Statistical Testing of DeMark Technical Indicators on Commodity Futures

Marco Lissandrin[1,*], Donnacha Daly, Didier Sornette

Chair of Entrepreneurial Risks, ETH Zürich, Scheuchzerstrasse 7, SEC F 7, CH - 8092 Zürich, Switzerland

Abstract

We examine the performance of three DeMark indicators (Sequential, Combo and Setup Trend), which constitute specific implementations of Technical Analysis often used by practitioners, over 21 commodity futures markets and 10 years of data. Our backtests characterise the predictive power of these indicators. 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 market-exit signals, randomization tests have been performed for benchmarking. We generate the distributions of three performance metrics (mean return, profit factor and risk-return ratio) over different trade holding horizons and compare them with their randomized versions. We have further checked the impact of the rolling strategy of future contracts on the performance of the DeMark indicators. For the period from Jan. 2004 to Jan. 2014, our results suggest statistically significant predictive power on a wide range of commodity futures.

Keywords: Technical Analysis, Back-Testing, Permutation Test, Financial Markets, Commodity Futures, Contract Roll-Over
JEL: C12, G14, G17

1. Introduction

In financial markets, Technical Analysis (TA) refers to a set of methods that examine past and present market activity such as price, volume and open interest to identify patterns that can predict future price movements. Recent books and literature surveys (Irwin, 2007; Menkhoff and Taylor, 2007; Pardo, 2008; Chan, 2013) tend to conclude that there is some value and predictive power in it, which contrasts with earlier more skeptical perceptions from academic researchers.

In TA, prices have always been the primary reference of past market activity with which so-called "technicians" could attempt to predict future market sentiment. A belief of TA is that, like physical objects, prices have inertia and, when at rest, they often stay approximately at rest. On the other hand, when in motion, they often stay in motion (Widner, 1998). This is exemplified for instance in the probabilistic mechanical view of market movements proposed by Andersen et al. (2000). In this setup, technical indicators can be seen as combinations of measurements of price velocities and price accelerations. Price levels and price ratio relationships between highs and lows give 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

*Corresponding author
Email addresses: lissandrin.marco@gmail.com (Marco Lissandrin), donnacha@ieee.org (Donnacha Daly), dsornette@ethz.ch (Didier Sornette)

[1]The ideas presented in this work originate from Marco Lissandrin's Master Thesis in fulfillment of his Master of Science Degree at ETH Zürich. Please find the full version at the following url: `http://www.er.ethz.ch/media/publications/phd-and-master-theses.html`

support and resistance levels. Prices tend to bounce on these levels (Garzarelli et al., 2014). On the other hand, the crossing of these levels is interpreted as prices being in motion and are likely to continue along the trend (Kosar, 1991). There is also another approach to predicting future market movements and this is time-based forecasting (Coles, 2011; Miner, 1991). This class of methods tries to identify patterns in time series that should repeat over time.

The statistical problems of insignificant evidence and data snooping were both known since early TA studies (James, 1968; Jensen, 1967). Major methodological innovations came in much later in parallel with increased and more accessible computational power. In statistical testing, the replacement of predefined returns distributions with simulated distributions goes back to the bootstrap approach (Brock et al., 1992). Data snooping used to be checked on out-of-sample data until White (2000) developed an in-sample "reality" check. In this procedure, the best performing trading signal is tested against the simulated performance distribution of the full universe of comparable trading signals. All these technical improvements make the study of TA's predictive power more rigorous, but we should keep in mind that the determination of the predictability in financial markets is far from being 100% accurate (Zhou et al., 2012).

In fact, there is an economic upper bound to TA's predictive power, which is based on a risk-return principle (Ross, 2005; Zhou, 2010). The Efficient Market Hypothesis (EMH) has been for a long time the building block of finance theory, and it contradicts the utility of TA. In an efficient market, prices are always a good estimate of the underlying product because a huge number of competing rational profit-maximizers elaborate all available information independently of each other and will immediately act on it when price divergences appear (Fama, 1995).

Nevertheless, recent empirical studies suggest the existence of speculative bubbles in several commodity markets (Sornette et al., 2009; Gilbert, 2010; Phillips and Yu, 2010; Sornette and Cauwels, 2014, 2015a). Market participants cannot always enter positions when a price correction is yet to come (Gromb and Vayanos, 2010). For example, market exposures in trading books are limited by capital constraints (Shleifer and Vishny, 1997) and by internal risk limits. In addition, even well-informed commodity traders must formulate price expectations based on partial or uncertain data (Gorton et al., 2007; Khan, 2009) and this stimulates the use of rational herding behaviours, which have been described by Devenow and Welch (1996), Bikhchandani and Sharma (2001), Hirshleifer and Teoh (2003). Herding behaviour can also be irrational. Noise traders keep or adjust their positions independently of any changes in commodity fundamentals, based on judgemental biases (Ariely, 2010; Grinblatt and Han, 2005; Penteado, 2013), on positive-feedback mechanisms (Sornette and Cauwels, 2015b), on simple TA rules which can be easily understood also by traders with no fundamental understanding (Gehrig and Menkhoff, 2006), and on cross-asset strategies (Tang and Xiong, 2010). All these aspects deviate commodity prices from fundamental values for periods long enough to disturb the normal decision making processes (UNCTAD, 2011; Filimonov et al., 2014) and may justify the use of TA. In fact, TA has always had a significant and consistent user cohort, also among commodity traders (Smidt (1964); Lukac et al. (1988); Billingsley and Chance (1996); Menkhoff and Taylor (2007)).

New TA techniques have always relied on visual intuition for their development (as readily found in all TA magazines), but this aspect is usually neglected during statistical tests. With the goal of keeping our approach visual, we examine one corner of this vast topic by studying the performance of DeMark chart indicators on commodity markets. In particular, we present a Monte-Carlo based backtesting framework (Aronson, 2007; Masters, 2010) to determine whether three of these individual indicators have statistically significant predictive power. This family of indicators is, commercially speaking, one of the most popular and it is also possible to make use of it as an upgrade in leading financial market terminals such as Bloomberg Professional® and Thomson Reuters® which, combined, cover roughly 60% of the market share (Stafford, 2015). Despite this, no previous study has analysed its effectiveness, although there are other studies available on simpler chart patterns, e.g. Lo et al. (2000). We focus our study on the testing of predictive power on a limited number of indicators knowing that one good market entry indicator is not sufficient for a trader to generate systematic profits. This person will know when to enter positions by combining multiple indicators and considerable market intuition but, most importantly, his exposures will be managed by thorough risk management rules (OriginalTurtles.org, 2003; Covel, 2009).

The first known backtesting of trading signals using historical data belongs to a professional astrologer from Antwerp, who in 1540 tried to distinguish himself by testing his astrological system, which he said could foretell local commodity prices (Ehrenberg, 1928; Lo and Hasanhodzic, 2010). The main idea was to use stars as a way to generate random patterns. 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.

Commodities as an asset class refers to those raw materials that serve as a starting point for our society to create valuable products and services. While this set of raw materials keeps evolving in parallel with society, physical commodity trading has been based on the exploitation of geographical and timing-based market opportunities since ancient times. This translates into using market and logistical knowledge to buy commodities at a discount from producers to then sell them to customers at a premium.

Starting from the 18th century (Hamori et al., 2001), these markets drastically evolved with the introduction of financial contracts (i.e. commodity futures contracts), which enabled producers and consumers to hedge their market risks and to adopt more transparent price formulae. Financial investors are part of the picture because they are willing to carry those exposures that physical participants do not want to have. Nowadays, futures contracts are the basic type of exchange traded financial instrument for commodities, right at the intersection of physical and financial trading.

However, there are still relatively few academic studies of technical analysis on commodity futures markets (Lukac et al., 1988; Lukac and Brorsen, 1990; Roberts, 2003). A possible explanation is that the construction of a continuous price series using futures data is not straightforward. Once a specific contract is used to determine the price for a given trading day, there are multiple ways on when (Carchano and Pardo, 2009; Ma et al., 1992) and how (Masteika and Rutkauskas, 2012; Masteika et al., 2012; NASDAQ, 2013; Pelletier, 2011) to roll-over to the next contract. In addition, contract rolls should also take into account mark-to-market and cash management principles. Commodity markets are much broader than futures contracts. Compared to the total volume of tradable commodities, liquid exchange based futures contracts cover a very small share of raw materials, but there are also hundreds of exchange cleared commodity swaps that cover a huge variety of more specific products. Additionally, their market exposure is the same as for futures, except for the month on which they are pricing. Technically speaking, the problem of dealing with discontinuous time series on forward contracts (either futures or swaps) is crucial for all traders who have exposures on a broad range of traded commodities and it should be addressed in any backtesting method, like here.

The paper is organized as follows. Section 2 describes the DeMark indicators used to generate entry and trading signals. Section 3 enters more into details regarding commodity futures and explains the problem of rolling commodity futures contracts. In addition, it suggests a method to create continuous daily returns starting from separate futures contracts. In section 4, different ways for evaluating the performance of entry/trading signals are discussed with a focus on Monte-Carlo permutation tests. Using a two-dimensional framework based on our Monte-Carlo permutation tests, the main results from DeMark backtests on rolled commodity futures are then presented. Section 5 concludes.

2. Definition of DeMark Indicators to be tested

This is a family of indicators developed over time by Tom DeMark (DeMark, 1997; Perl, 2008). It collects many revised versions of traditional price-based indicators such as moving averages, trendlines, price ratios and Elliot waves, but it also includes other indicators, among which, *Sequential* is arguably

Algorithm 1 Entry Signals

```
 1: for each (new) t-th price bar:
 2: procedure DeMark(P_c,P_h,P_l,n,m,q,p,k,options)
 3:     update Setup's counter (s = s + 1)          ▷ eq. 1
 4:     if s = m then
 5:         compute Setup's range & R               ▷ eq. 2, 3
 6:         update Support/Resistance levels
 7:     end if
 8:     if indicator = ST then
 9:         if no open positions then
10:             if P_c(t) > res(t) then
11:                 open new long position at t + 1
12:             else if P_c(t) < sup(t) then
13:                 open new short position at t + 1
14:             end if
15:         end if
16:     else if indicator = Sequential then
17:         if s = m then
18:             if no active Countdown phase then
19:                 activate new Countdown phase
20:                 reset Coutdown's counter (c = 0)
21:             end if
22:         end if
23:         if active Countdown phase then
24:             check recycle and ending conditions
25:             if Combo then
26:                 update c                          ▷ eq. 7
27:             else (normal Countdown)
28:                 update c                          ▷ eq. 5, 6
29:             end if
30:             if c = p then
31:                 open new position                 ▷ eq. 8
32:             end if
33:         end if
34:     end if
35: end procedure
```

the most renowned. *Sequential*, *Combo* (Sequential's main variation) and *Setup Trend* (ST) all generate long and short entry signals based on Algorithm 1. This pseudocode uses the following inputs: historical closing, high and low prices (P_c, P_h, P_l), a set of parameters given by DeMark (n, m, q, p, k) and the option to choose among slightly different versions. Our description focuses only on long entry signals. Short entry signals can be derived by applying symmetrical conditions.

In row 3 of Algorithm 1, whenever a new price bar is available at time t, a counter will be updated. Its value s, which is initially set to zero, is increased by one ($s = s+1$) each time the following long condition is fulfilled:

$$\forall \, (new) \, t, \; P_c\,(t) < P_c\,(t-n)\,, \tag{1}$$

with $n = 4$ according to DeMark. The counter increases only if there are consecutive closing prices satisfying eq. 1, otherwise it is set back to zero. A parallel counter is running based on a symmetrical short condition. If the long counter increases, then the short counter must be reset to zero, and vice versa. DeMark sets $m = 9$ and whenever $s = m$ then a *Setup* is complete. However, this will not reset the counter back to zero. A Setup ends only when the number of consecutive closes cannot be increased. Fig. 1 shows an example of a long Setup. Each time that $s = m$,

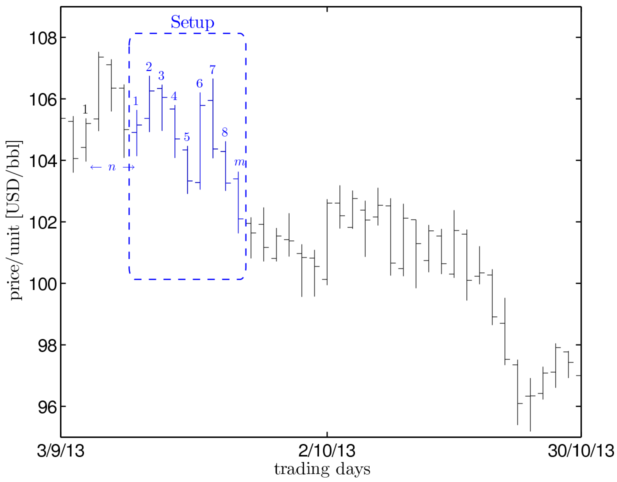


Figure 1: Setup on the Light Crude Oil Futures contract, also known as the *West Texas Intermediate* (WTI) contract.

the price bars related to the newly completed Setup are used in row 5 to determine the Setup's range:

$$\left[\min_{s=1,..,m} \left(P_l\right)_s ; \max_{s=1,..,m} \left(P_h\right)_s \right], \tag{2}$$

and the Setup's width:

$$R_w \triangleq \max_{s=1,..,m} (P_h)_s - \min_{s=1,..,m} (P_l)_s \,. \tag{3}$$

The same price bars also update support (sup) and resistance (res) levels:

$$\forall \text{ Setup}, \begin{cases} (\textit{long}) \ \operatorname{res}(t) = \max\limits_{s=1,..,m} (P_h)_s \,, & \text{(4a)} \\ (\textit{short}) \ \sup(t) = \min\limits_{s=1,..,m} (P_l)_s \,. & \text{(4b)} \end{cases}$$

New long Setups update resistance levels (see fig. 2), whereas short Setups update support levels.

Rows 8-15 of Algorithm 1 describe long and short entry strategies for ST. For example, a new long position will be entered on the next traded day $(t+1)$, at $P_c(t+1)$, as soon as the latest closing price breaks its resistance level $(P_c(t) > \operatorname{res}(t))$, like in fig. 3.

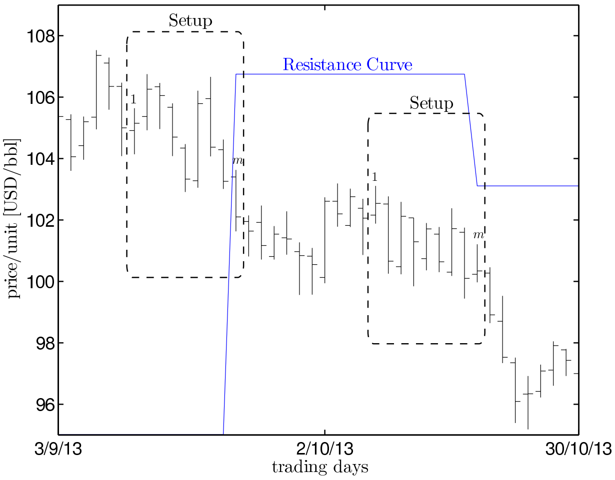


Figure 2: Each completed *Setup* updates WTI's resistance level.

Starting from row 16, the pseudocode explains how Sequential can generate entry signals. At each new completed Setup $(s = m)$, a new *Countdown* phase starts unless one is already active (rows 17-22). Like the Setup phase, also the Countdown can be long or short and has its counter that is set to zero as soon as a new Countdown begins $(c = 0)$. For each new t-th price bar, counter c will increase by one on a long

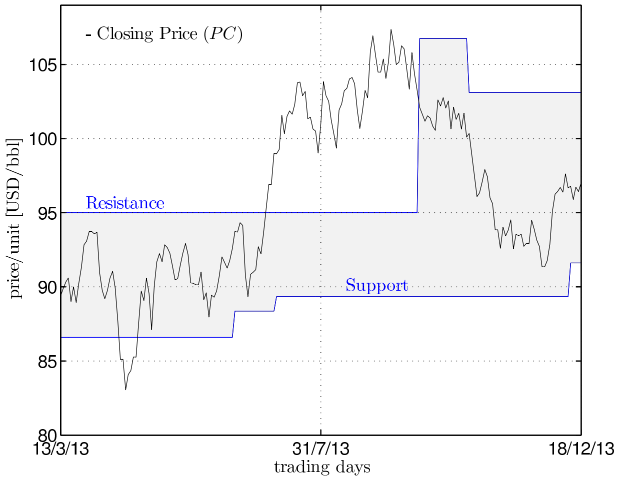


Figure 3: In June 2013, WTI's P_c breaks its resistance level $(P_c(t) > \operatorname{res}(t))$. Therefore, ST (Setup Trend) generates a new long entry signal.

Countdown only if the following conditions are met (rows 27-28):

$$\text{Given } t, \begin{cases} (\textit{standard}) \ P_c(t) \le P_l(t-u) \,, & \text{(5a)} \\ (\textit{aggressive}) \ P_l(t) \le P_l(t-u) \,, & \text{(5b)} \end{cases}$$

and $u = 2$, as suggested by DeMark. Only one of the two variants in eq. 5 should be used. We opt for the aggressive version in our backtests. These conditions are very similar to eq. 1: P_l replaces P_c, u replaces n and here the equality condition is also accepted. In addition, for both cases, the p-th bar completes the Countdown if, given the k-th price bar,

$$P_l(p) \le P_c(k) \,. \tag{6}$$

DeMark suggests to set $p = 13$ and $k = 8$. If eq. 6 is not verified immediately, then the completion of this phase is postponed until this condition is met in one of the later bars. Unlike the Setup, what matters for the Countdown to be completed is the total number p of bars fulfilling eq. 5, not the consecutive number, as in fig. 4. Counter c increases based on conditions that are independent of the Setup, although, during the Countdown (or its alternative, the *Combo*), there are additional checks based on the Setup that can

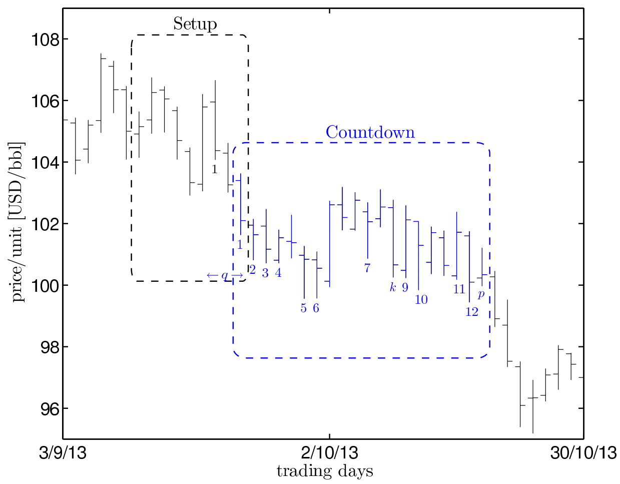


Figure 4: A completed *Countdown* on Light Crude Oil.

restart ("recycle") the phase or, in the worst case, end it before its completion (i.e. row 24):

1. New opposite Setups: A completed Setup in the opposite direction will restart the existing Countdown. For example, a short Setup is completed while a long Countdown is still building up. The long Countdown will end immediately, replaced by a short Countdown with $c = 0$;
2. Crossed support and resistance levels: For example, in case of a buy, if $P_l(t) > \mathrm{res}(t)$, then the Countdown is stopped and cancelled. This condition is similar to the one that generates entry signals using ST, but here closing prices are replaced by high and low prices;
3. New Setups: if there is a new completed Setup in the same direction as the old one, then the Countdown is reset ($c = 0$) only if ${R_w}^{\text{new}} > {R_w}^{\text{old}}$. Nevertheless, if the new Setup has a P_c within the range of the old Setup, then the current Countdown is kept going.

Instead of using eq. 5-6, it is possible to increase counter c by one if all the following *Combo* conditions are fulfilled (rows 25-26):

$$\text{Given } t, \begin{cases} P_c\,(t) \leq P_l\,(t-u)\,(= \text{Eq. } 5)\,, & \text{(7a)} \\ P_l\,(t) \leq P_l\,(t-1)\,, & \text{(7b)} \\ P_c\,(t) < P_c\,(\text{previous Combo bar})\,, & \text{(7c)} \\ P_c\,(t) < P_c\,(t-1)\,. & \text{(7d)} \end{cases}$$

A major difference with the normal Countdown is that the bar check in eq. 7 starts from the first bar of the Setup instead of the last (fig. 5).

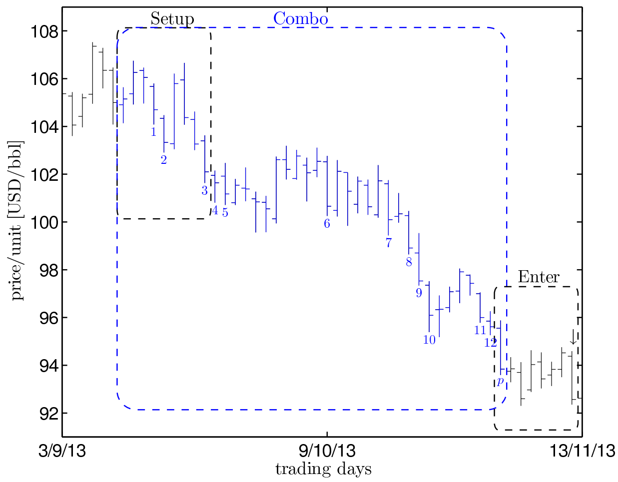


Figure 5: A completed *Combo* on Light Crude Oil.

Once the Countdown (or Combo) is complete at time t, an entry strategy determines when to start a long entry signal (row 31). The *aggressive* strategy enters a long position on the next traded day $(t+1)$, at $P_c(t+1)$. On the other hand, the *conservative* strategy enters at $P_c(t+1)$ as soon as,

$$\mathrm{P}_c\,(t) > P_c\,(t-n)\,. \tag{8}$$

Eq. 8 is similar to the Setup check (eq. 1), but with a reversed test direction. We chose the conservative strategy in our backtests. Anyhow, the choice of the entry strategy does not seem to be the key factor for the performance of the indicator because the turning point has already been identified by a completed Countdown. The entry strategy tries to optimize the timing by delaying an entry signal only for a few price bars (fig. 5).

Sequential (in its traditional Countdown or its alternative Combo version) is a time-based indicator, which tries to identify areas of trend exhaustion that will lead to price reversals. It is made of up to two sequential phases: the first one is the Setup, which tries to capture price momentum, and it is followed by the Countdown that looks for momentum exhaustion. ST (Setup Trend) is a price-based indicator that uses the Setup to determine support and resistance levels. When these are crossed, then prices are in motion and should continue along the trend. The Setup phase (rows 3-7) is a common starting point for both Sequential and ST entry signals.

To generate a long entry signal with Sequential, the Setup has first to identify a bearish momentum in the market. To do so, closing prices (or settlement prices for derivatives) are compared to the close n bars earlier (eq. 1). The idea of using n-days momenta to avoid noise is not unique and can be found also in Chan and Lin (2004). A long Setup is completed when there are m consecutive closes, each one less than the corresponding close n bars earlier. Here the goal is to identify a continuous negative price velocity, i.e. a negative trend. There can still be price oscillations, but a n-days momentum guarantees that the amplitude is small enough and the period is short enough so that there is no interruption of the Setup. According to DeMark, a negative price velocity is measured over four time periods ($n = 4$) and it has to be maintained for nine consecutive periods ($m = 9$) in order to identify a trend. We think that the choice of a 4-days momentum is in line with the behaviour of many traders who are not interested in daily price moves as much as they are for price moves over a few trading days. As no surprise, risk management reports built for traders are often sent out on a weekly basis to limit noisy information and to be in line with the trader's way of thinking. If $m \leq 5$, then we are sure that the closing price of the last Setup bar is lower compared to the first bar so that a negative price velocity has actually moved the price down.

Setting m to 9 means adding additional trend checks to all the first 5 bars: the closing price will trend down with continuity and the price from the first to the last bar will have gone further down because $P_c(9) < P_c(5)$ while $P_c(5) < P_c(1)$. According to this interpretation, m depends on n, while the latter is given a value in line with the traders' way of looking at price moves, all this makes the Setup a very general method to identify incipient trends. Larger oscillations are tolerated during the Countdown because the market price is still trending (with a negative price velocity), but it is decelerating. Deceleration is by no means constant, and this translates into an alternation of negative and positive price velocities. Eq. 5 deterministically suggests that, after $p = 13$ negative velocities, the trend is finished and is ready to revert as soon as the current price level of bar p is below or approximately the same as it was on bar $k = 8$ (eq. 6). Therefore, prices can continue to be bearish or they can go sideways while in the meanwhile the trend exhaustion pattern is building up. Since larger oscillations are tolerated, price velocities have to be measured over a shorter time frame to capture velocity oscillations. This might be the reason why the Countdown's and Combo's $u = 2$ is smaller than the Setup's $n = 4$. On the other hand, the fixed number of p negative moves before prices start to recover might be dependent on specific markets within a defined asset class. It is reasonable to think that DeMark did not focus his studies on commodity markets, but rather on stock markets, given the systematic use of *closing prices* instead of *settlement prices*, which are a common term for commodity futures and swaps.

Note that long Setup Trend signals are based on bearish momenta. There is an acceleration from a zero price velocity to a sustained negative velocity each time a long Setup is complete. Despite a market force pulling the price down, if the current price has the strength to push itself above the resistance level (the highest price level within the latest completed long Setup), then the price is supposed to have enough inertia to continue its motion along the upward trend. When there is an ongoing long open position, we could also use a long completed Setup (i.e. a bearish market force that sets a new resistance level) to exit the position. This is visible in fig. 3: when the price returns into the shaded area, the current long position should be closed. Unfortunately, for Sequential there is no exit signal, therefore we will backtest all of our indicators only on their entry signals.

Lastly, a symmetrical algorithm generates a short entry signal for both Sequential and ST. In reality, there is no symmetry between uptrends and downtrends in the markets (Benyamini, 2009). In fact, selling pressure can take long pauses, but when prices drop they tend do it at a high and constant velocity. On the other hand, buying pressure is relatively even and generates slower and longer uptrends. In commodities, bottoms have larger price oscillations than tops. Given these differences, Sequential seems more suited for long entry positions just after fast price drops have occurred. In case of long-lasting self-sustaining uptrends, the risk for Sequential is to repeatedly suggest short entry positions while the trend is still ongoing.

3. Commodity Futures Markets

3.1. Description

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, exchange for physical (EFP), investment or speculative purposes. 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 producers and deliver them to customers with a premium. In parallel, financial instruments are used as a reference for pricing formulas and as a support for hedging and for EFP-based deals.

Contrary to physical trading, financial investments need exposure to market prices to unlock profit opportunities. Besides an opposite approach to market risk, financial instruments are at the intersection of physical and financial trading. If it was not possible to hedge physical exposures on financial markets (this still applies to some mature, but specific commodity markets), then every physical purchase should be covered by a corresponding physical sale to eliminate market risk. Investors are willing to accept exposures to forward price movements that physical producers, traders (to a certain extent) and buyers do not want to have and their participation in financially based commodity markets increases the liquidity of the related financial instruments. On the physical side, this makes financial hedging more reliable and, regarding deals, a liquid financial contract builds trust on market players, therefore they will be more willing to use price formulae based on such financial contracts.

Futures contracts on commodities constitute the basic type of exchange traded financial instrument. Currently, the most popular contracts cover soft agriculturals, grains, energy products, industrial and precious metals. A commodity futures contract is an agreement to buy or sell a standardized quantity and quality of raw material through an exchange at a future date and at a price agreed upon entering the contract. These instruments can start being traded at least one year before their expiry, but there is no rule for it. 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 the European ICE exchange and covers the physical market for Low Sulphur Gasoil (Diesel 10 ppm) 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 type of delivery, or a different geographical scope (e.g. the Mediterranean) will not have a dedicated futures contract. Despite these differences, a physical trade would probably use the Gasoil ICE futures contracts both as a reference for pricing and as an instrument for hedging. It is still possible to get financial exposures to more specific products by entering exchange-cleared swaps. There are hundreds of these products covered by the ICE and NYMEX exchange. Exposures of such instruments are identical to futures contracts except for the month of expiry when, differently from futures, market exposure decreases linearly along the month until there is no exposure left. This means that the methodology described in this paper is very general because it can potentially cover most of the financially traded commodities.

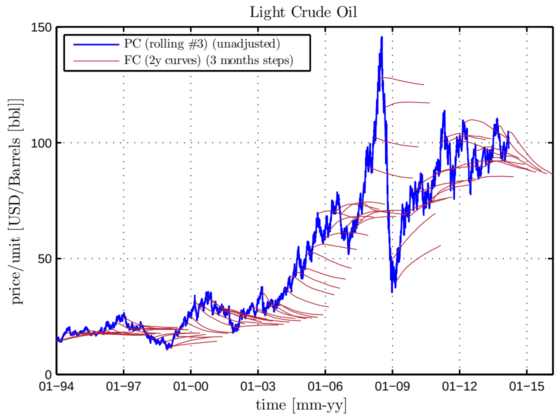
Table 1 provides a visual interpretation of the two dimensions of futures contracts: the *term structure* or *forward curve* (horizontal) and *flat price* (vertical). The latter captures day-to-day price moves within individual contracts, while forward curves show, on a specific date, market prices for future contracts sorted by nearest expiry date.

Table 1: The two dimensional nature of futures contracts on NYMEX Light Crude Oil on $d = 07/02/2014$. Each column contains a separate contract: the front contract M (Mar 2014, first to expire) and the following ones $M+1$ (Apr 2014), $M+2$ (May 2014) and $M+n$ (n=3, Jun 2014). Row d-1 and d-2 show flat prices ($/bbl) for the prior days.

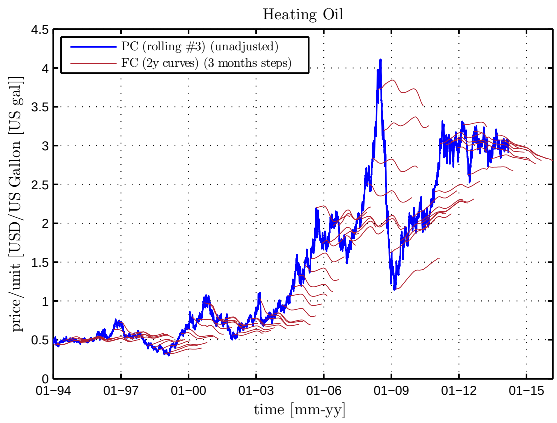
	M	**M+1**	**M+2**	**M+n**	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
d-2	97.38	96.76	96.02	95.22	$\cdots$
d-1	97.84	97.32	96.66	95.94	$\cdots$
d	99.88	99.35	98.62	97.84	$\cdots$

The front part of the curve is the most sensitive to spot price changes. Volatility is higher compared to the end part of the curve and correlations between nearby contracts on the front is lower compared to correlation of nearby contracts at the end of the curve. The most straightforward way for investors to be exposed to commodity market prices is with outright exposure. This means that the forward market structure is ignored and long or short positions are entered just on one column of prices in Table 1, for example on the front contracts to maximize volatility. This trading style is the most common for technical traders. Blue curves in fig. 6 represent continuous flat prices while forward curves are in red.

Despite the product type, the delivery period is a additional dimension in commodity trading. The same product delivered in two different months can be considered as a different asset. A market is in *contango* when a contract that follows the current one is trading at a higher price on the red curve. Otherwise, when a following contract is trading at a lower price, the market is *backwardated.* Given a physical storage, there is a lower time pressure to sell when the forward curve is in contango because, after each month, the evaluation of the product is being rolled to a higher price.



(a)



(b)

Figure 6: Price evolution for (a) Light Crude Oil and (b) Heating Oil. The blue curve represents continuous closing flat prices (PC). It has been constructed by rolling the front contract M to the next M+1 in the last month before expiring as soon as the open interest of M+1 is higher than that of M. Red lines are forward curves, plotted every three months.

For financial outright exposures, the forward curve is not predictive, in fact contract rolls do not generate any PnL from a mark-to-market perspective. Nevertheless, changes on the forward curve might effect flat prices and vice versa. Typical commodity traders speculate mostly without outright, which means that their daily profits do not depend (to some extent) on continuous flat price moves. Physical trading of raw materials needs an understanding of how markets behave when different type of shocks occur (demand/supply, geopolitical, macroeconomic, regulatory, legal and natural disasters). This knowledge helps to understand the behaviour of forward curves, so as to enter physical or speculative time-spreads by betting on relative movements of the curve.

If spreads involve different curves (e.g. long M+1 Heating Oil and short M+1 Light Crude), then there is an additional inter-product risk. In general, outright, time and inter-product exposures can be combined at the same time depending on the traded commodity.

In this paper, we limit our analysis to outright risk. DeMark indicators are being tested over 21 commodity futures markets and 10 years of data (Jan 2004 - Jan 2014). Table 2 lists all futures contracts chosen for the backtesting. Some contracts had to be excluded. DeMark indicators are demanding and require complete price bars, but these were not available for the London Metal Exchange contracts. CBOT Soybean Oil, Soybean Meal and NYMEX Gasoline US could not cover the whole tested period.

Table 2: List of commodity futures used for backtesting.

	Name	**Type**	**Exchange**
1:	Wheat	Grain	CBOT
2:	Corn	Grain	CBOT
3:	Oats	Grain	CBOT
4:	Soybean Meal	Grain	DM.
5:	Cocoa	Soft	ICE US
6:	Coffee C	Soft	ICE US
7:	Sugar #11	Soft	ICE US
8:	Cotton #2	Soft	ICE US
9:	Light Crude	Energy	NYMEX
10:	Nat. Gas	Energy	NYMEX
11:	Heating Oil	Energy	NYMEX
12:	Brent Crude	Energy	ICE EU
13:	Nat. Gas	Energy	ICE EU
14:	Gasoil	Energy	ICE EU
15:	Aluminium	Ind. metal	SHFE
16:	Copper	Ind. metal	COMEX
17:	Copper	Ind. metal	SHFE
18:	Gold	Prec. metal	COMEX
19:	Silver	Prec. metal	COMEX
20:	Platinum	Prec. metal	NYMEX
21:	Palladium	Prec. metal	NYMEX

3.2. Rolling Futures Contracts

A continuous price series of a traded futures must have one contract selected for each trading day. Prices belong to the front part of the curve (M, M+1, M+2) to capture spot price movements. This part of the curve is also the backbone of physical trading because it is used to evaluate and hedge unsold material. It is not possible to stick to one contract for the whole backtested period because futures contracts can be traded only for a few years and cannot cover the whole backtest. When a contract starts being traded, it represents the most long-term future price expectation with a weak dependency on front price changes, a lower volatility compared to the front contract and a lower liquidity. The simplest way to build a continuous flat price curve is to roll following the last traded day of the latest expired contract. A more sophisticated method should anticipate the roll because contracts in their last weeks of life show abnormal volatility (Samuelson, 1965). Carchano and Pardo (2009); Ma et al. (1992) suggest to roll the front contract 1-2 weeks before maturity, or on the first day of the expiring month, or, as an 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 on M+n+1 is higher than on M+n. Data providers suggest similar solutions (Reuters, 2010) with the addition of rolling methods that roll from M to M+1 based on weighted values. In this paper, the following rolling strategies have been considered:

1. from M to M+1, following M's expiry;
2. from M to M+1, 10 days before M's expiry;
3. from M to M+1, when $OI(M) < OI(M+1)$;
4. always on the contract with maximum OI.

It is not allowed for any strategy to roll back to the previous contract, i.e. from M+1 to M.

Results presented in this paper are based on rolling #3. While rolling #3 seems a reasonable choice for our backtests (see section 4.2 where we study the impact of the different rolling methods on the tests), there is still room to further investigate the impact of rolling strategies.

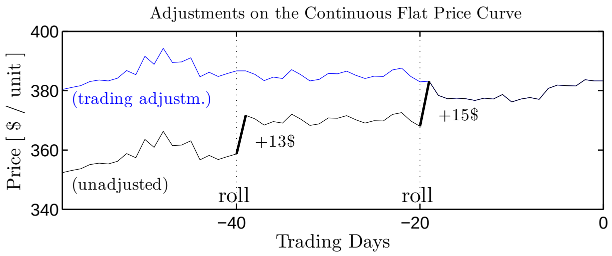


Figure 7: The black (unadjusted) curve is a generic continuous flat price curve, generated by one of the rolling strategies described in the text. Given the price level and the market's contango, it could also represent a very steep front contango market on ICE Gasoil after a flat price crash. Trading day 0 is the most recent trading day of the time period and the plot goes backwards in time, from right to left. The black curve shows price discontinuities on the rolling day. The blue curve, which is obtained as explained in the text to be neutral to contract jumps, is used to generate signals and to test their performance.

This is an additional issue that we have to address before presenting the extensive results of our tests. When each trading day has a contract linked to it, then continuous settlement prices still show misleading price discontinuities on the rolling day, like on the black curve in fig. 7. DeMark indicators use daily flat price movements to generate entry signals and a price movement in the wrong direction could stop a signal before its completion. Therefore, any DeMark indicator should not consider price discontinuities coming from contract rolls as price moves that can generate profits.

A simple way to remove discontinuities would be to use local adjustments by weighting price values across multiple trading days. Nevertheless, there are more computationally intensive adjustment techniques which fit better with our backtests. The black unadjusted curve in fig. 7 shows a steep front contango market, in fact the front contract M is rolled twice to a higher valued M+1 contract. Assuming that the black curve is related to a long position that we want to keep, we should sell a lower valued contract and buy the next contract that always trades at a higher price (+13$ and +15$). From a mark-to-market perspective, changing the contract creates no Profit & Loss (PnL) if we exclude transaction costs. For this reason, entry signals are being generated on the blue trading curve, which is neutral to discontinuities coming from the rolling process.

There are multiple ways to adjust discontinuities on continuous flat price time series (Masteika and Rutkauskas, 2012; Masteika et al., 2012; NASDAQ, 2013; Pelletier, 2011). There is general consensus that prices in the old contract should be adjusted to prices in the new contract. In this way, continuous prices that are later in time should be less affected by adjustments. On the other hand, this backward-adjustment is more computationally intensive than a forward-adjustment because a change on a single roll-over requires additional changes on all previous trading days.

Fig. 7 uses backward-adjustments using a discrete translation of the price by the adjustment Δ on the day of the roll such that:

$$\Delta = Po_d - Pc_{d-1}. \tag{9}$$

The blue curve, which is neutral to rolling, is obtained by adding the adjustment Δ to all prices (P_c, P_o, P_h, P_l) prior to the rolling day.

Another method is the proportionally adjusted method. Instead of adding a fix quantity Δ, the proportionally adjusted method multiplies all the prices prior to the rolling day by the quantity ρ:

$$\rho = {}^{Po_d}/_{Pc_{d-1}}. \tag{10}$$

Using such ratios has the advantage that negative prices are not possible by construction. On the other hand, they may create large reconstructed price fluctuations. This unwanted features lead practitioners to favour the use of the Δ approach over the ρ approach.

In this paper, continuous flat price curves have been backward-adjusted using Δ quantities. This method generates negative prices on the trading series (in blue) 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 trading time series because DeMark indicators only use relative prices to build up entry signals. However, this is a problem when evaluating the performance of strategies in comparison with that of the commodity, defined by its daily market returns at day d:

$$r_d = \frac{Pc_d - Pc_{d-1}}{Pc_{d-1}} \ . \tag{11}$$

The blue curve in Fig. 7 cannot be used directly to compute daily returns because settlement prices (P_c) can be negative or close to zero values and this would distort daily returns. A simple solution is to use the blue curve to compute the nominator in eq. 11 while the denominator uses settlement prices from the black unadjusted continuous flat price curve. In fact, the blue curve shows correct changes in relative prices, while the black curve refers to the market's absolute price levels.

Finally, cash management plays a role on the PnL: rolling to a contract with a higher value means that initial margin requirements will be higher whether the position is long or short. Without considering any portfolio effect, higher collaterals may force trading businesses to borrow more money for which additional interest rates should be paid. In this paper, performance is measured on the blue trading curve (i.e. daily market returns), while transaction and financing costs are not included.

4. Market performance of DeMark indicators

4.1. Methodology

A positive trade record requires profit generation, but trades should also beat the markets on a risk adjusted basis in order to be attractive for investors. If indicators generate signals that outperform the market, then those indicators are informative, in other words, they have predictive power.

In this paper, we study the predictive power of each separate indicator. This is not sufficient to guarantee the profitability of an indicator, but it is already a first step towards it. Profitability requires a much broader discussion that we do not address here. In fact, traders use their intuition and their research skills to enter new positions (e.g. by combining several trusted fundamental and technical indicators), but profitable traders are also excellent exposure managers. They handle multiple diversified strategies while keeping their portfolio volatility always under control and they know when it is the best moment to cut losses and to lock-in profits.

Predictive power can be studied on a single indicator by comparing conditional distributions of returns (conditioned on entry signals) with unconditional distributions of returns representing the market's performance (Lo et al., 2000). A further step in the analysis could be to substitute conditional returns with a conditional return-to-risk ratio, for example the *Risk-Return-Ratio* (*RRR*) (Johnsson, 2010). The conditional versus unconditional returns approach cannot combine longs and shorts on the same graph, but it is easy to implement and it offers graphical intuition. For example, fig. 8 suggests that long Sequential entry signals might over-perform the Cocoa futures market when holding days are fixed to 13-14 days.

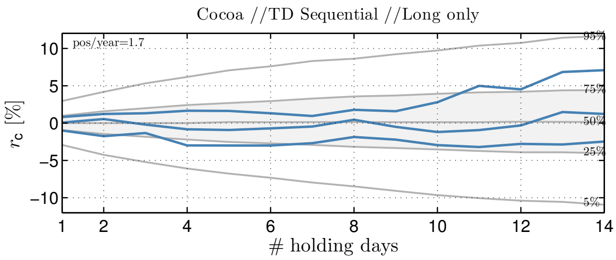


Figure 8: The conditional distribution of returns on Cocoa for long Sequential entry signals represented as 25%, 50% and 75% quantiles (blue curves) is compared to the market's unconditional returns distribution (in grey).

Unfortunately, with a p-value of 0.64 in the Kolmogorov-Smirnov test, the conditional distribution does not seem significantly different from the unconditional distribution when the number of holding days is set to 14. A limit of this test is that it only tests the maximum gap between the two cumulative distributions without considering that each conditional quantile is overperforming its corresponding market quantile. In addition, the test assumes Independent and Identically Distributed (IID) returns, which is not plausible for financial data (Lo et al., 2000). So the failure of the Kolmogorov-Smirnov test to reject the null hypothesis, that the long Sequential entry signals do not over-perform the Cocoa futures market, may just be a result of its lack of power.

As no surprise, Monte-Carlo permutation tests (Aronson, 2007; Masters, 2010) in figures 9, 10, 11 provide a different conclusion, much more in line with the previous graphical intuition. According to the null hypothesis H_0 the signal's long, short and neutral positions are paired randomly with daily market returns. The alternative hypothesis H_A supports the idea that the current pairing improves performance beyond what could be expected from randomness, which also means that the indicator is intelligent and has predictive power over the market. The value of such random strategies that have the same characteristics as the prediction system to test, except for the timing, is also explained by Daniel et al. (2009) in the context of avoiding selection biases, survival biases and look-ahead biases during backtests.

Randomized signals have several constraints. All possible permutations, including the candidate signal, must have both equal chances of appearing in real life and in the randomization process. For example, the total number of trades and the total number of trading days has always to be the same. An additional constraint in our study is that each trade must be held for the same fixed number of days because we examine the predictive power of entry signals by sweeping the number of holding days. This means that, by design, the candidate signal contains trades with a fixed number of holding days. Moreover, to avoid possible interactions, trades should not overlap in time. Lastly, each daily return needs to be paired with a long, short or neutral market position.

On the one hand, compared to the bootstrap method, permutation tests have more requirements to fulfil. On the other hand, these tests overcome many bootstrap weaknesses. There is no assumption on the null-hypothesis distribution used for the test, while in bootstrap tests the empirical distribution of the obtained sample is assumed to be representative of the whole population. Furthermore, bootstrap tests re-sample returns with replacement while the trading signal is kept unchanged. This is how the null hypothesis distribution is generated. Instead, during permutation tests, daily returns are kept on their historical positions while the trading signal is permuted (without replacement). In this way, the intelligent part of the returns is left untouched together with its behavioural and statistical dependencies (e.g. autocorrelations).

The total number of signal permutations grows very quickly given the number of trading days, given the number of entered trades and given the number of holding days for each trade. Let us assume that there are a total of 30 trading days and that the signal can only be long or out of the market. If, with all the mentioned constraints, there is only one trade that lasts 3 holding days, then there are only 28 possible permutations. If instead there are 5 trades with the same duration, then the number of signal permutations goes quickly up to $\sim 10^{4.2}$. In this paper, 400 randomized trading signals are simulated from each candidate signal, on each test. It is a trade-off between quantile smoothness, size of confidence intervals and computational power. Permuted distributions are approximated, therefore observed quantiles need to be adjusted according to confidence intervals before being used for hypothesis testing. Let q and $\hat{q}$ be respectively the true and the observed quantile. These transformations use a normally approximated binomial method in accordance with Conover (1999). A limitation of this method is that it requires large samples, but this is not a problem for our permutation tests. Its strength is that it can be applied to any quantile. Given the number of randomized signals n_0, the observed quantile $\hat{q}$, the confidence level α and Z_α as the Z-statistic (e.g., $Z_{1-\alpha} \sim 1.65$ when α=95%), then the true quantile q has the following confidence interval $\hat{q} - \varepsilon \leq q \leq \hat{q} + \varepsilon$ with:

$$\varepsilon = \frac{Z_{1-\alpha} \cdot \sqrt{n_0 \cdot \hat{q} \cdot (1 - \hat{q})}}{n_0}. \qquad (12)$$

With $n_0 = 400$, α=95% and $\hat{q}$=95% then $93.2\% \leq q \leq 96.8\%$. When a p-value refers to a theoretical $q = 95\%$ quantile, we will conservatively pick a $\hat{q}$=97% measured quantile from the empirical distribution. To avoid approximation, it is generally recommended to use the "exact" Clopper-Pearson (1934) confidence interval, which is well described by Agresti and Coull (1998). Both methods provide the same rounded results for ε. All the quantile adjustments adopted in the permutation tests are listed in tab. 3.

Table 3: Quantile transformations in the Monte-Carlo permutation test. Null-hypothesis distributions are approximated, therefore observed quantiles need to be conservatively adjusted according to confidence intervals. q and $\hat{q}$ are respectively the true and the observed quantile.

Left Tail		
$q = 2.5\%$	$\rightarrow$	$\hat{q} = 1\%$
$q = 5\%$	$\rightarrow$	$\hat{q} = 3\%$
$q = 10\%$	$\rightarrow$	$\hat{q} = 7\%$
Right Tail		
$q = 90\%$	$\rightarrow$	$\hat{q} = 93\%$
$q = 95\%$	$\rightarrow$	$\hat{q} = 97\%$
$q = 97.5\%$	$\rightarrow$	$\hat{q} = 99\%$

Table 4: Measured quantiles for long Sequential on Cocoa.

	$\mathbf{Profit_{trade}}$	P_f	$RRR_{\mathbf{trade}}$
13 holding days:	95%	95%	87%
14 holding days:	95%	92%	90%

Each trading signal uses an aggregate measure to determine its performance. In our case, those measures can be mean values over single positions like mean returns and mean RRRs (called respectively $\mathrm{Profit_{trade}}$ and RRR_{trade}), but also measures that are computed only over multiple positions, like the Profit Factor. This is a profit-to-loss ratio which is defined as

$$P_f \triangleq \frac{\sum \text{Profits}}{\sum \text{Losses}}. \tag{13}$$

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

$$P_f = \frac{\sum r^+}{\sum r^-} = \frac{N^+ \cdot \overline{r^+}}{N^- \cdot \overline{r^-}}, \tag{14}$$

where N^+ and $\overline{r^+}$ are respectively the number of win-

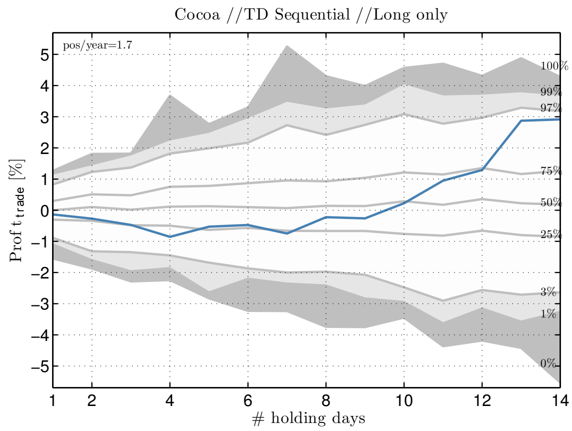


Figure 9: $\mathrm{Profit_{trade}}$ permutation tests for long Sequential on Cocoa.

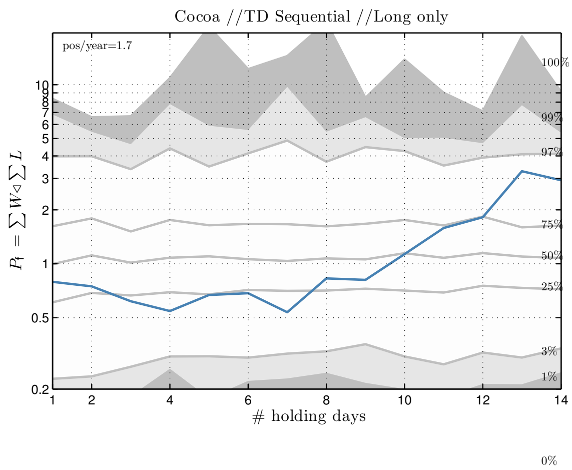


Figure 10: P_f permutation tests for long Sequential on Cocoa.

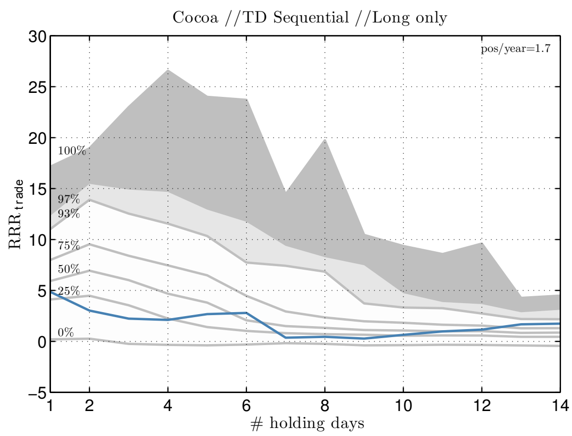


Figure 11: RRR_{trade} permutation tests for long Sequential on Cocoa.

ning trades and their mean return. On the other hand, N^- and $\overline{r^-}$ are the number of loosing trades and their mean return. A trade is considered a winner when its gross return is strictly positive. Next, given the definitions of *Payoff Ratio* and *Win Ratio* as:

$$P_r = \frac{\overline{r^+}}{\overline{r^-}} \quad \text{and} \quad W = \frac{N^+}{N^+ + N^-}, \tag{15}$$

the profit factor can be then written to obtain

$$P_f = \frac{W}{1 - W} \cdot P_r. \tag{16}$$

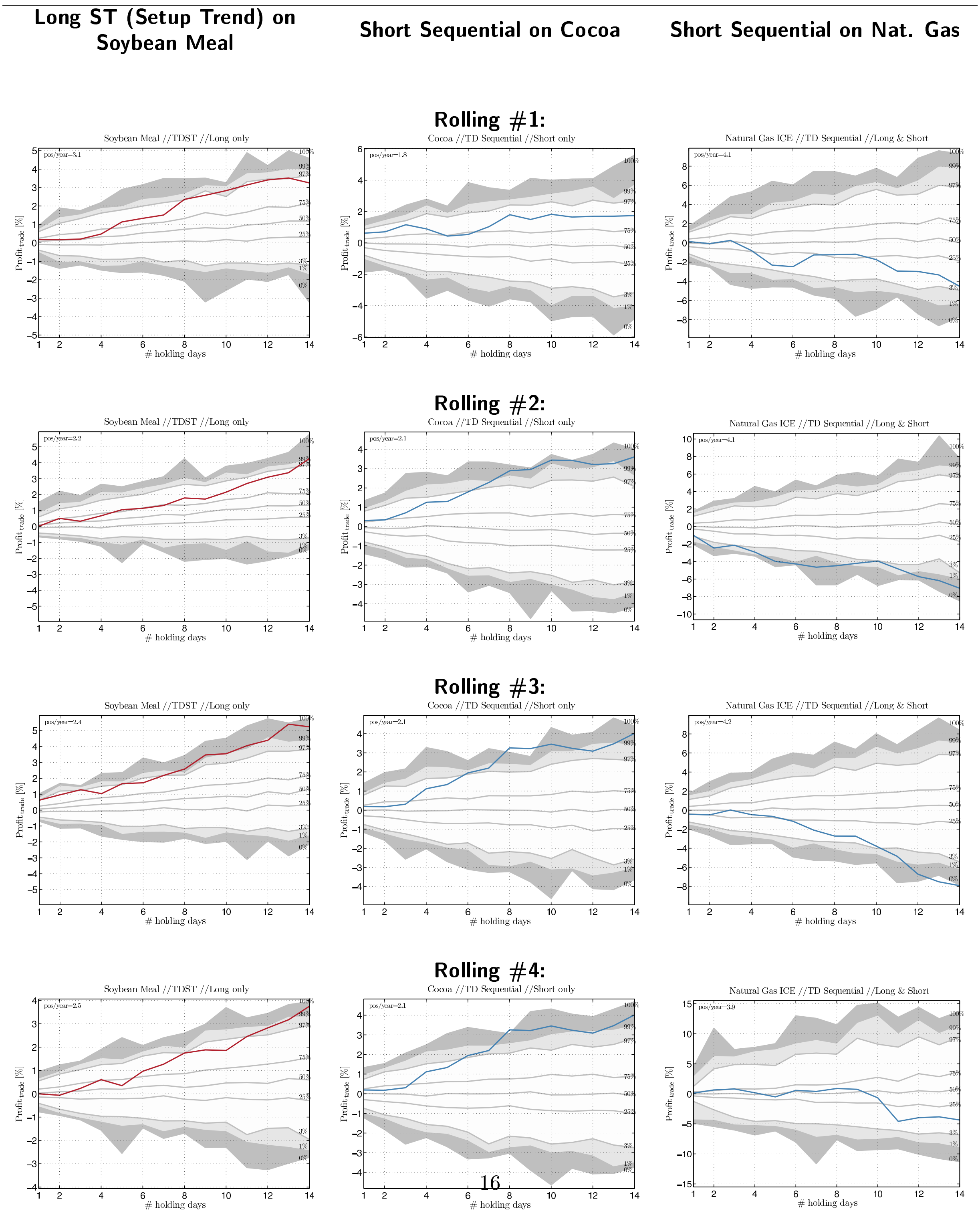
A trading signal generates profits when $P_f > 1$ and, by definition the break-even is reached when $P_f = 1$. The desired condition is often $P_f > 2$ (Harris, 2009). Eq. 16 highlights how signals with low P_r must have a high W to generate profits, like for intra-day trading. On the other hand, high P_r values coupled with a low win ratio (i.e. $W < 50\%$) can still be profitable. $\text{Profit}_{\text{trade}}$, P_f and RRR_{trade} are computed for each candidate and randomized signal. Performance measures such as $\text{Profit}_{\text{trade}}$ and P_f use double-sided tests. When the observed Profit Factor is below the break-even ($P_f = 1$) and the random distribution is performing significantly better, then this condition is informative and it might suggest to flip the direction of the trades. RRR_{trade} uses a single-sided test instead. RRR's definition uses the maximum drawdown 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.

Results for long Sequential tests on Cocoa are shown in figures 9, 10 and 11. In the current backtesting, a p-value of 10% translates into a *possibly significant* rejection of H_0 and a value of 5% into a *statistically significant* rejection. The light shaded area represents statistical significance and the dark shaded area represents high statistical significance at the 99% level, both of them use conservatively adjusted quantiles by taking into consideration confidence intervals. Looking at possible significance, the unadjusted quantiles on single-sided tests ($q = 90\%$) and double-sided tests ($q_1 = 5\%$, $q_2 = 95\%$) have been adjusted to $\hat{q}$=93%, $\hat{q_1}$=97% and $\hat{q_2}$=97%. This means that the candidate indicators (in figures 9, 10 and 11, and in table 4) is never statistically significant because the use of conservative quantile transformations on all performance measures is always decisive to downgrade possibly significant results to non significant when each position is being held for 13 or 14 days. However, the indicators are clearly *possibly significant*, i.e. above the 90% confidence level, except for RRR_{trade} for 13 holding days.

4.2. Impact of the rolling method on performance results

Fig. 12 examines the impact of different rolling strategies for long ST (Setup Trend) on Soybean Meal, short Sequential on Cocoa and long & short Sequential on Natural Gas ICE. When the roll is done on the expiry day, then the number of statistically significant holding days decreases compared to rolling #3 and in two examples there is even no statistical significance left. In practice, the rolling strategy plays a key role in evaluating the statistical performance of a trading indicator similar to DeMark on commodity futures markets. Rolling #1 is more conservative: it tends to accept (rigorously speaking, it tends not to reject) the null hypothesis more easily compared to the other rolling strategies. This also means that a basic backtesting on commodity prices series using rolled prices based on rolling #1 might underestimate the predictive power of an indicator. Furthermore, rolling #1 always has the lowest $\text{Profit}_{\text{trade}}$ values (in magnitude) among the four rolling strategies. Rolling #2 and #3 have a similar number of trades, similar $\text{Profit}_{\text{trade}}$, 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 #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. In contrast, rolling #3 and #4 show very different behaviours for Natural Gas ICE. This can be explained by the strong seasonality of Natural Gas contracts. The roll is discontinuous: some contracts are more important and can be used for long periods

Figure 12: Impact of rolling strategies on the indicator's $\text{Profit}_{\text{trade}}$ performance and statistical significance.



in the continuous flat price curve, while others can be completely ignored.

4.3. Synthesis of the performance tests

Three DeMark indicators have been tested to generate long and short outright entry signals on commodity futures, which, as mentioned before, are the basic type of exchange based financial instrument within traded commodities. The number of holding days is swept and this allows to study trade performances during the days that follow entry signals, just like in fig. 9, 10 and 11.

Given a fixed number of holding days, as before, we consider the following performance metrics introduced in section 4.1: (i) mean return: $\text{Profit}_{\text{trade}}$; (ii) Profit Factor P_f defined in equation (13) and (iii) *Risk-Return-Ratio* RRR_{trade}. If such a performance metric is measured to go beyond what can be expected by randomness by having a value within the shaded area of previous figures, then the indicator is predictive. Ideally, measured performance should be within the dark shaded quantiles for each holding day. In this way, positions can be entered with a delay while still being on the correct side of the trade and they can be closed whenever the trader feels more comfortable.

Let us assume instead that there is a limited range of holding days for which the indicator has predictive power and that we want to replicate the encouraging results by trading with it. As soon as the entry signal is given, a position will be taken. This might not be possible if limits on margin requirements or market exposures have been reached already. Regarding when to exit the position, we might face an uncomfortable situation in which a choice has to be made between taking profit or waiting till the end of the ideal holding period. None of these problems would arise if the indicator would provide statistically significant entry signals regardless of the holding period.

By design, Monte-Carlo permutation tests described in section 4 use the performance metrics to test the indicator's predictive power over the holding days. To quantify the information contained in such figures as fig. 9, 10 and 11, we introduce the *stability* σ of an indicator's predictive power over a specific market, which is defined as the percentage of holding days following entry signals that show statistically significant performance, with 100% representing the optimal case. Nevertheless, the most suited indicator for a given market should jointly maximize σ, this is the priority, and the performance measure, also called profit potential. The latter is computed as the average of $\text{Profit}_{\text{trade}}$ values that belong to statistically significant holding days, if there are any. It seems a simple but effective way to limit the data mining effect that would occur if the maximum $\text{Profit}_{\text{trade}}$ value within the 14 holding days would be chosen to represent the profit potential of an indicator.

This two-dimensional evaluation framework with σ (% of significant holding days) on the horizontal axis and profit potential on the vertical axis (i.e average $\text{Profit}_{\text{trade}}$) has been applied to the DeMark indicators over the 21 commodity futures markets listed in table 2. In fig. 13(a), indicators are evaluated only for their long entry signals, in 13(b) only for short entry signals and in 13(c) for both longs and shorts. The most interesting indicator-market combinations are represented in the shaded areas of fig. 13. DeMark indicators have sparse entry signals. Across all commodities, each indicator enters the markets with the frequencies shown in table 5.

Table 5: Across all commodities, each indicator enters the markets with the following frequencies (total positions/ 10 years):

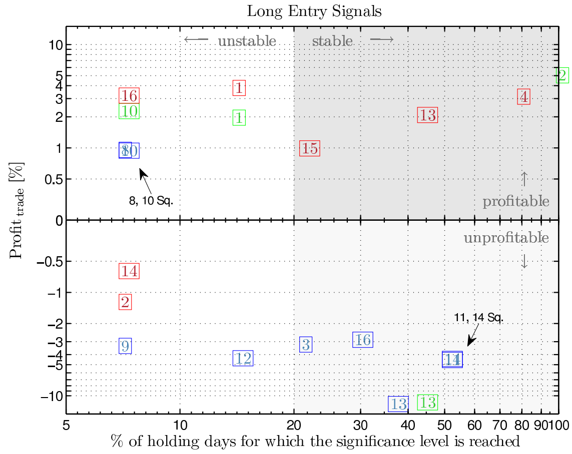
pos./year	**Sequential**	**Combo**	**ST**
minimum:	3.0	1.2	3.4
mean:	3.9	1.7	4.5
maximum:	4.6	2.2	6.1

ST (Setup Trend) provides the highest number of trades and is predictive on a few markets, especially with long entry signals. In addition, it captures correctly the direction of the market and this is always true when both long and short positions are possible. Sequential and its Combo variant generate less entry signals compared to ST. As explained in section 2, Sequential and Combo need every time a new Setup before completing an entry signal while ST uses the Setup phase only to update support and resistance curves. If we also consider that these two indicators need to fulfil additional conditions during the Countdown phase or the Combo phase, then it takes ideally around one month before an entry signal can be completed on Sequential or Combo. Quite interestingly, Sequential has shown statistically significant performance for either long or short positions on all commodity futures apart from Platinum. Unfortunately, just like its Combo version, this indicator does not seem to be able to forecast the correct market direction. This limits our trading strategies, but forthcoming price move in an unknown direction can still be profitable, for example by using a long straddle option strategy. Looking at fig. 13(a), long Sequential on energy products is more a trend detector that a turning point detector and the market is more likely to continue its negative trend rather than to shift to a new positive trend.

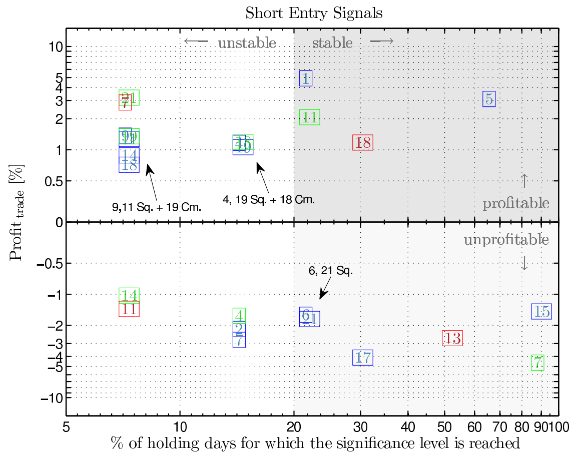
Therefore, Setup can identify negative trends, but markets seem to have a price inertia that is too big to be captured by the current Countdown parameters. Indicator-market combinations that in fig. 13 appear in the stable but unprofitable area can still be informative if interpreted correctly.

So far, fig. 13 has provided a high level overview of the behaviour of the indicators on several markets. The framework can be studied in more details by comparing results from different performance measures within the permutation test setup, like in fig. 14. Conditional distributions of returns versus unconditional returns can be used as an additional countercheck.

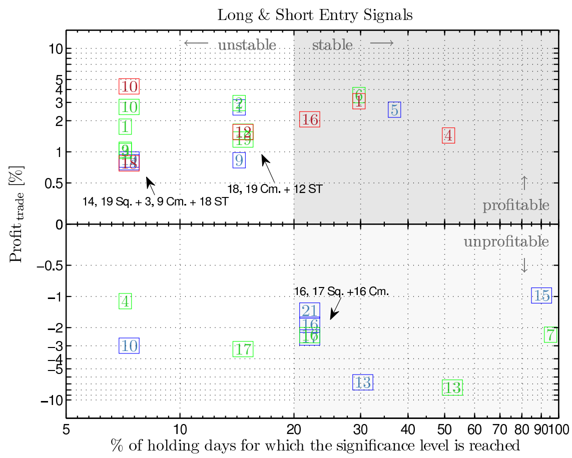
For example, fig. 13(a) highlights a good performance for long ST (Setup Trend) entry signals on soybean meal. A more in depth analysis based on fig. 14 shows that $\text{Profit}_{\text{trade}}$ grows regularly over the holding period and, at the same time, it stays steadily in the shaded area which represents a possible or statistically significant result. Furthermore the quantiles of conditional returns are also steadily above the corresponding quantiles generated from unconditional market returns. These observations are partially confirmed by additional permutation tests using different metrics. While results are similar when P_f is being used, RRR_{trade} is in the region of significance too, but less frequently. The same kind of thinking should be applied to the other examples. Sequential on Cocoa is shown as a successful example for short



(a)



(b)



(c)

Figure 13: Evaluation framework that shows for each indicator its σ on the horizontal axis (% of significant holding days) and the corresponding profit potential on the vertical axis (average $\text{Profit}_{\text{trade}}$). The commodity market is specified by a number which refers to table 2. Blue colour represents Sequential (Sq.), green is for Combo (Cm.) and red is for ST.

positions. If we also recall the interesting results for long positions previously described in section 4, then Sequential is a predictive indicator also when both long and short positions are considered, as confirmed by fig. 13(c). The last example in fig. 14 shows statistical significance for long and short ST entry signals on wheat and it shows how results might depend on the holding period. Across all metrics, there is nice predictive power, but this only appears 11-12 days after the position has been entered.

5. 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 Setup Trend) has been tested from Jan. 2004 to Jan. 2014 on 21 commodity futures belonging to the following classes: grains, softs, energy, industrial metals and precious metals.

Most academic studies test indicators that are not frequently used in practice. An original aspect of the present work is to test indicators that are currently among the most popular ones. These indicators are indeed found and can be used on leading financial platforms.

Mr. 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. ST (Setup Trend) 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.

For each trading day, DeMark indicators need bar charts (starting, highest, lowest and closing prices). With the average trade frequency (3.9 trades per year for Sequential, 1.7 for Combo and 4.5 for ST), 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 have been computed from discontinuous commodity futures markets. This has been 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 also the most straightforward way for investors to get exposed to commodity 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 flat price movements of a contract and it is usually referred to as outright trading. Traditional commodity trading instead makes a limited use of this trading style. In fact, typical commodity traders speculate mostly without outright exposures (time-spreads and inter-product spreads are their bread and butter), which means that their daily profits do not depend, to a certain extent depending also on correlations between contracts, on flat price movements.

The predictive power of each indicator has been studied in two steps. The first one was to compare conditional returns on entry signals to exact unconditional return distributions (which represent the market). An over-performance of the conditional distribution compared with the market suggests that the tested indicator might have predictive power. The second step has used approximated permutation tests to check if the initial suggestion is correct.

We find that all three indicators exhibit predictive power on some commodity futures. Most of the times, the entry signals provided by the indicators show 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. For all the commodity futures apart from Platinum, DeMark's Sequential 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. ST 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 ST 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 because it is not sufficient to over-perform the market if still no profits can be made. This seems a more complete approach compared to only choosing markets where indicators are just over-performing, 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 the present work is the study of 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 that are close to expiry to capture both spot price and short term expectations, but there are potentially infinite ways to roll over to the next contract. Four different rolling 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 rolling strategy, the average profit per trade is always the lowest compared to the other rolling strategies.

A natural extension of this work should investigate the data snooping bias, or selection bias (Daniel et al., 2009). The higher is the number of indicators being tested on the same historical data-set or the higher is the number of data-sets tested on the same indicator, the higher is the probability that luck had an impact on the most statistically interesting results. Regarding this topic, Kuang et al. (2014) review and compare a broad range of data snooping tests on 25'988 trading rules over 9 currency markets, from White's reality check (White, 2000) to more elaborated schemes. In our case, results do not seem to be driven by luck due to a relatively low number of combinations (3 trading rules over 21 data-sets). In fact, Sequential provides statistically relevant results (either for long, short or long/short entry signals) on almost every tested market.

Further developments of this work should include backtests of the ST 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 inter-product 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 A. Parameters used during the backtests

Table A.6: List of parameters used during the backtests

Backtested Period	
start:	1/1/2004
end:	1/1/2014
trading days:	2610
DeMark Parameters	
m	9
n	4
Sequential & Combo:	
p	13
q	4
k	8
aggressive Countdown	on
Entry Strategy:	
conservative	on
Rolling Strategy	
Main:	#3
Others:	#1, #2, #4
Quality Measures	
Main:	$\text{Profit}_{\text{trade}}$
Others:	P_f or RRR_{trade}
List of Commodities	
See tab. 2	
Significance Test	
Type:	
Permutation Test	
Parameters:	
# of holding days	1-14
# of permuted signals	400
p-values	5%, 10%
Confidence intervals:	
Clopper-Pearson	
Conover (1999)	

References

Agresti, A., Coull, B. A., 1998. Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician Vol. 52, 119–126.

Andersen, J. V., Gluzman, S., Sornette, D., 2000. Fundamental framework for technical analysis. European Physical Journal Vol . B 14, 579–601.

Ariely, D., 2010. Pedictably irrational. New York and London: Harper Perennial.

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

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

Bikhchandani, S., Sharma, S., 2001. Herd behavior in financial markets. IMF Staff Papers Vol. 47:3, 279–310.

Billingsley, R., Chance, D., 1996. Benefits and limitations of diversification among commodity trading advisors. Journal of Portfolio Management Vol. 23, 65–80.

Brock, W., Lakonishock, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance Vol. 47, 1731–1764.

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

Chan, E. P., 2013. Algorithmic trading: winning strategies and their rationale. John Wiley & Sons.

Chan, L., Lin, L., 2004. Boosting profitability. Technical Analysis of Stocks & Commodities Vol. 22:8, 28–32.

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

Conover, W. J., 1999. Practical Nonparametric Statistics. Wiley.

Covel, M. W., 2009. The Complete Turtle Trader. HarperCollins.

Daniel, G., Sornette, D., Woehrmann, P., 2009. Look-ahead benchmark bias in portfolio performance evaluation. Journal of Portfolio Management Vol. 36:1, 121–130.

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

Devenow, A., Welch, I., 1996. Rational herding in financial economics. European Economic Review Vol. 40:3-5, 603–615.

Ehrenberg, R., 1928. Capital and Finance in the Age of the Reinassance: A Study of the Fuggers. Hardcourt Brace.

Fama, E., 1995. Random walks in stock market prices. Financial Analysts Journal Vol. 51:1, 75–80.

Filimonov, V., Bicchetti, D., Maystre, N., Sornette, D., 2014. Quantification of the high level of endogeneity and of structural regime shifts in commodity prices. The Journal of International Money and Finance Vol. 42, 174–192.

Garzarelli, F., Cristelli, M., Zaccaria, A., Pietronero, L., 2014. Memory effects in stock price dynamics: evidences of technical trading. Nature, Scientific Reports Vol. 4, Article number: 4487.

Gehrig, T., Menkhoff, L., 2006. Extended evidence on the use of technical analysis in foreign exchange. International Journal of Finance and Economics Vol. 11:4, 327–338.

Gilbert, C., 2010. Speculative influences on commodity futures prices 2006-2008. Tech. rep., UNCTAD Discussion Paper No.197.

Gorton, G., Hayashi, F., Rouwenhorst, K., 2007. The fundamentals of commodity futures returns. working paper no. 13249. Tech. rep., National Bureau of Economic Research (NBER).

Grinblatt, M., Han, B., 2005. Prospect theory, mental accounting, and momentum. Journal of Financial Economics Vol. 78:2, 311–339.

Gromb, D., Vayanos, D., 2010. Limits of arbitrage. Annual Review of Financial Economics Vol. 2:1, 251–275.

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

Harris, M., 2009. Profitability and Systematic Trading. Wiley.

Hirshleifer, D., Teoh, S., 2003. Herd behaviour and cascading in capital markets; a review and synthesis. European Financial Management Vol. 9:1, 25–66.

Irwin, S. H., 2007. What do we know about the profitability of technical analysis? Journal of Economc Surveys Vol. 21, 786–826.

James, F., 1968. Monthly moving averages - an effective investment tool? Journal of Financial and Quantitiative Analysis Vol 3:3, 315–326.

Jensen, M., 1967. Random walks: reality or myth - comment. Financial Analysts Journal Vol. 23, 77–85.

Johnsson, R., 07 2010. A simple risk-return-ratio. URL `http://www.richardcbjohnsson.net/pdf/risk_return_ratio.pdf`

Khan, M., 2009. The 2008 oil price "bubble". policy brief 09/19. Tech. rep., Peterson Institute of International Economics, Washington DC.

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

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

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

Lo, A. W., Hasanhodzic, J., 2010. The Evolution of Technical Analysis. Wiley.

Lukac, L. P., Brorsen, B. W., 1990. A comprehensive test of futures market disequilibrium. The Financial Review Vol. 25:2, 593–622.

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

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

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

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

Masters, T., 2010. Monte-carlo evaluation of trading systems.
URL http://www.evidencebasedta.com/montedoc12.15.06.pdf

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

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

NASDAQ, July 2013. NASDAQ Commodity Index Methodology.

OriginalTurtles.org, 2003. The original turtle trading rules.
URL http://www.dailystocks.com/turtlerules.pdf

Pardo, R., 2008. The Evaluation and Optimization of Trading Strategies. John Wiley & Sons.

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

Penteado, M. A., 2013. Efficient markets, behavioral finance and a statistical evidence of the validity of technical analysis. Master's thesis, Cornell University Library.

Perl, J., 2008. DeMark Indicators. Wiley.

Phillips, P., Yu, J., 2010. Dating the timeline of financial bubbles during the subprime crisis. Master's thesis, Cowles FUndation Discussion Paper No. 1770, Yale University.

Reuters, August 2010. Futures Continuous Series, Methodology and Definitions. 0th Edition.

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

Ross, S., 2005. Neoclassical Finance. Princeton University Press, Princeton.

Samuelson, P., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review Vol. 6, 41–51.

Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance Vol. 52:2, 737–783.

Smidt, S., 1964. Amateur speculators. Master's thesis, Graduate School of Business and Public Administration, Cornell University.

Sornette, D., Cauwels, P., 2014. 1980-2008: The illusion of the perpetual money machine and what it bodes for the future. Risks Vol. 2, 103–131.

Sornette, D., Cauwels, P., 2015a. Financial bubbles: mechanisms and diagnostics. Review of Behavioral Economics Vol. 2 (3), 279–305.

Sornette, D., Cauwels, P., 2015b. Managing risks in a creepy world. Journal of Risk Management in Financial Institutions Vol. 8 (1), 83–108.

Sornette, D., Woodard, R., Zhou, W.-X., 2009. The 2006-2008 oil bubble: evidence of speculation, and prediction. Physica A Vol. 388, 1571–1576.

Stafford, P., 2015. Symphony is just the latest attack on bloomberg fortress. Tech. rep., Financial Times.

Tang, K., Xiong, W., 2010. Index investment and financialization of commodities. working paper no. 16385. Tech. rep., National Bureau of Economic Research (NBER).

UNCTAD, 2011. Price formation in financialized commodity markets: the role of information. Tech. rep., United Nations.

White, H., 2000. A reality check for data snooping. Econometrica Vol. 68:5, 1097–1126.

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

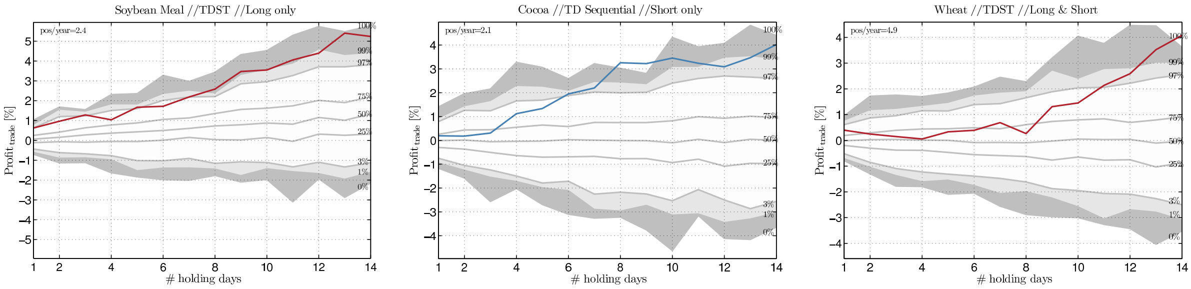
Zhou, G., 2010. How much stock return predictability can we expect from an asset pricing model? Economic Letters Vol. 108, 184–186.

Zhou, G., Zhu, Y., Qiang, S., 2012. Asset allocation: can technical analysis add value? Int. J. Portfolio Analysis and Management Vol. 1:1, 43–58.

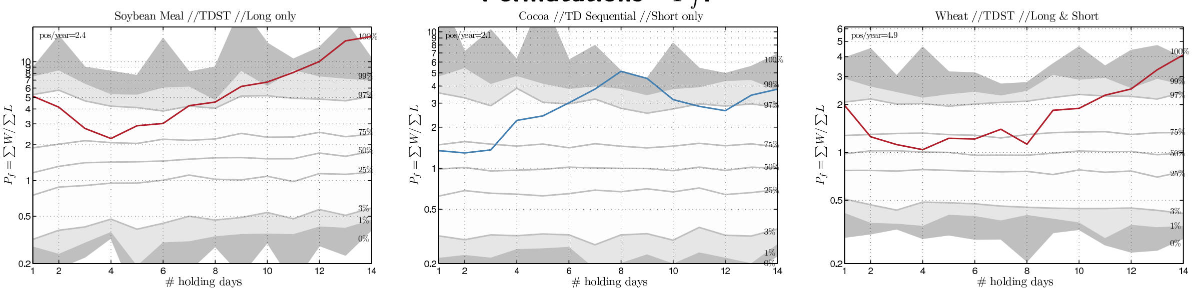
Figure 14: Permutation tests and conditional versus unconditional returns are being applied to three signal-market examples.

Long ST (Setup Trend) on Soybean Meal	Short Sequential on Cocoa	Long/Short ST on Wheat

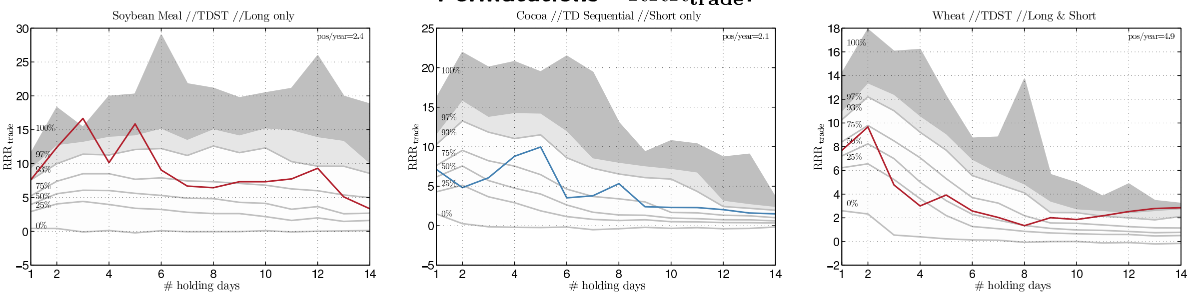
Permutations - $\mathbf{Profit_{trade}}$:



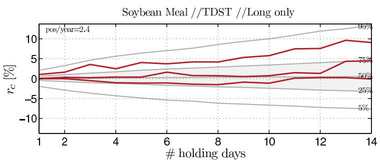
Permutations - P_f:



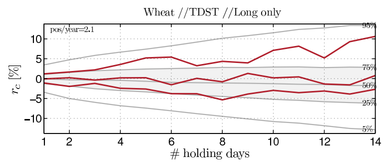
Permutations - $RRR_{\mathbf{trade}}$:



Conditional returns:



only short entry signals are being tested on Cocoa



only long entry signals are being tested on Soybean Meal