a minimum of 0% to a maximum of 100% (best case).

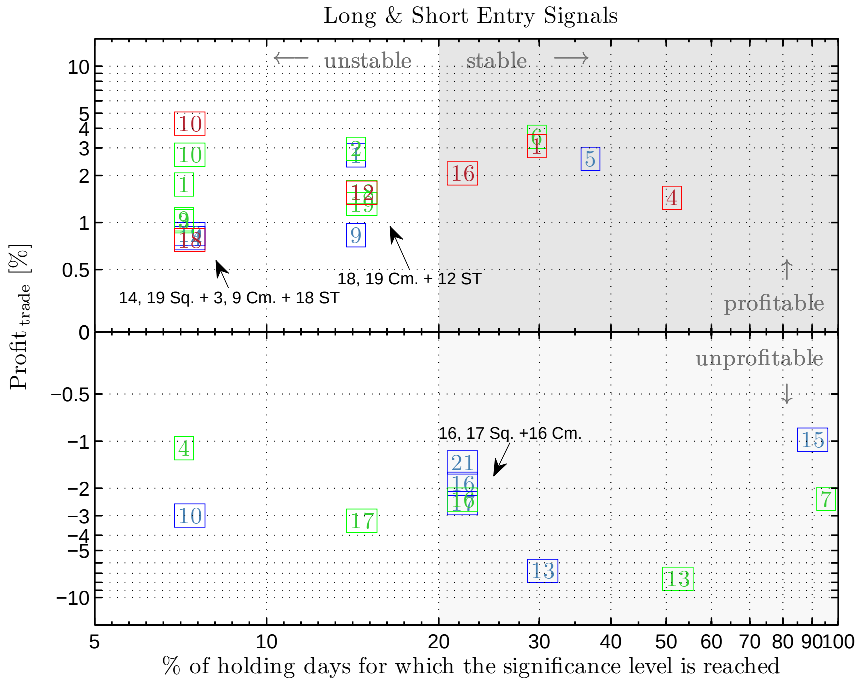
Predictive power is being tested by the permutation test which uses $\text{Profit}_{\text{trade}}$ as the aggregated measure of quality for the entire signal. However, the $\text{Profit}_{\text{trade}}$ value is not guaranteed when predictive power is confirmed by statistical significance because profit quantification is not part of the test. Anyway, it is important in practice to assign a number to the profit potential because the final choice of using the signal considers both the stability of the indicator's predictive power and the profit potential. The computation of the average of $\text{Profit}_{\text{trade}}$ values that belong to statistically significant holding days seems a simple, but effective way to limit the data mining bias which is an overstatement of the expected performance of an indicator based on the level of performance that allowed the system rules to be selected in the optimization process [3]. If the choice to trade a signal on a specific market is based solely on the maximization of the observed profit potential, then it is likely that the measured performance will be lower compared to expectations. The best indicator for a commodity market should jointly maximize the stability of predictive power over the holding days (main priority) and the profit potential. These two dimensions are shown in Fig. 4.1.



(a)



(b)



(c)

Figure 4.1: *Stability of predictive power over the holding days is the main criteria for choosing an indicator on a specific futures market. Stability in this context means that there should be predictive power not just for one fixed number of holding days, but ideally for all the swept holding days that follow the entry signal. Stability is measured by the % of holding days that show statistically significant performance of the pairings between the trades and the daily market returns. It goes from a minimum of 0% to a maximum of 100% (best case). For the commodity futures numbered from 1 to 21 this graph shows the % of holding days for which the significance level is reached versus the profit potential, i.e. the averarge* $\text{Profit}_{\text{trade}}$ *(for further details refer to 4.1). The optimal case would be to have a stability of 100%, which means that the indicator overperforms the market independently of the number of holding days, and* $\text{Profit}_{\text{trade}}$ *should have a value which is the highest possible (either positive or negative). In the figure, the red colour represents TDST (ST), green is for Combo (Cm.) and blue is for Sequential (Sq.), while the most interesting cases, according to the framework, are in the shaded area. In Fig. a) indicators are evaluated only for their long entry positions, in b) only for their short entry positions and in c) for both long and short.*

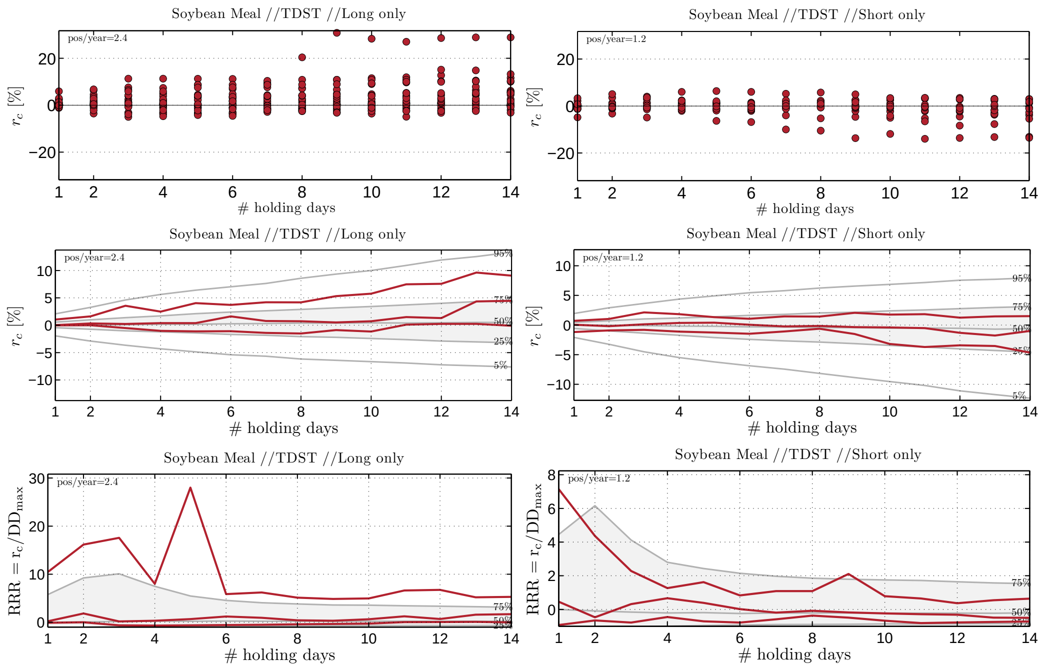
4.3 Examples

The framework in Fig.4.1 is a useful overview on possible couplings of indicators with futures markets. Fig. 4.1a shows the performance of only long entry positions on the markets, Similarly, Fig. 4.1b shows performance for only short entries, while Fig. 4.1c considers both long and short positions on a single market. For each of them an indicator coupled with a market will be further analysed. The choices are: long TDST positions on Soybean Meal, short Sequential positions on Cocoa and both long and short Sequential positions on Natural Gas ICE. These combinations have been chosen due to their high stability of predictive power together with high $\text{Profit}_{\text{trade}}$ values.

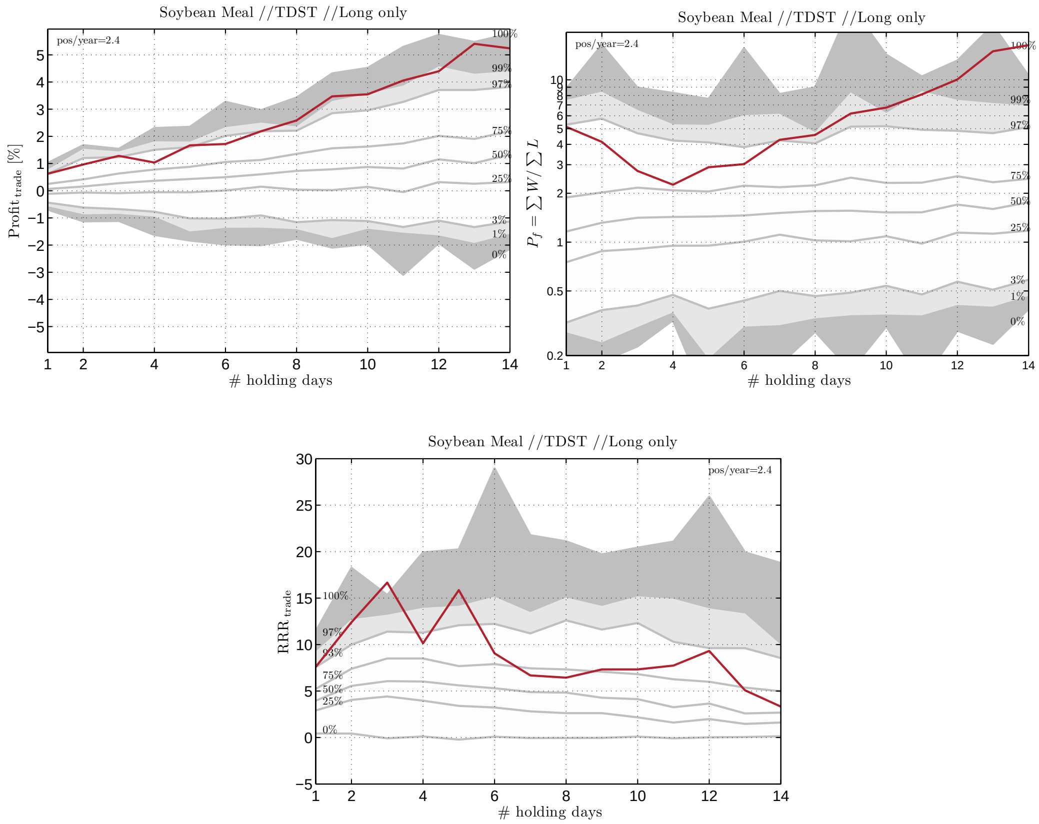
TDST on Soybean Meal is as an example of good performance on long positions. The first step of the analysis is to look at Section A in Fig. 4.2a. The quantiles of conditional returns are steadily above the corresponding quantiles generated from unconditional market returns. The same is true for conditional risk-adjusted returns (using RRR_{trade}). This suggestions of overperformance compared to the market are confirmed by the permutation test on long entry positions. The observed values of $\text{Profit}_{\text{trade}}$ and P_f are steadily in the statistical significance area, which means that long TDST entry signals have predictive power on Soybean Meal. RRR_{trade} is in the region of significance too, but less frequently. The same kind of thinking should be applied also to the the other examples. Sequential on Cocoa is shown as a successful example for short positions. In addition, Sequantial and Cocoa seems a good match also for long positions. Section A in Fig. 4.2b shows that for both long and short positions conditional returns (simple and risk-adjusted) overperform the market, but there is not confirmed by statistical significance. Permutation tests on short positions using $\text{Profit}_{\text{trade}}$ and P_f are coherent the results in Section A, but RRR_{trade} is not although the observed values are always above the market's median performance. Finally, let's consider Sequential on Natural Gas ICE. This case has been chosen as an example where long and short positions, combined, exhibit predictive power. Conversely to the other cases, this example shows statistically significant market *underperformance* which can be still used in practice to generate profits once it is clear that positions have to be reverted (from long/short to short/long) before entering the market. In Section A long positions steadily underperform compared to the market, while short positions slightly overperform. All the combined positions underperform in the permutation test mainly because long Sequential entry positions are much more frequent compared short ones. This means that the overall performance depends more on the result of long positions. The permutation test shows statistical significance for a small number of holding days. Nevertheless, there is statistical significance for high values of holding days. When a position is entered, it takes some time for the indicator to show its predictive power, which is still good in practice because the trader has time to enter the trade without loosing profit potential. The ideal case is when there is statistical significance for all the holding days because it means that the indicator has predictive power independently from the number of holding days of a trade.

TDST Entry Signals on Soybean Meal

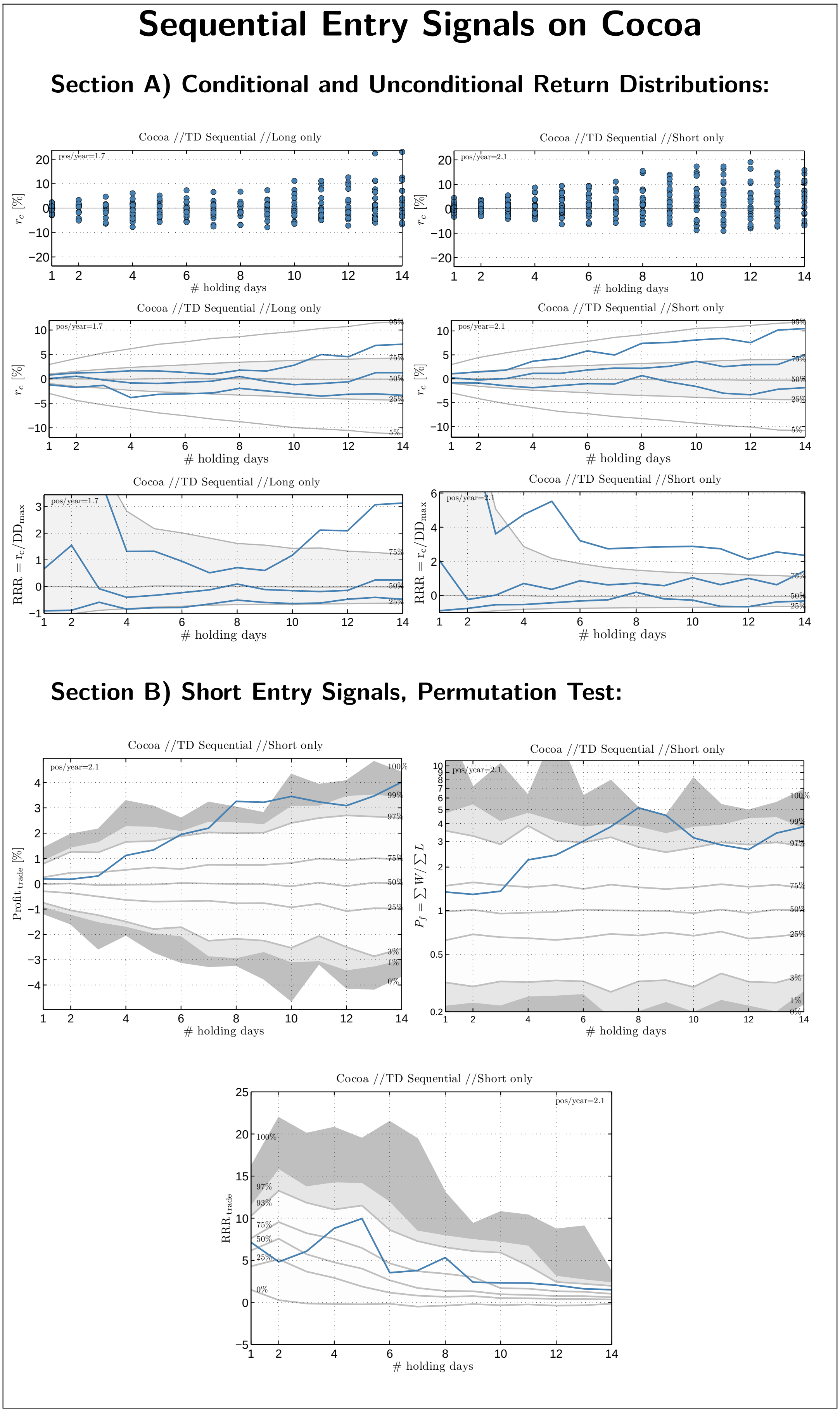
Section A) Conditional and Unconditional Return Distributions:



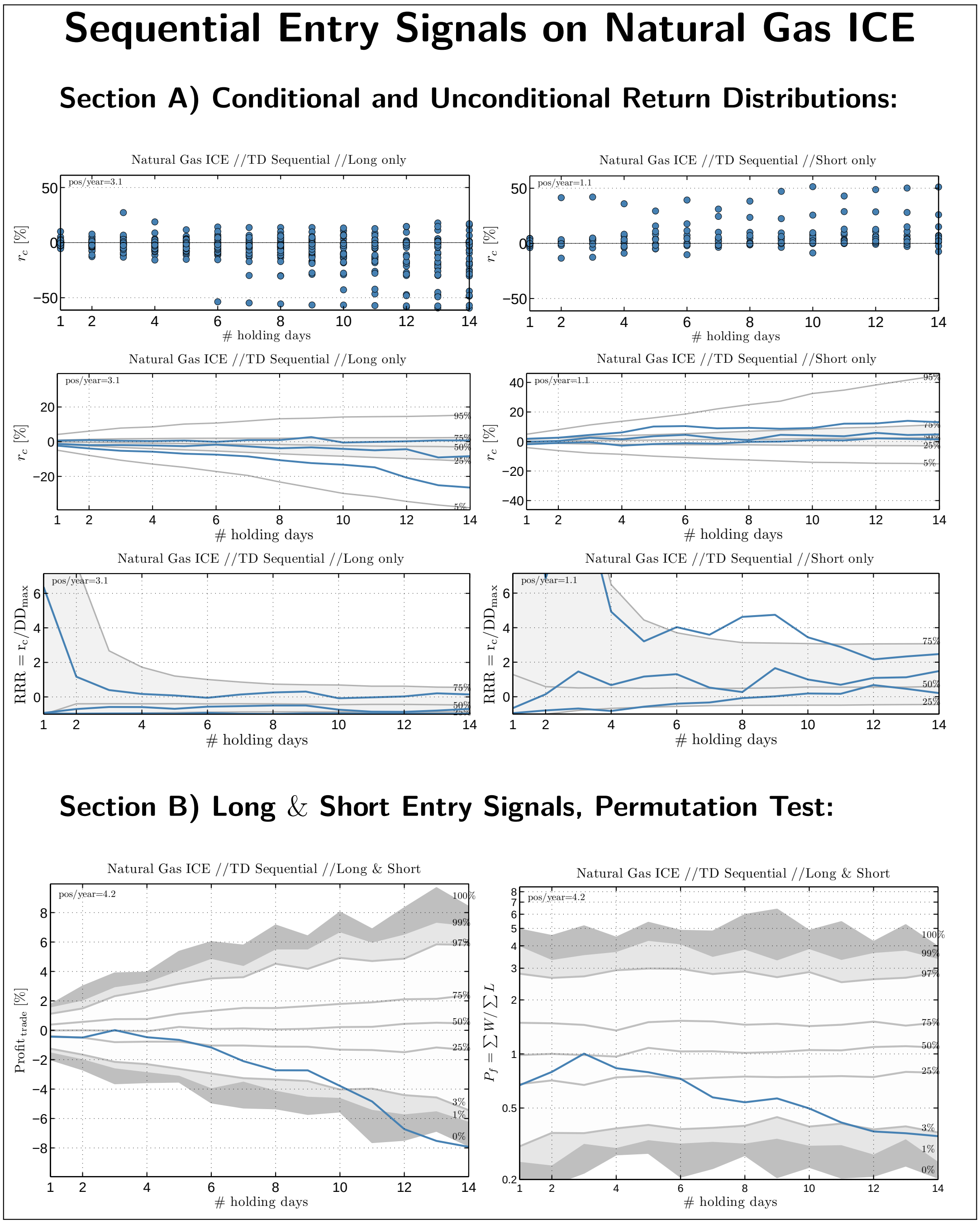
Section B) Long Entry Signals, Permutation Test:



(a)


Sequential Entry Signals on Cocoa
Section A) Conditional and Unconditional Return Distributions:
Cocoa //TD Sequential //Long only
Cocoa //TD Sequential //Short only
pos/year=1.7
pos/year=2.1
r_c [%]
RRR = r_c/DD$_{max}$
# holding days
Section B) Short Entry Signals, Permutation Test:
Profit$_{trade}$ [%]
$P_f = \sum W / \sum L$
RRR$_{trade}$
100%
99%
97%
93%
75%
50%
25%
3%
1%
0%
95%
5%

(b)



(c)

Figure 4.2: *Performances of a) TDST entry signals on Soybean Meal, b) Sequential entry signals on Cocoa and c) Sequential entry signals on Natural Gas ICE. Section A is the same for all the examples. The left/right side shows conditional returns on long/short entry positions. If we focus on the left plots, the first is a scatter plot of all conditional returns, the second compares the conditional return distribution (25%, 50% 75% quantiles) to the unconditional return distribution (shaded area) and the third is similar to the second, but instead of simple returns the distributions use risk-adjusted returns. Section B shows the results of the permutation test by using three different aggregated quality measures fo the entire signals:* Profit$_{\text{trade}}$, P_f, *and* RRR_{trade}. *In a) the permutation test focuses only on long entry signals, in b) it focuses on short entry signals and in c) it considers both long and short entries. In c) there is no permutation test on* RRR_{trade} *because the daily returns captured by the positions are mostly negative and* RRR_{trade} *is not a symmetrical measure as* Profit$_{\text{trade}}$ *and* P_f *are. In this example, the idea is to use the indicator to identify trends rather trend reversals.*

4.3.1 Sensitivity to Different Rolling Strategies

All the continuous price curves used in the backtest are based on the following rolling strategy: the front contract M is rolled to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M. To review the "roll over" concept and to check the other rolling strategies refer to §3.2.1. A continuous price series of a futures contract must have one contract selected for each trading day, as in Fig. 3.2. In general it uses prices from contracts which are close to expiry (i.e. M, M+1, M+2) to capture both spot price and short term expectations. It is not possible to stick to one contract for the whole backtest period, therefore, a roll over is needed before the contract expires. The simplest way to build continuous price curves is to roll a contract on its expiry day to the next expiring contract (rolling #1 in Tab. 3.3). The roll can be anticipated more and more until there is no connection with the expiry date, for example by rolling such that the price curve always uses the most liquid contract (rolling #4 in Tab. 3.3). The rolling strategy used for the backtests can be positioned in between these two extremes since contracts are not rolled on the expiry date, but yet there is still a link to it.

The goal of this section is to understand how the choice of a rolling strategy can influence the results of the backtests. This is done by simulating again the permutation tests for the three examples discussed in §4.3. The results are shown in Fig. 4.3. When the roll is done on the expiry day, then the number of statistically significant holding days decreases and in two out of three examples there is even no statistical significance left. Furthermore, $\text{Profit}_{\text{trade}}$ is always the lowest of the four rolling strategies. Rolling #2 and rolling #3 have similar numbers of trades, similar profit potentials, although the stability of predictive power across the number of holding days may vary. Rolling #4 always picks the most liquid contract, but the impact strongly depends on the futures market. Rolling #3 on Cocoa is very similar to rolling #4. The reason is that the front contract is the most liquid until 20-30 days before its expiry, therefore the two strategies roll very near in time. For Natural Gas ICE rolling #3 and rolling 4# show very different behaviours. This can be explained by the strong seasonality of Natural Gas contracts. The roll is discontinuous: some contracts can be used for long periods, while others can be completely left out of the continuous price curve. In a nutshell, rolling on the expiry date seems to penalize the predictive power and the profit potential of DeMark indicators. The data suggests to anticipate the roll before the expiry date of a contract to, ideally, 10 to 20 days (1 month) before the expiry. A possible continuation of this project may focus on finding optimal timings for the roll overs. Continuous prices curves constructed using the most liquid contract can be tricky because seasonality factors can make the roll very discontinuous and some contracts may be totally ignored.

4.4 Summary

The results of the backtest have been presented. Permutation tests are the main statistical tool to identify the predictive power of DeMark indicators on commodity futures markets. Sequential is the indicator that has shown predictive power the highest number of times (20 out of 21 markets). Long entry signals are mostly trend followers, like for energy products, while short entry signals are better for identifying trend reversals. Some patterns can be seen for commodity classes, like for energy, but also within the same class each market can have a very different behaviour, driven by different supply and demand dynamics. It is encouraging to see that similar products such as Brent Crude and Light Crude offer similar performances with Sequential. The same is valid for Heating Oil and Gas Oil, but also for Gold and Silver. Combo has many similarities with Sequential, but the main difference is that the number of trades is on average 30-40% lower. TDST has the highest number of trades and provides the correct direction of the trade, but it shows predictive power only on a lowest number of futures (6 out of 21 markets). For each indicator it is possible to suggest classes of commodities or specific futures for which entry signals are predictive (either long or short). A more general framework should combine predictive power with profit potential to identify the best combinations of indicators and markets. An indicator might be predictive, but if the expected gross profit is near to zero, then there is less interest in trading the signals. Based on these two parameters three market-indicator combinations were chosen to be further studied. The analysis starts from the comparison between conditional returns (simple and risk-adjusted) and unconditional return distributions and tires to find coherent results by looking at permutation tests. Lastly, for the same three examples, the effect of different rolling strategies on permutation tests is discussed. While the rolling strategy used in the test seems reasonable, there is still room to develop an optimal rolling strategy.

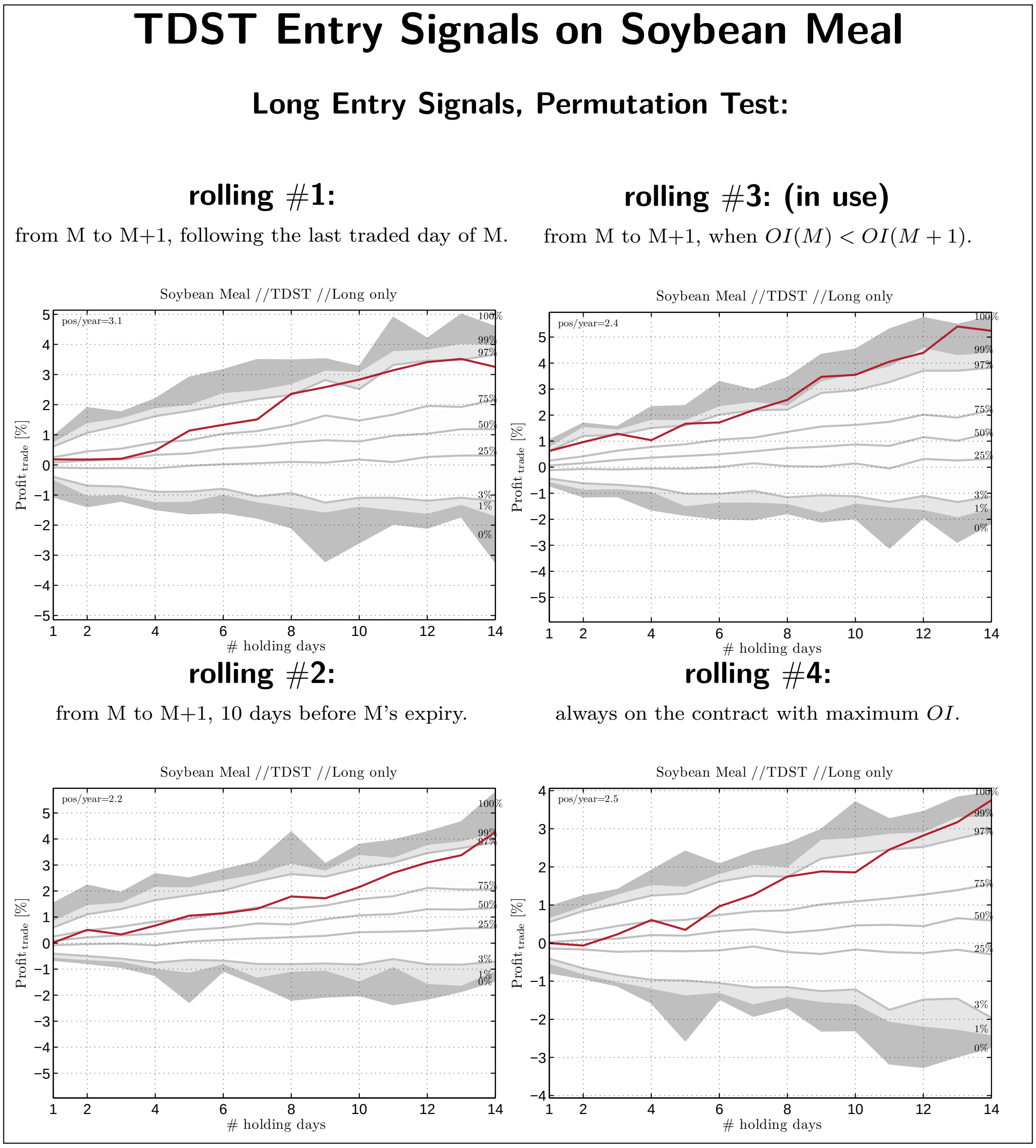


Figure 4.3: *Impact of different rolling strategies on the permutation test for TDST on Soybean Meal. When* Profit$_{\text{trade}}$ *reaches the statistically significant area in grey, then the tested indicator has predictive power on the underlying futures market. The simplest way to build continuous price curves is to roll a contract on its expiry day to the next expiring contract (rolling #1 in Tab. 3.3). The roll can be anticipated more and more until there is no connection with the expiry date, for example by rolling such that the price curve always uses the most liquid contract (rolling #4 in Tab. 3.3). The rolling strategy used for the backtests is #3 and can be positioned in between these two extremes since contracts are not rolled on the expiry date, but yet there is still a link to it.*

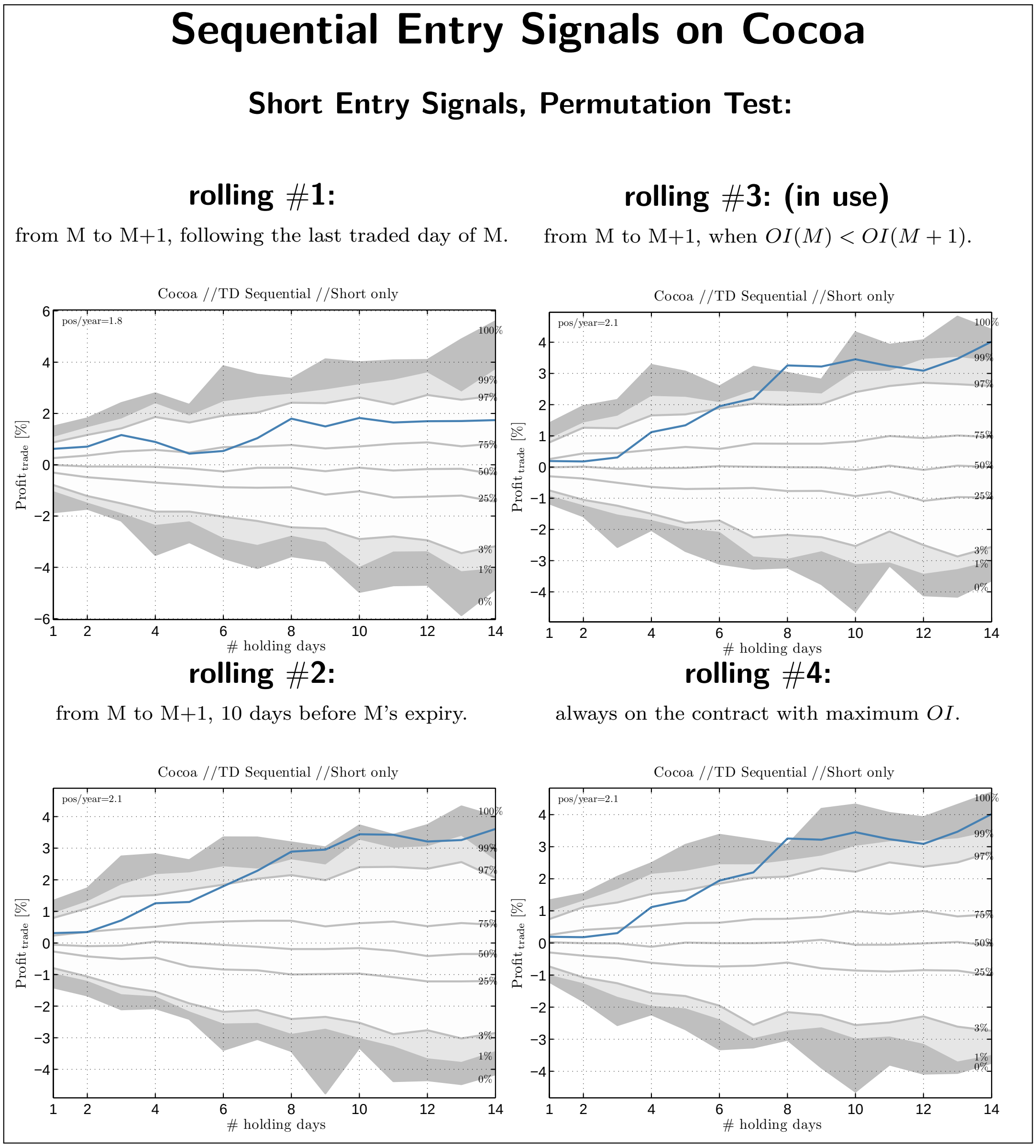


Figure 4.4: *Impact of different rolling strategies on the permutation test for Sequential on Cocoa. The aggregated quality measure for the test is* $\text{Profit}_{\text{trade}}$. *Rolling #3 was used for the backtest.*

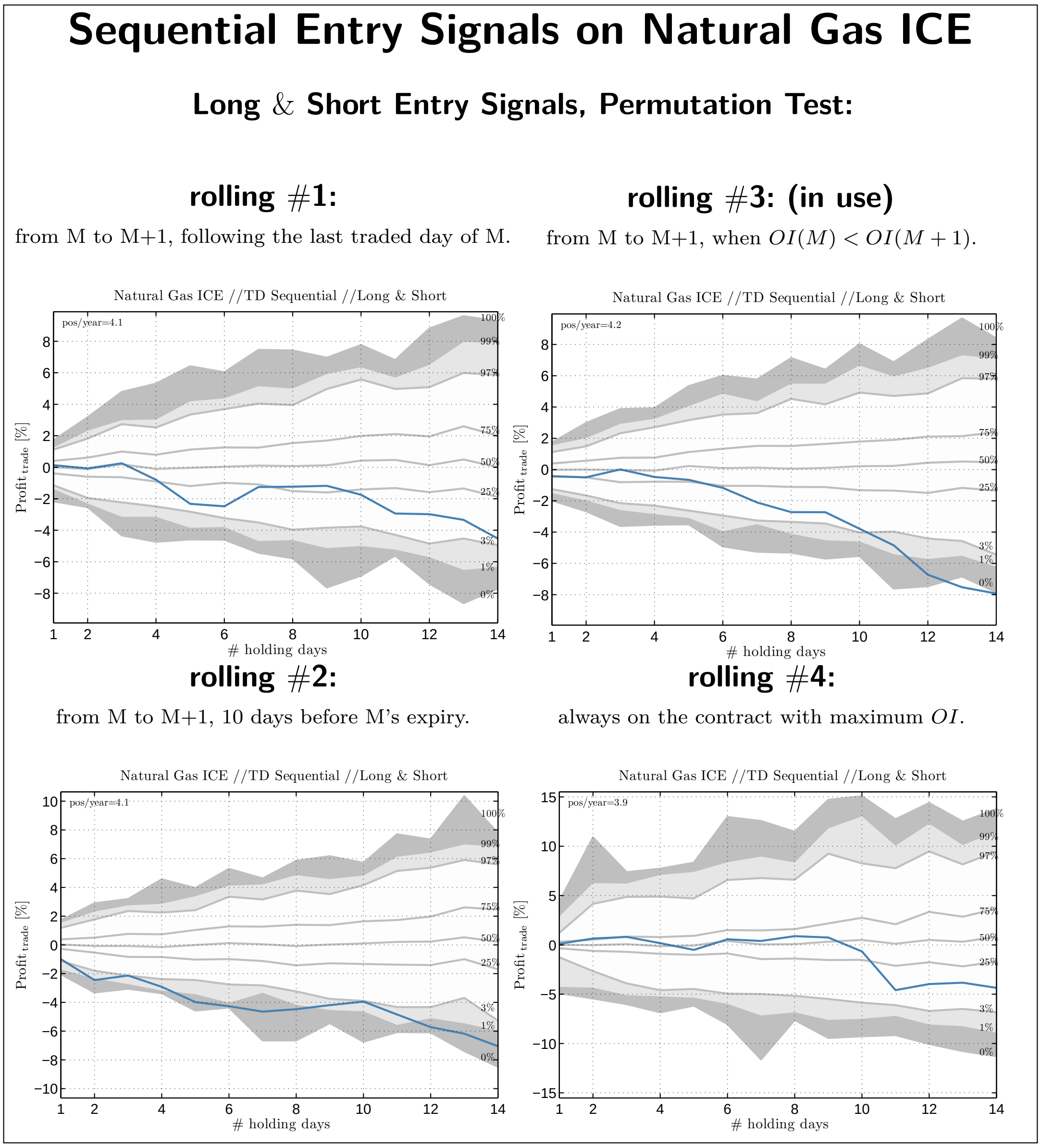


Figure 4.5: *Impact of different rolling strategies on the permutation test for Sequential on Natural Gas ICE. The aggregated quality measure for the test is* Profit$_{\text{trade}}$. *Rolling #3 was used for the backtest.*

Conclusion

Currently, there are relatively few studies of technical analysis on commodity futures markets. In this work, the predictive power of three DeMark technical indicators (Sequential, Combo and TDST) has been tested on 21 commodity futures belonging to the following classes: grains, softs, energy, industrial metals and precious metals. An original aspect is that, compared to most academic studies which test indicators that are not frequently used in practice, the indicators tested here are currently among the most popular ones and it is possible to use them in leading financial platforms. DeMark is mainly renowned for his Sequential. It is a time-based indicator that identifies potential turning points from trends (Setup phase) and then forecasts the beginning of price reversals (Countdown phase). Combo is the main variant to Sequential: while it uses different rules to forecast the timing of price reversals, turning point identification stays the same via the Setup phase. TDST is complementary to both because it uses the Setup as well, but it tries to capture sustainable trends instead of searching for price reversal patterns. These are all indicators that provide entry signals, they are not trading systems.

The tested period starts on 1/1/2004 and ends on 1/1/2014. The end of the time period is close to the present, but there should be still a time period between the last trading day and the current date. This period determines the out-of-sample data which is precious because the results of the backtest can be retested before making trading decisions on them. A cross-validation on this new data helps to identify recent changes in market behaviour and also to identify data mining biases which overestimate the profitability of an indicator in processes where large data is selected/optimized. For each trading day, DeMark indicators need bar charts (starting, highest, lowest and closing prices). Just by looking at the average trade frequency (3.9 trades per year for Sequential, 1.7 for Combo and 4.5 for TDST), we can conclude that the entry signals are sparse. This means that for most of the tested period the trade signal would have been out of the market. In practice, this limits the possibilities regarding performance measurement because such sparse positions block the use of measures which show profit evolution over the trading period (e.g., Net Asset Value).

Before running the backtest, daily returns are computed from discontinuous commodity futures markets. This is done by rolling the individual contracts into one continuous futures price series. This approach is the most common for technical traders who trade in commodity futures markets. It is the most straightforward way for investors to get exposure to market prices. Long or short positions are entered just on one contract, for example always on the rolled front contract. This trading style makes profits when positions are properly matched with daily price movements of a contract at it is usually referred to as *outright exposure*. Traditional commodity trading instead, makes a limited use of this exposure. In fact, typical commodity traders speculate also without outright (e.g. on multiple contracts with time-spread and inter-product exposures), which means that their daily profits do not depend, to a certain extent depending on the correlation between contracts, on price movements of the continuous futures prices. The rolling should complete three sequential steps: the timing strategy, the adjustment strategy and the transformation from prices to returns.

The predictive power of each indicator has been studied in two steps. The first is to compare conditional returns on entry signals to exact unconditional return distributions (which represent the market). This can be visualized (Fig. 2.4 and 4.2) or tested by the Kolmogorov-Smirnov test which is derived under the assumption that returns are Independent and Identically Distributed (IID), but this is not plausible for financial data [29]. An overperformance of the conditional distribution compared the market suggests that the tested indicator might have predictive power. The second step uses approximated permutation tests to check if the initial suggestion is correct.

All three indicators exhibit predictive power on some commodity futures. Most of the times the entry signals provided by the indicators shows predictive power only for long or only for short positions. If we consider that long and short entry signals are generated by symmetrical algorithms, then this confirms the fact that uptrends and downtrends are asymmetrical in the markets. In reality instead, Sequential is DeMark's most famous indicator. For all the commodity futures apart from Platinum the indicator has shown statistically significant predictive power for either long or short entry positions. It is informative on energy for long positions and on the other commodity classes for short positions. Light Crude, Heating Oil and Gas Oil are an exception because both long and short positions are informative. Although Sequential is described as a time-based indicator that identifies turning points, there are products or even commodity classes, like energy products, where entry signals (long and/or short) identify continuing trends instead of turning points. Combo has many similarities with Sequential, but the main difference is that the number of trades is on average 30-40% lower. Compared to Sequential there are more cases in which the signal is predictive for both long and short signals, for example with Gold and Silver. TDST has the highest number of trades (4.5 trades/year on average) and provides the correct trade direction, but it shows predictive power on the lowest number of futures (6 out of 21 markets) and mainly for long positions. Long TDST signals are informative in particular for grains while for combinations of long and short positions there is statistical significance for Wheat, Soybean Meal, Natural Gas NYMEX, Brent, Gold and Copper NYMEX.

The choice of which indicator to use to generate trades on a given market should be mainly driven by the stability of predictive power over the holding days. Nevertheless, it is also important to include a measure of expected profit potential (for example $\text{Profit}_{\text{trade}}$) because it is not sufficient to overperform the market if still no profits can be made. This seems a more complete approach compared to only choosing markets where indicators are overperforming, but it also carries further complications. If the choice to trade a signal on a specific market is based mainly on the maximization of the observed profit potential, then it is likely that the measured performance will be lower compared to expectations. For this reason, and also because market conditions might change during the tested period, results should be cross-validated on the most recent out-of-sample-data.

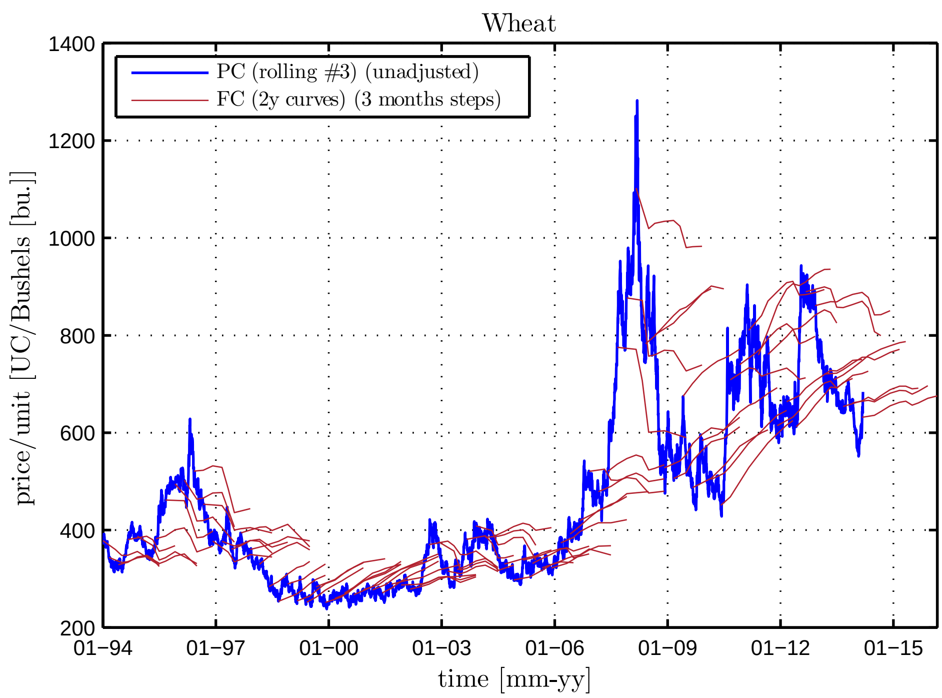
Another original element of this work is the study on the sensitivity to different rolling strategies for predictive power and profit potential. A continuous price series of a futures contract must have one contract selected for each trading day. In general it uses prices from contracts which are close to expiry (i.e. M, M+1, M+2) to capture both spot price and short term expectations, but there are potentially infinite ways to roll over to the next contract. Four different strategies have been tested. The main conclusion is that when the roll is done on the expiry day, then the stability of predictive power across the holding days decreases and in two out of three examples there is even no statistical significance left. Furthermore, for this strategy $\text{Profit}_{\text{trade}}$ is always the lowest compared to the other rolling strategies.

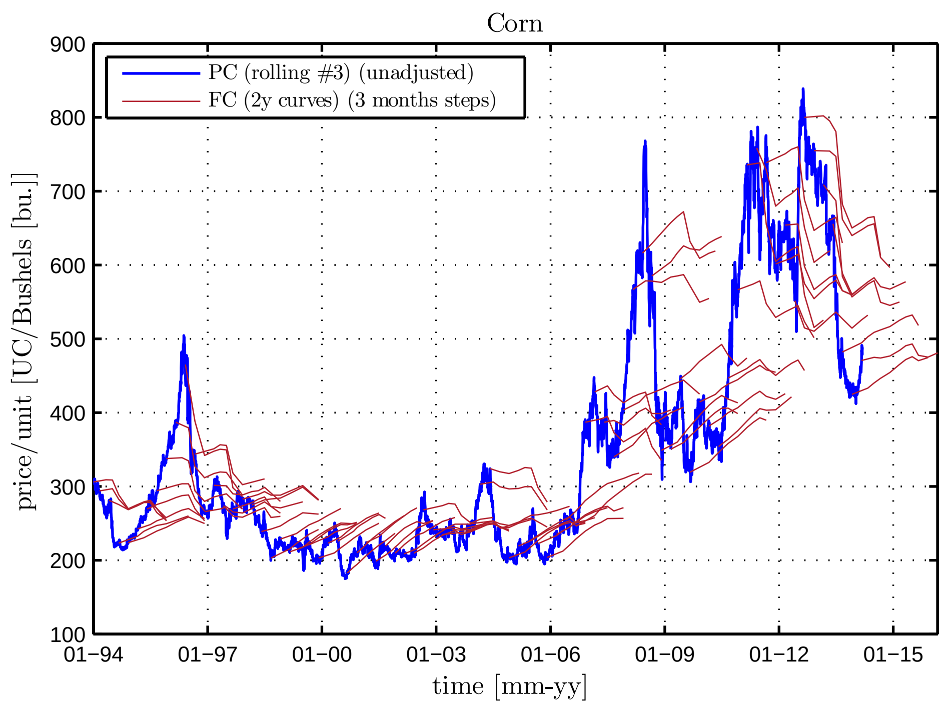
Further developments of this work should include backtests of the TDST indicator using complete entry and exit signals, a more general study on the effects of rolling strategies on backtests and, specifically for backtests on commodity futures, time-spreads and interprduct-spreads should be added as different sources of market exposures. Parameter optimization was not included in this work because the first step when testing indicators is to analyse their natural trading potential (the stability of predictive power combined with profit potential) on specific markets. Only then, if it is worth the effort, parameters can be optimized to maximize trading potential.

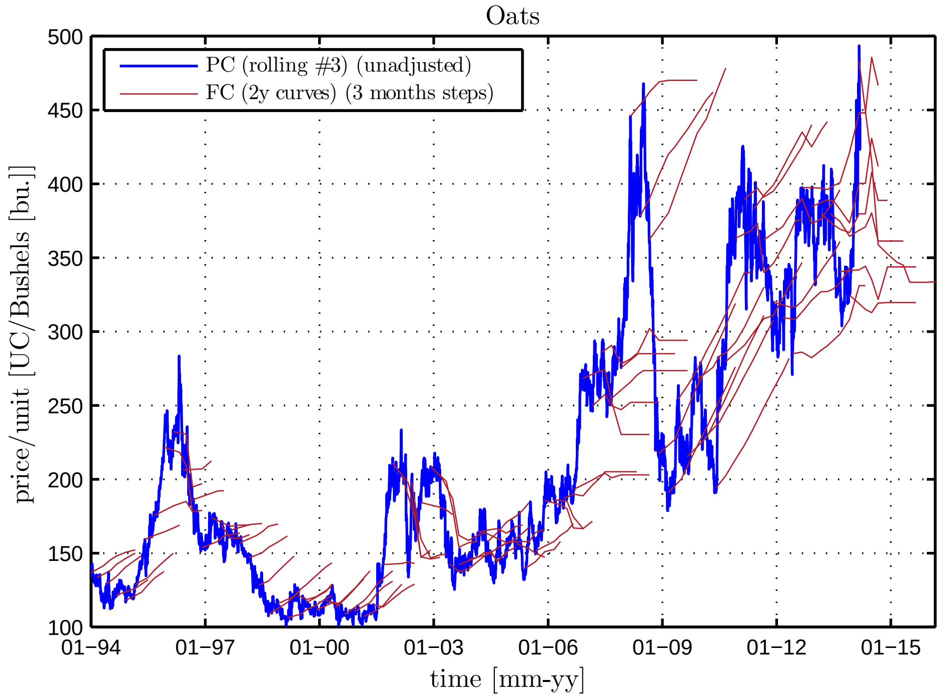
APPENDIX $\mathcal{A}$

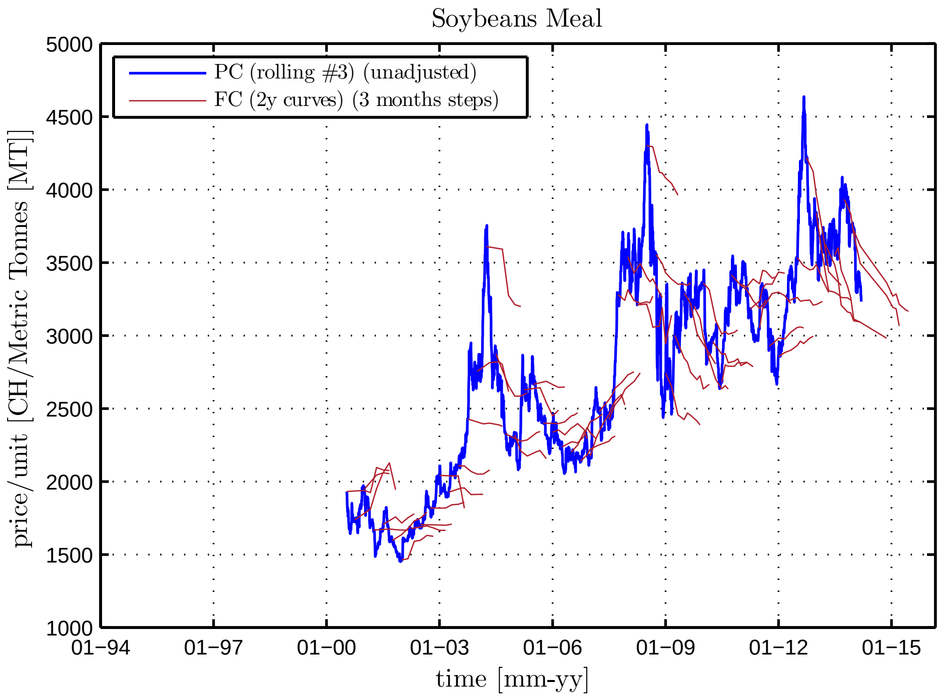
Continuous Price Curves and Forward Curves

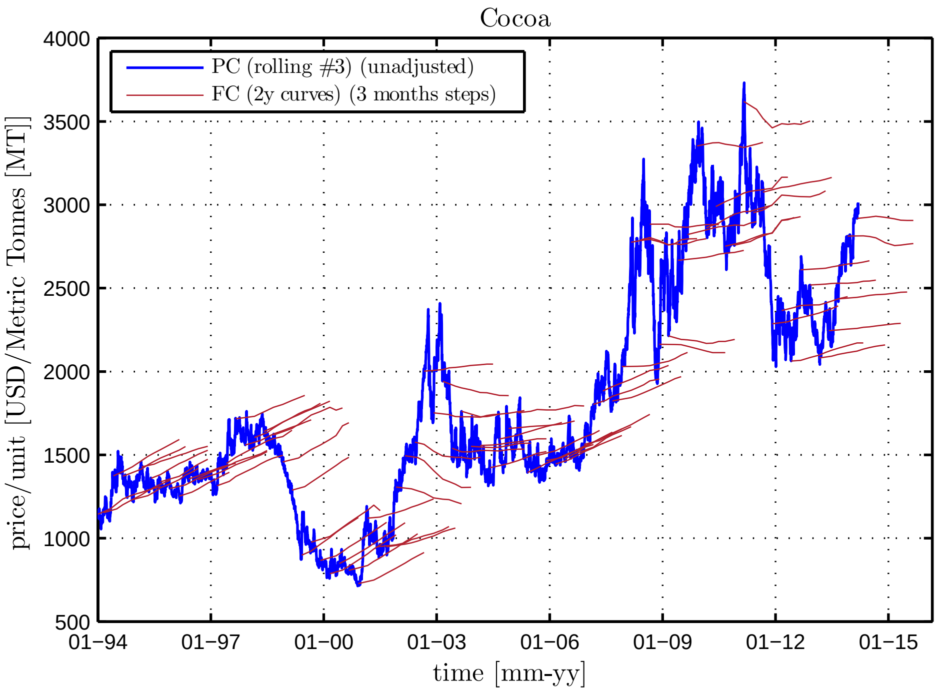
The whole section is dedicated to the commodity futures used in the backtests. For each of them, there is a figure which includes the continuous price curve and the forward curves. The blue curves are the continuous closing prices (PC) constructed by rolling the front contract M to the next M+1 in the last month before expiry, as soon as the open interest of M+1 is higher than that of M. The red lines are forward curves. For each date there is a forward curve (FC) linked to it. One curve every three months was plotted. Furthermore, each curve was built using only the contracts that had expiry dates within 2 years from the date to which each forward curve is related.

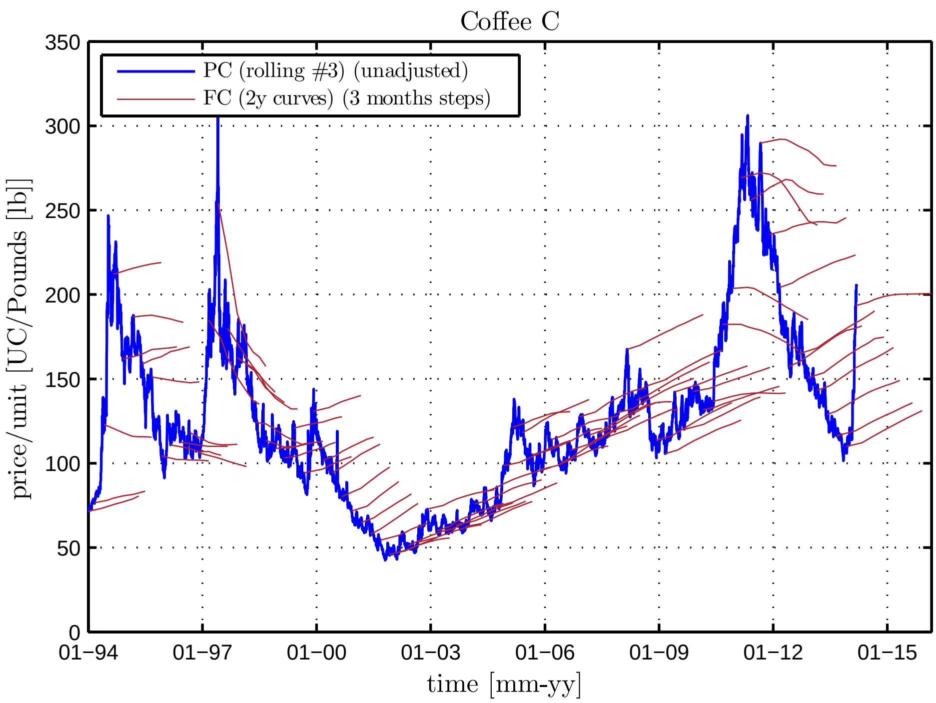

Wheat
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Bushels [bu.]]
time [mm-yy]
1400
1200
1000
800
600
400
200
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

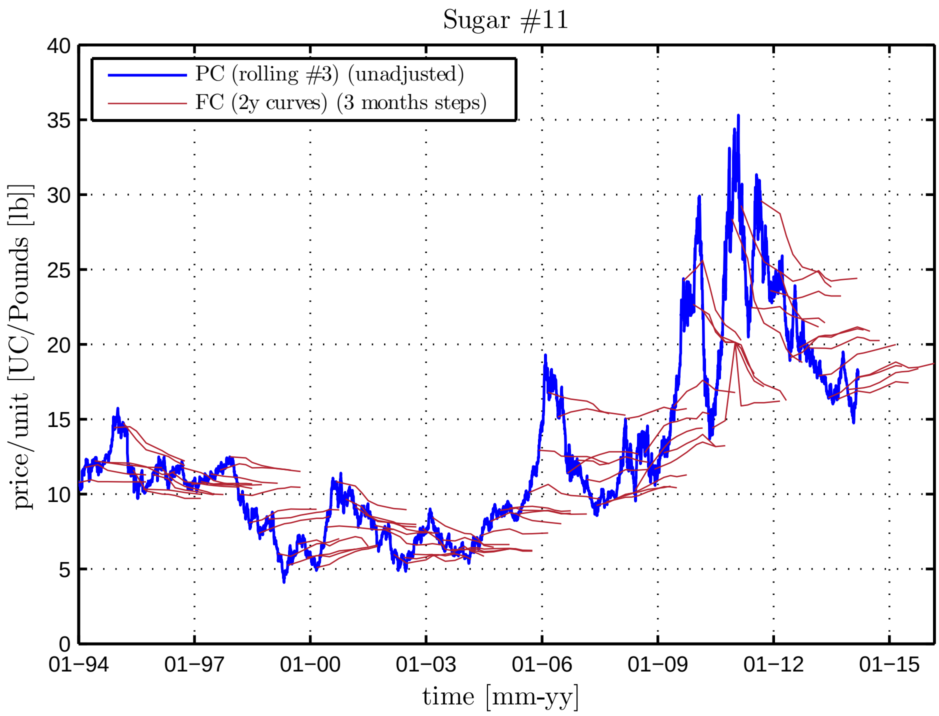

Corn
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Bushels [bu.]]
time [mm-yy]
900
800
700
600
500
400
300
200
100
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

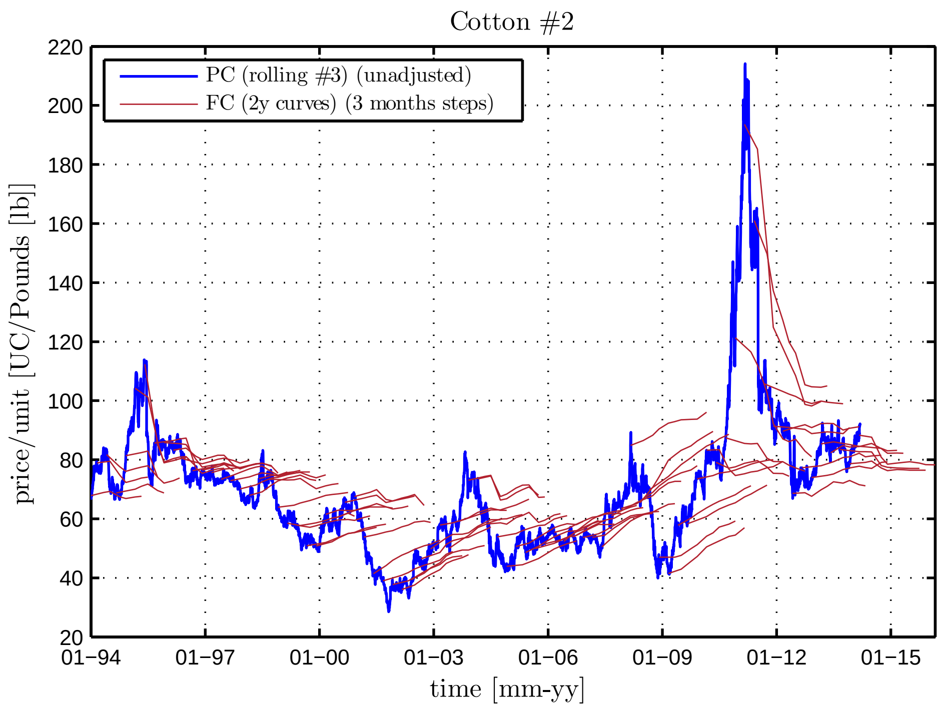

Oats
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Bushels [bu.]]
time [mm-yy]
100
150
200
250
300
350
400
450
500
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

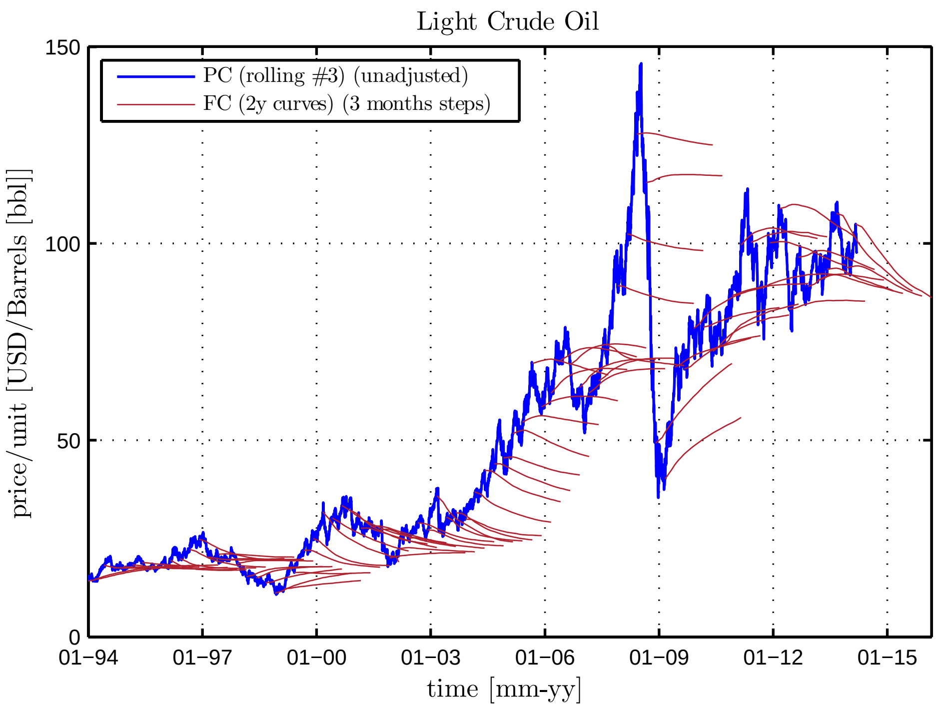

Soybeans Meal
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [CH/Metric Tonnes [MT]]
time [mm-yy]
1000
1500
2000
2500
3000
3500
4000
4500
5000
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

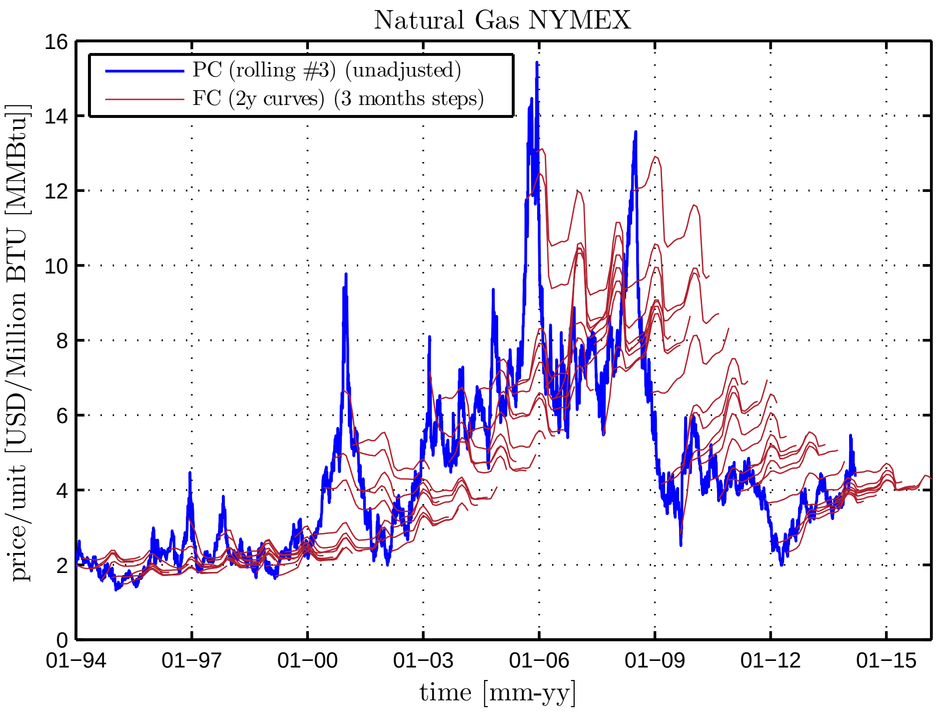

Cocoa
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Metric Tonnes [MT]]
4000
3500
3000
2500
2000
1500
1000
500
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]

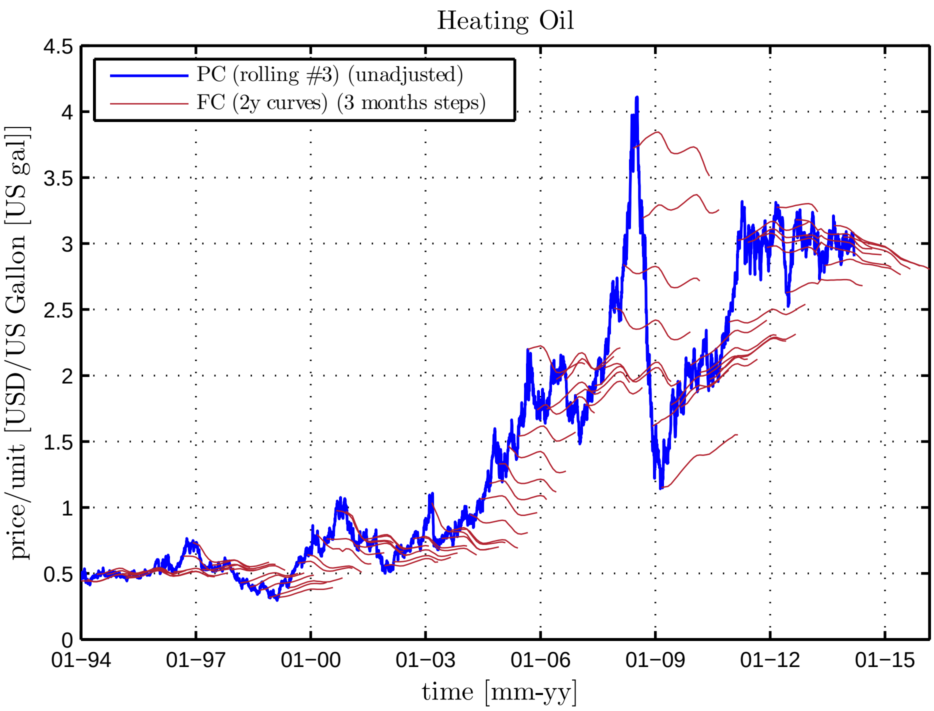

Coffee C
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Pounds [lb]]
350
300
250
200
150
100
50
0
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]

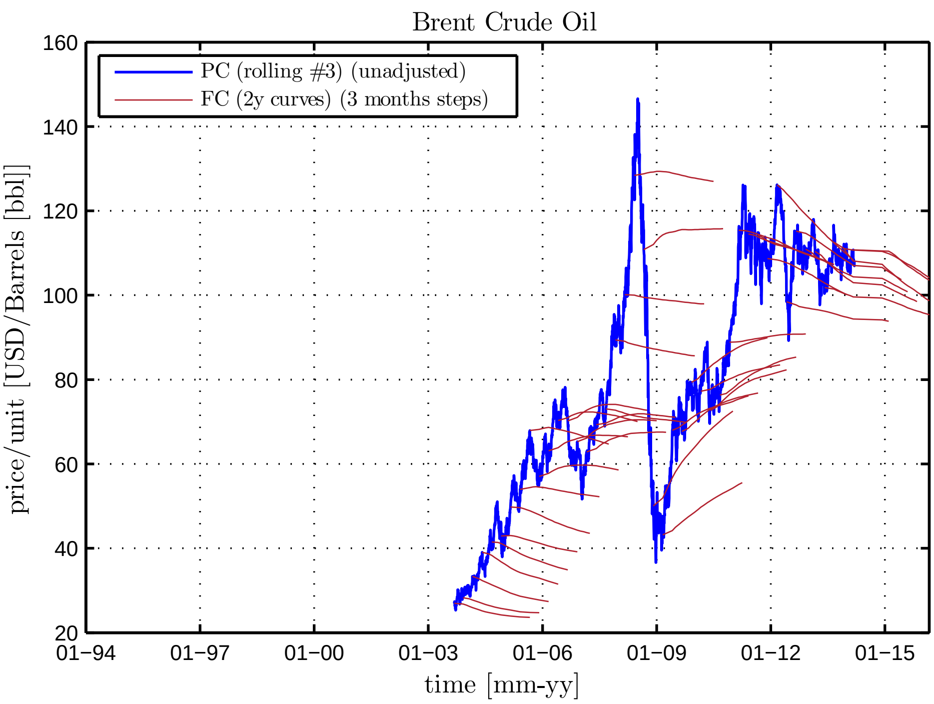

Sugar #11
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Pounds [lb]]
time [mm-yy]
0
5
10
15
20
25
30
35
40
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

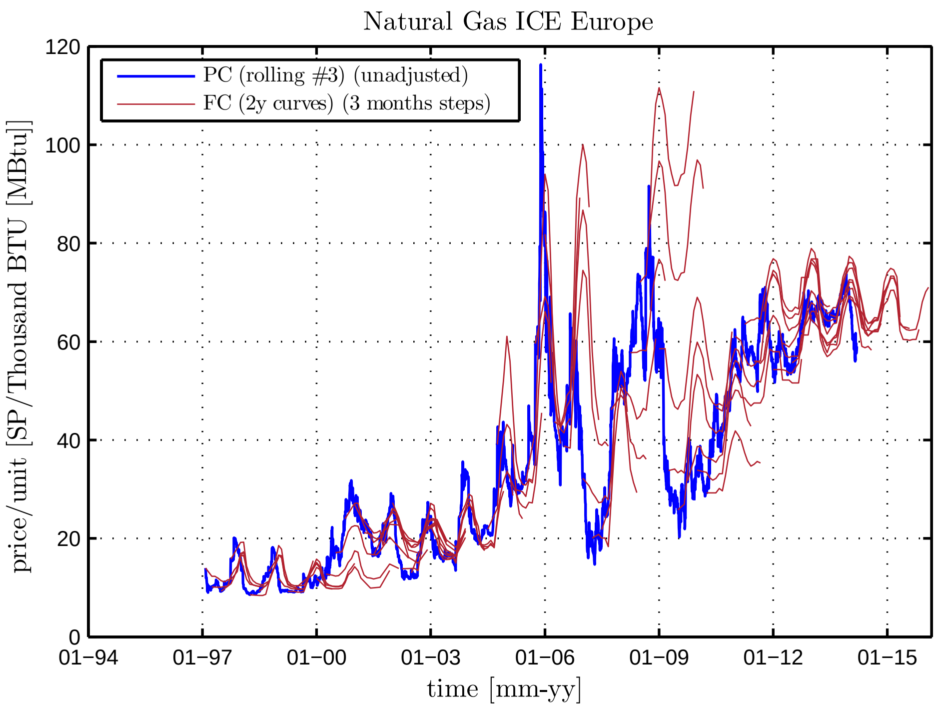

Cotton #2
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [UC/Pounds [lb]]
time [mm-yy]
20
40
60
80
100
120
140
160
180
200
220
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

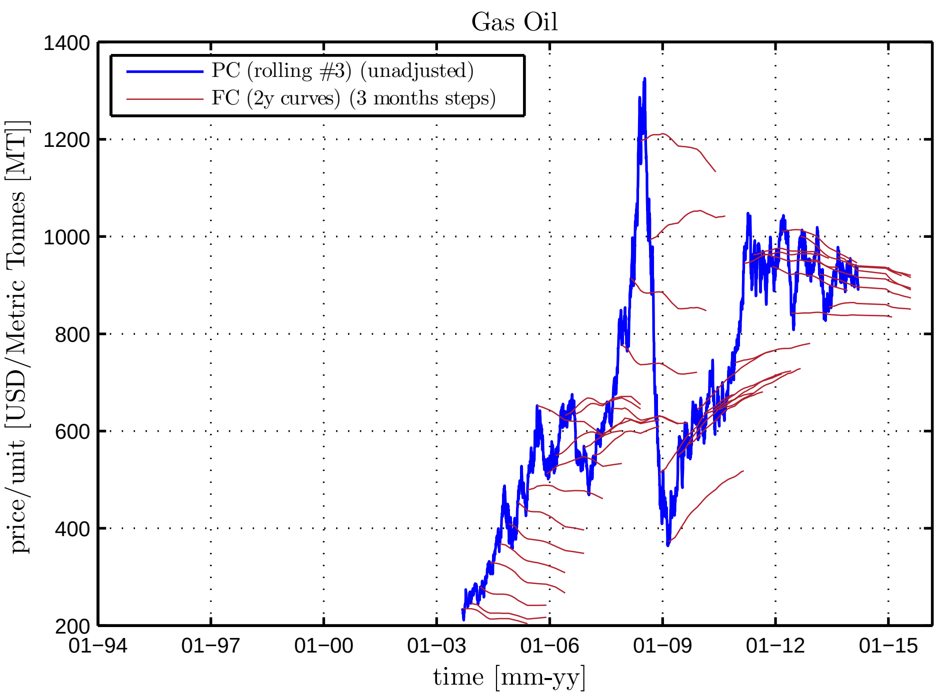

Light Crude Oil
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Barrels [bbl]]
time [mm-yy]
0
50
100
150
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

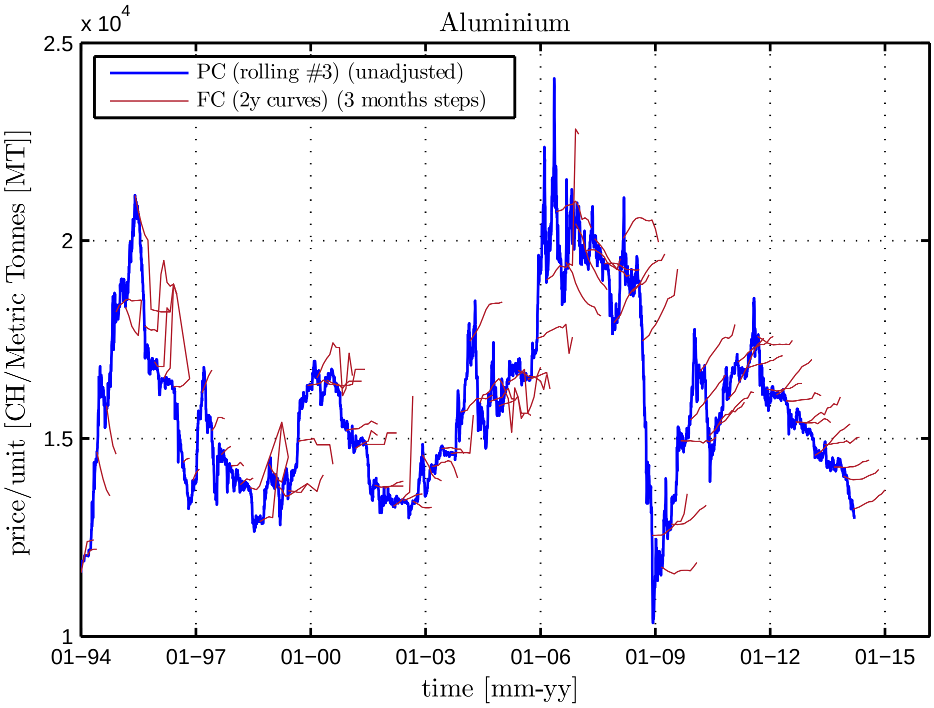

Natural Gas NYMEX
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Million BTU [MMBtu]]
time [mm-yy]
0
2
4
6
8
10
12
14
16
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15


Heating Oil
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/US Gallon [US gal]]
time [mm-yy]
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

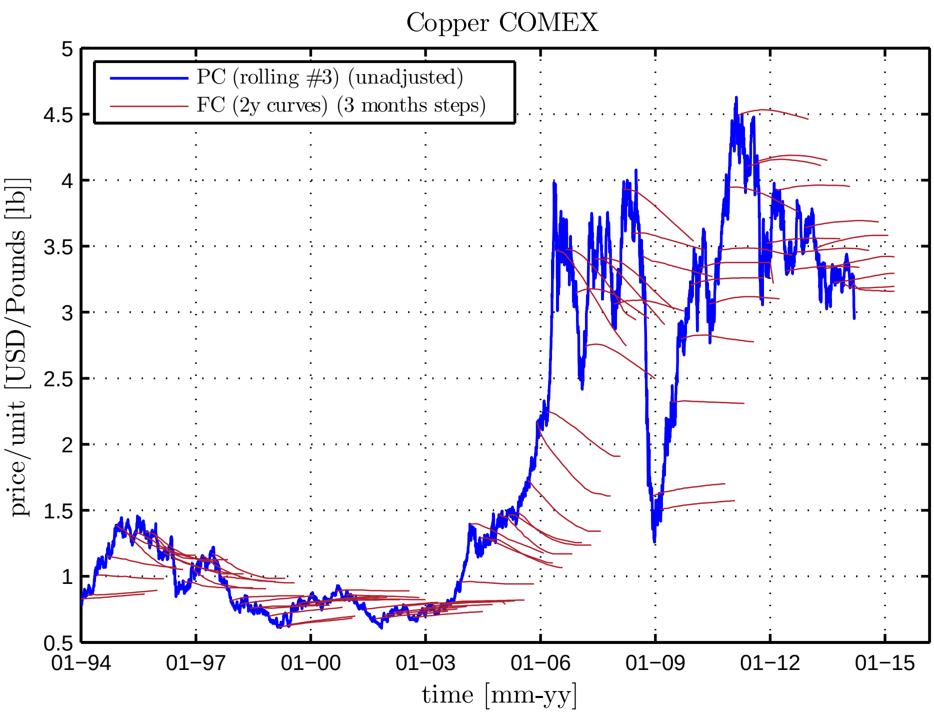

Brent Crude Oil
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Barrels [bbl]]
time [mm-yy]
20
40
60
80
100
120
140
160
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15


Natural Gas ICE Europe
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [SP/Thousand BTU [MBtu]]
time [mm-yy]
0
20
40
60
80
100
120
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

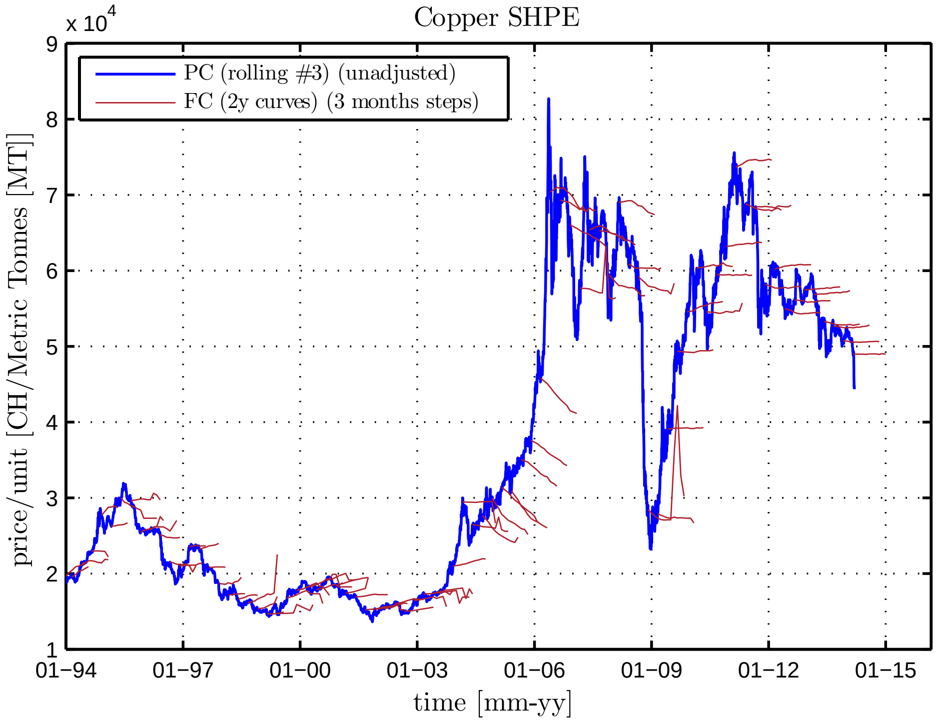

Gas Oil
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Metric Tonnes [MT]]
time [mm-yy]
200
400
600
800
1000
1200
1400
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15

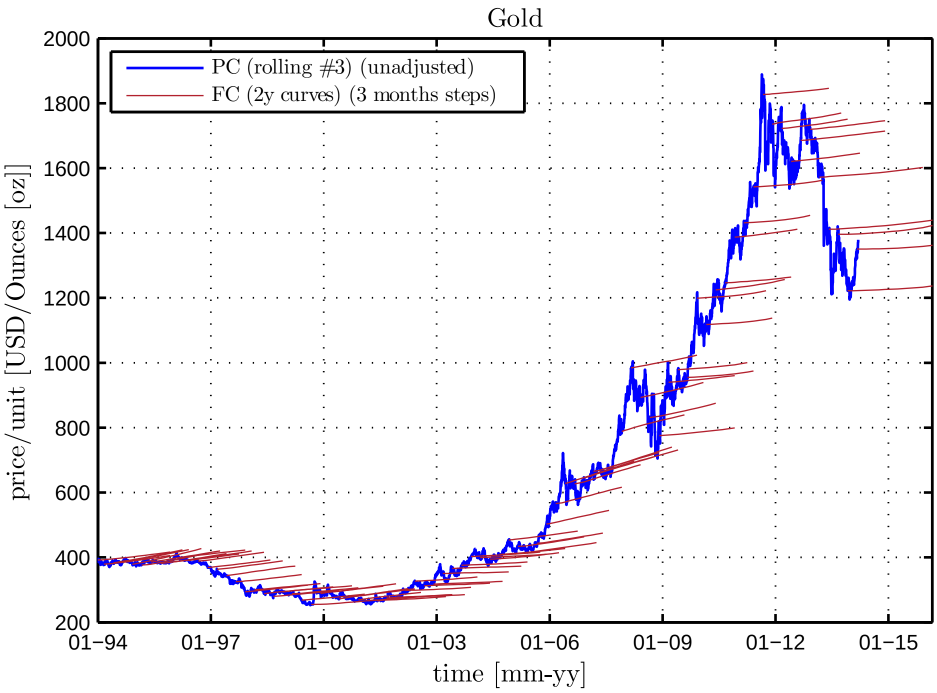


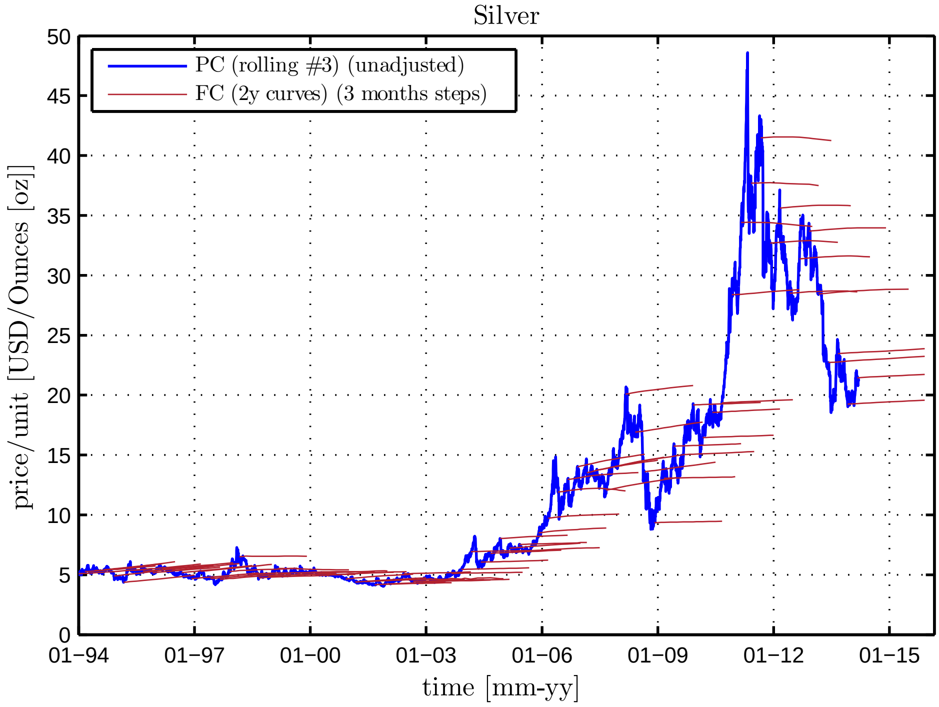
Aluminium
x 10^4
2.5
2
1.5
1
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]
price/unit [CH/Metric Tonnes [MT]]
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)

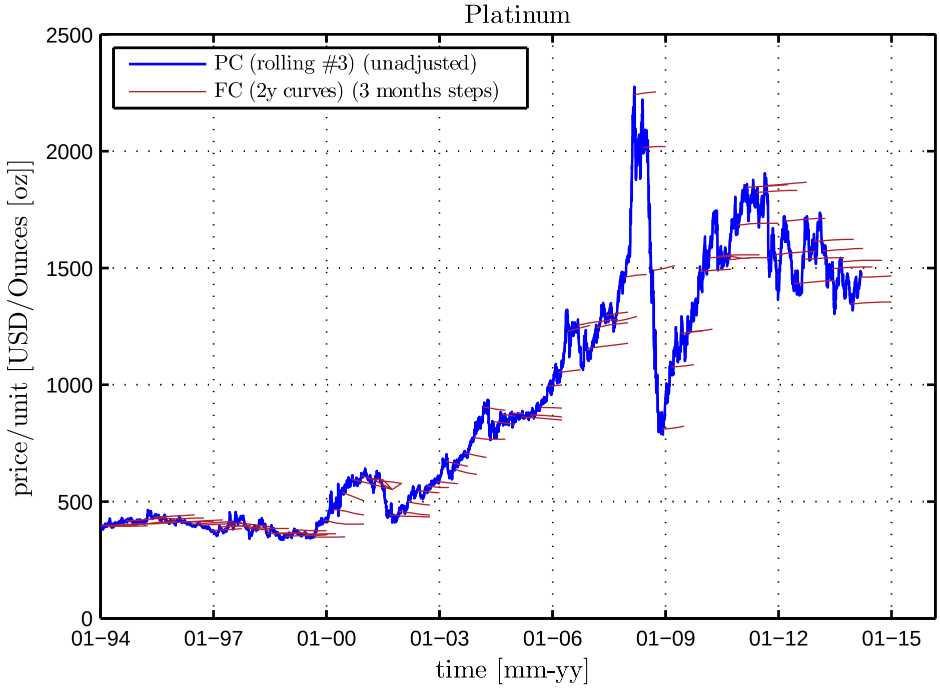


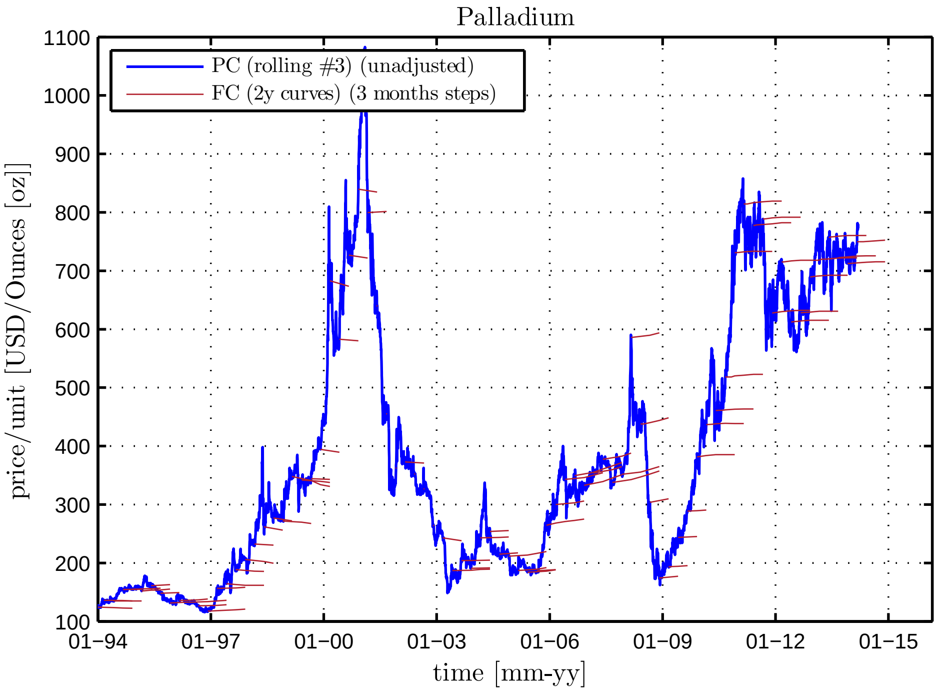
Copper COMEX
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]
price/unit [USD/Pounds [lb]]
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)


Copper SHPE
x 10^4
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [CH/Metric Tonnes [MT]]
1
2
3
4
5
6
7
8
9
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]


Gold
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Ounces [oz]]
200
400
600
800
1000
1200
1400
1600
1800
2000
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]


Silver
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Ounces [oz]]
time [mm-yy]
50
45
40
35
30
25
20
15
10
5
0
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15


Platinum
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Ounces [oz]]
time [mm-yy]
2500
2000
1500
1000
500
0
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15



Palladium
PC (rolling #3) (unadjusted)
FC (2y curves) (3 months steps)
price/unit [USD/Ounces [oz]]
1100
1000
900
800
700
600
500
400
300
200
100
01−94
01−97
01−00
01−03
01−06
01−09
01−12
01−15
time [mm-yy]

Bibliography

[1] Agresti, A. e B. A. Coull: *Approximate is better than "exact" for interval estimation of binomial proportions.* The American Statistician, 52:119–126, 1998.

[2] Andersen, J. V., S. Gluzman e D. Sornette: *Fundamental framework for techincal analysis.* European Physical Journal, B 14:579–601, 2000.

[3] Aronson, D. R.: *Evidence-Based Techical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals.* Wiley, 2007.

[4] Association, S. B.: *Switzerland as a commodity trading centre.* Swiss Bankers Association, 2013. `http://shop.sba.ch/999930_e.pdf`.

[5] Atilgan, Y., T. G. Bali e O. Demirtas: *The performance of hedge fund indices.* Borsa Istanbul Review, 13:30–52, 2013.

[6] Bacon, C.: *Practical Portfolio Performance Measurement & Attribution.* Wiley, 2004.

[7] Beja, A. e B. Goldman: *On the dynamic behaviour of prices in disequilibrium.* The Journal of Finance, 35-2:235–248, 1980.

[8] Benyamini, Z.: *Going long vs. going short.* Technical Analysis of Stocks & Commodities, 27:7:10–14, 2009.

[9] Brorsen, B. W. e S. H. Irwin: *Futures funds and price volatility.* The Review of Futures Markets, 6-2:119–135, 1987.

[10] Carchano, O. e A. Pardo: *Rolling over stock index futures contracts.* The Journal of Futures Markets, 29:7:684–694, 2009.

[11] Chan, L. e L. Lin: *Boosting profitability.* Technical Analysis of Stocks & Commodities, 22:8:28–32, 2004.

[12] Chekhlov, A., S. Uryasev e M. Zabarankin: *Drawdown measure in portfolio optimization.* International Journal of Theoretical and Applied Finance, 8-1:13–58, 2005.

[13] Coles, A.: *Automated techniques for intrady traders.* Technical Analysis of Stocks & Commodities, 29:8:14–31, 2011.

[14] Coles, A.: *Td sequential and ermanometry for intraday traders.* Technical Analysis of Stocks & Commodities, 29:9:12–22, 2011.

[15] Conover, W. J.: *Practical Nonparametric Statistics.* Wiley, 1999.

[16] DeMark, T.: *New Market Timing Techniques: Innovative Studies in Market Rythm & Price Exhaustion.* Wiley, 1997.

[17] Eling, M. e F. Schuhmacher: *Does the choice of performance measure influence the evaluation of hedge funds?* Journal of Banking & Finance, 31:2632–2647, 2007.

[18] Eling, M. e F. Schuhmacher: *Sufficient conditions for expected utility to imply drawdown-based performance rankings.* Journal of Banking & Finance, 35:2311–2318, 2011.

[19] Eling, M. e F. Schuhmacher: *A decision-theoretic foundation for reward to-risk performance measures.* Journal of Banking & Finance, 36:2077–2082, 2012.

[20] Fama, E. F. e M. E. Blaume: *Filter rules and stock-market trading.* Journal of Business, 39-1:226–241, 1966.

[21] Hamori, S., N. Hamori e D. A. Anderson: *An empirical analysis of the efficiency of the osaka rice market during japan's tokugawa era.* The Journal of Futures Markets, 21:861–874, 2001.

[22] Hannam, M.: *The commodity risk premium.* Nel *Essays*, 2009. `http://www.markhannam.com/essays/essay2a.htm`.

[23] Harris, M.: *Profitability and Systematic Trading.* Wiley, 2009.

[24] Hodges, S.: *A generalization of the Sharpe ratio and its applications to valuation bounds and risk measures.* Coventry : Financial Options Research Centre, Warwick Business School, Univ. of Warwick, 1998.

[25] Hollander, M. e D. A. Wolfe: *Nonparametric Statistical Methods.* Wiley, 1999.

[26] Johnsson, R.: *A simple risk-return-ratio*, lug. 2010. `http://www.richardcbjohnsson.net/pdf/risk_return_ratio.pdf`.

[27] Kosar, J.: *Support and resistance levels.* Technical Analysis of Stocks & Commodities, 9:1:17–19, 1991.

[28] Lo, A. W.: *The statistics of sharpe ratios.* Financial Analysts Journal, 58:4:36–52, 2002.

[29] Lo, A. W., M. Harry e J. Wang: *Fundations of technical analysis: Computational algorithms, statistical inference and empirical implementation.* The Journal of Finance, 55:4:1705–1765, 2000.

[30] Lo, A. W. e J. Hasanhodzic: *The Evolution of Technical Analysis.* Wiley, 2010.

[31] Lukac, L. P. e B. W. Brorsen: *A comprehensive test of futures market disequilibrium.* The Financial Review, 25-2:593–622, 1990 1990.

[32] Lukac, L. P., B. W. Brorsen e S. H. Irwin: *A test of futures market disequilibrium using twelve different technical trading systems.* Applied Economics, 20:623–639, 1988.

[33] Ma, C. K., J. M. Mercer e M. A. Walker: *Rolling over futures contracts: A note.* The Journal of Futures Markets, 12:2:203–217, 1992.

[34] Magnus, K.: *Impressive signals from demark.* The Technical Analyst, 1:6:20–24, 2004.

[35] Masteika, S. e A. V. Rutkauskas: *Research on futures trend trading strategy based on short term chart pattern.* Journal of Business Economics and Management, 13(5):915–930, 2012. `http://www.tandfonline.com/doi/abs/10.3846/16111699.2012.705252`.

[36] Masteika, S., A. V. Rutkauskas e J. A. Alexander: *Continuous futures data series for back testing and technical analysis.* Nel *2012 International Conference on Economics, Business and Marketing Management*, 2012.

[37] Masters, T.: *Monte-carlo evaluation of trading systems*, 2010. `http://www.evidencebasedta.com/montedoc12.15.06.pdf`.

[38] Miner, R.: *Time as a trading tool.* Technical Analysis of Stocks & Commodities, 9:3:113–115, 1991.

[39] Murphy, J. J.: *Techical Analysis of the Financial Markets.* The New York Institute of Finance, 1986.

[40] NASDAQ: *NASDAQ Commodity Index Methodology*, July 2013.

[41] Oberlechner, T.: *Importance of technical and fundamental analysis in the european foreign exchange market.* International Journal of Finance and Economics, 6:81–93, 2001.

[42] Osler, C. L.: *Identifying noise traders: The head-and-shoulders pattern in u.s. equities.* Nel *Federal Reserve Bank of New York*, 1998.

[43] Park, C. H. e S. H. Irwin: *The profitability of technical trading rules in us futures markets: A data snooping free test.* 2005.

[44] Pelletier, B.: *Computed contracts: Their meaning, purpose and application.* Nel *An Essay on Computed Contracts*, 2011. http://www.csidata.com/cgi-bin/getManualPage.pl?URL=essay.htm.

[45] Perl, J.: *DeMark Indicators.* Wiley, 2008.

[46] Reuters: *Futures Continuous Series, Methodology and Definitions*, 0.1 version ed., August 2010.

[47] Roberts, M. C.: *Technical analysis in commodity markets: Risk, returns, and value.* Nel *Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, 2003.

[48] Samuelson, P.: *Proof that properly anticipated prices fluctuate randomly.* Industrial Management Review, 6:41–51, 1965.

[49] Schuster, M. e B. R. Auer: *A note on empirical sharpe ratio dynamics.* Economics Letters, 116:124–128, 2012.

[50] Siligardos, G. E.: *Evaluating algorithmic trading signals.* Technical Analysis of Stocks & Commodities, 32:6:10–15, 2014.

[51] Skouras, S. http://www.aueb.gr/users/skouras/tabiblio.htm, 1998.

[52] Smidt, S.: *Amateur speculators.* tesi di laurea, Graduate School of Business and Public Administration, Cornell University, 1964.

[53] Sullivan, R., A. Timmerman e H. White: *Data snooping, technical trading rule performance, and the bootstrap.* Journal of Finance, 54:1647–1691, 1999.

[54] Sullivan, R., A. Timmerman e H. White: *Forecast evaluation with shared data sets.* International Journal of Forecasting, 19:217–227, 2003.

[55] Swiss Federal Council: *Report on commodities*, mar. 2013. https://www.news.admin.ch/message/index.html?lang=en&print_style=yes&msg-id=48319.

[56] Szymanowska, M., F. de Roon, T. Nijman e R. van den Goorbergh: *An anatomy of commodity futures risk premia.* The Journal of Finance, 69:1:453–482, 2014.

[57] White, H.: *A reality check for data snooping.* Econometrica, 68-5:1097–1126, 2000.

[58] Widner, M.: *Automated support and resistance.* Technical Analysis of Stocks & Commodities, 16:5:225–232, 1998.

[59] Zakamouline, V. e S. Koekebakker: *Portfolio performance evaluation with generalized sharpe ratios: Beyond the mean and variance.* Journal of Banking & Finance, 33:1242–1254, 2009.