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

In this paper, 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 twenty-one commodity futures markets and ten years of daily data. Our work addresses price behaviour following new entry signals by studying whether, for short holding periods, the entry signals generated by these indicators can time the market moves and suggest the right side of the market (long or short). For example, we want to know how long we should hold or delay a trade before the price is expected to move significantly. The signals are sparse, as they mostly suggest entering the markets between one to five times per year. To adjust for the limited number of total days in which the trade signals are in-the-market, we generate the distributions of multiple performance metrics (mean return, profit factor and risk-return ratio) over different trade holding horizons, and compare them with their randomized versions, which have the same number of entry signals and the same number of holding days. The rolling strategy, which creates continuous futures data from separate contracts, plays a role in evaluating the statistical performance of these indicators. Overall, this paper gives more clarity to the predictive performance of these indicators as well as practical guidance on how to use them in a trading environment as generators of market entry ideas.

Keywords: technical analysis, backtesting, permutation test, financial markets, commodity futures, Contract rollover.
JEL: C12, G14, G17

1. Introduction

DeMark indicators are chart-based indicators that generate market entry signals. This family of indicators is, commercially speaking, quite popular. It is also possible to make use of them as an upgrade in leading financial market terminals such as Bloomberg Professional® and Thomson Reuters® which, combined, account for roughly 60% of the market share (Stafford, 2015). Despite this, no previous study has analysed the effectiveness of these indicators, although there are other studies available on simpler chart patterns (see, for example, Lo et al. (2000)). This paper gives more clarity to the predictive performance of these indicators by testing if they can time exceptional price moves and suggest the right side of a trade. Our aim is to understand what we should expected from the price development just after a completed signal. This should help to manage a trade right from the start, e.g., by delaying the mar-

[*]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

ket entry, or by holding to a temporary initial loss that is expected to pay back shortly.

The first step towards an observed price predictability is to compute the performance of each market entry signal for a sliding number of holding days. To complete the picture, the results have to be compared with the market performance. This is done via permutation tests: we create batches of random signals that carry the same features as the original signal and bypass several limitations of a simpler buy-and-hold comparison. For example, the number of market entries and holding days are the same as in the tested signal. If, statistically speaking, the performance of the original signal goes significantly beyond what can be expected from equivalent sets of random signals, then the indicator can time a sizeable price move that is yet to come within a limited number of holding days. The tests carry three performance measures, but the results are all quite similar. Therefore, the paper focuses on one of them, the (mean) returns per trade. The most suited indicator for a given market should jointly maximize, for example, the (mean) returns per trade and the percentage of holding days following entry signals that show statistically significant performance. In most cases, there is a limited range of holding days for which an indicator has predictive power. This might sound counterintuitive, but it means that significantly exceptional price moves are yet to come, as the entry signal suggests; however, they may come with a delay and only last for a few days.

DeMark indicators are a small corner of a vast topic called technical analysis (TA). They are tested on commodity futures contracts, the basic type of exchange traded financial instrument for commodities that is right at the intersection of physical and financial trading. Academic studies of TA on commodity futures markets are certainly not something new; some are quite old (Lukac et al., 1988; Lukac and Brorsen, 1990; Roberts, 2003), while others are more recent and rather optimistic about the technical trading content of commodity futures prices (Szakmary et al., 2010; Miffre and Rallis, 2007). Other modern approaches are more pessimistic (Marshall et al., 2008; Chinn and Coibion, 2014). We observe in our study that different conclusions can be due to different constructions of continuous price series. Once a specific contract is used to determine the price for a given trading day, there are multiple options regarding 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 particular, a roll on the expiry should be avoided because it may underestimate the predictive power of DeMark signals.

We continue with an introduction to TA and its use in commodity markets; this is followed by a contextualization of the backtesting of predictive market signals. 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. 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 TA. This contrasts with earlier, more skeptical perceptions of academic researchers based on various forms of the efficient market hypothesis (EMH, see (Fama, 1995)). One belief of TA is that, like physical objects, prices have inertia: when at rest, they often stay approximately at rest, and 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). 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. For example, prices tend to bounce on support and resistance levels (Garzarelli et al., 2014). However, the crossing of these levels is interpreted as prices moving in a way that is 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.

Multiple factors contribute to the deviation of commodity prices from fundamental values for periods long enough to disturb the normal decision-making processes (UNCTAD, 2011; Filimonov et al., 2014). This statement is supported also by recent empirical evidence on the existence of speculative bubbles in several commodity markets (Sornette et al., 2009; Gilbert, 2010; Phillips and Yu, 2010; Sornette and Cauwels, 2014, 2015a). In this context, the use of TA can be justified as a tool for price discovery. In fact, TA has always had a significant and consistent user cohort among commodity traders (Smidt (1964); Lukac et al. (1988); Billingsley and Chance (1996); Menkhoff and Taylor (2007)). As we have said, there are several factors that could push the price away from its fundamental value. For example, market participants cannot always enter positions when a price correction is yet to come (Gromb and Vayanos, 2010) because 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); 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, and based on judgemental biases (Ariely, 2010; Grinblatt and Han, 2005; Penteado, 2013); positive-feedback mechanisms (Sornette and Cauwels, 2015b); simple TA rules, which can also be easily understood by traders with no fundamental understanding (Gehrig and Menkhoff, 2006); and cross-asset strategies (Tang and Xiong, 2010).

The first known attempt at backtesting trading signals using historical data was undertaken by 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). His main idea was to use stars as a way to generate random patterns. Centuries later, the statistical problem of insignificant predictive evidence in TA became well known (James, 1968; Jensen, 1967). Major methodological innovations arrived later still, 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 using out-of-sample data until White (2000) developed an in-sample "reality" check. 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). There is an economic upper bound to TA's predictive power, which is based on the risk-return principle (Ross, 2005; Zhou, 2010).

The paper is organized as follows. Section 2 describes the DeMark indicators used to generate entry and trading signals. Section 3 goes into more detail regarding commodity futures and explains the problem of rolling commodity futures contracts. In addition, it suggests a method for creating 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. It collects many revised versions of traditional price-based indicators such as moving averages, trend lines, price ratios and Elliot waves, but it also includes other indicators, among which, *Sequential* is arguably the most renowned. When reading (DeMark, 1997; Perl, 2008) one may feel that everything is described but nothing is explained. This further motivates us to explore the performance of these indicators, although we are aware that the rationale behind some choices will stay unexplained (e.g. the values assigned to the parameters). The indicators *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

Algorithm 1 Entry signals

1: **for each** (new) t-th price bar:
2: **procedure** DeMark($\mathbf{P_c}$,$\mathbf{P_h}$,$\mathbf{P_l}$,n,m,q,p,k,**options**)
3:   update Setup's counter ▷ 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** $P_c(t) > \mathrm{res}(t)$ **then**
11:         open new long position at $t+1$
12:       **else if** $P_c(t) < \mathrm{sup}(t)$ **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 Countdown'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**

among slightly different versions. Our description focuses only on long entry signals; short entry signals can be derived by applying symmetrical conditions. See Appendix A for a summary of the parameters used in the backtests. 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.

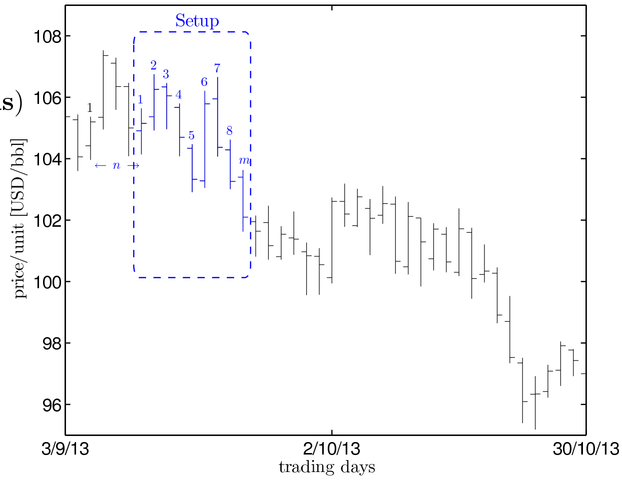


Figure 1: Setup on the light crude Oil futures contract, also known as the West Texas Intermediate (WTI) contract.

Each time $s = m$, the price bars related to the newly completed Setup are used in row 5 to determine the range of the Setup,

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

and the width of the range,

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

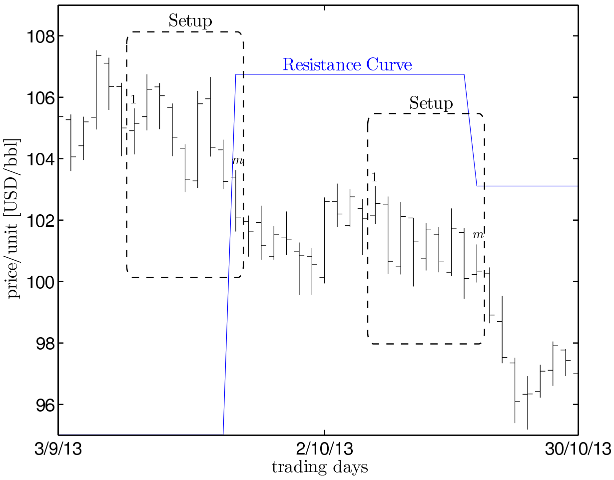


Figure 2: Each completed long *Setup* updates the resistance level of WTI.

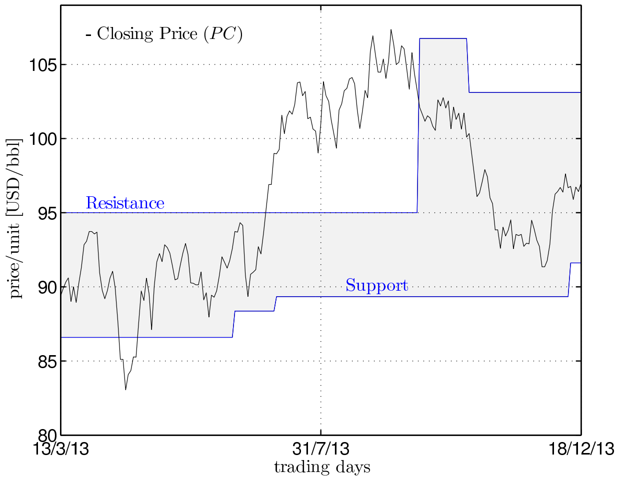


Figure 3: In June 2013, WTI's P_c breaks its resistance level ($P_c(t) > \mathrm{res}(t)$). Therefore, ST generates a new long entry signal.

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

$$\mathrm{res}\,(t) := \max_{s=1,..,m} (P_h)_s \;, \tag{4a}$$

$$\sup\,(t) := \min_{s=1,..,m} (P_l)_s \;, \tag{4b}$$

for each completed Setup. 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 ($P_c(t) > \mathrm{res}(t)$), as in Fig. 3.

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, 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 Countdown only if the following conditions are met (rows 27-28). Given t,

$$(\textit{standard})\;\; P_c\,(t) \le P_l\,(t-u)\,, \tag{5a}$$

$$(\textit{aggressive})\;\; P_l\,(t) \le P_l\,(t-u)\,, \tag{5b}$$

and $u = 2$, as suggested by DeMark. Only one of the two variants in eq. 5 should be used. Unfortunately, as already mentioned in the abstract, signals are sparse; we opt for the aggressive version in our backtests because it maximizes the number of entry signals. 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 in the Setup, what matters in the completion of the Countdown is the total number p of bars fulfilling Eq. 5, and not the consecutive number (see Fig. 4). Counter c increases based on conditions that are independent of

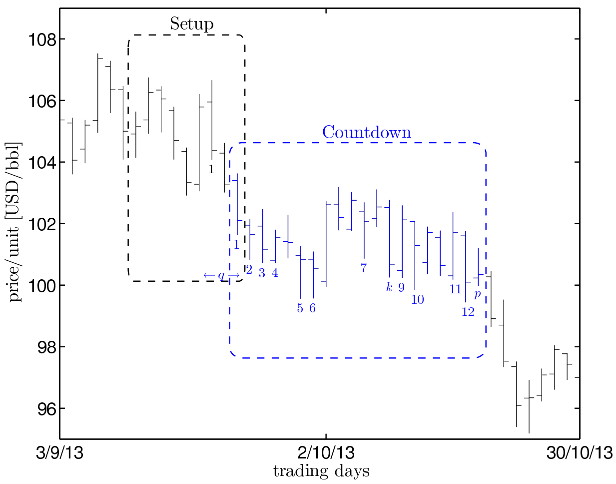


Figure 4: A completed *Countdown* on Light Crude Oil.

the Setup, although, during the Countdown (or its alternative, the Combo), there are additional checks based on the Setup that can restart ("recycle") the phase or, in a worst-case scenario, 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, to be 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 that which generates entry signals using ST, but, here, closing prices are replaced by high and low prices;
3. *New Setups.* If the new Setup has a P_c within the range of the old Setup, then the current Countdown is kept going. Otherwise, if there is a new completed Setup in the same direction as the old one, the Countdown is reset ($c = 0$) only if ${R_w}^{\text{new}} > {R_w}^{\text{old}}$.

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

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

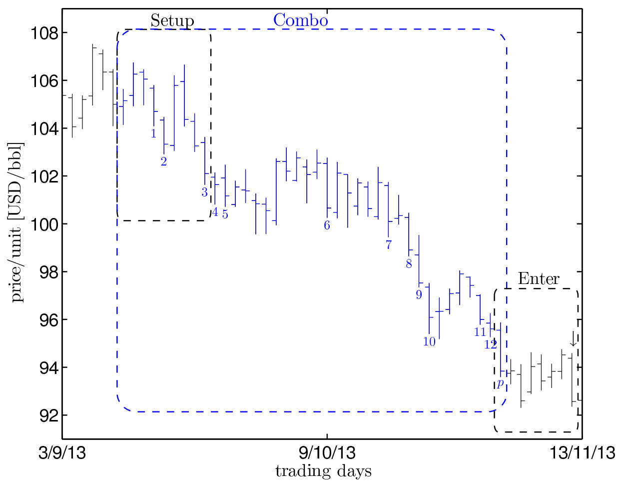


Figure 5: A completed Combo on WTI.

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 (see Fig. 5).

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 immediately on the next traded day $(t+1)$, at $P_c(t+1)$. However, 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 because it is the default configuration and it tries to optimize the timing by delaying an entry signal for a few price bars, if needed (Fig. 5). Anyway, the choice of the entry strategy does not seem to be the key factor for the performance of the indicator, because, at this point, an entry signal can only be delayed, not cancelled.

Sequential (in its traditional Countdown or its alternative Combo version) is a time-based indicator that tries to identify areas of trend exhaustion that will lead to price reversals. It is made of up to two sequential phases. The first is the Setup, which tries to capture price momentum; this is followed by the Countdown, which 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.

We now share some observations about Algorithm 1. To generate a long entry signal with Sequential, the Setup must 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 also be found in Chan and Lin (2004). A long Setup is completed when there are m consecutive closes, with 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 an 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. Based on this view and the parameters suggest by 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 necessarily interested in daily price moves as much as they are for price moves over a few trading days. Unsurprisingly, 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 five 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 of 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 move downwards 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 $n = 4$. However, 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, there is no exit signal for Sequential; therefore, we will only backtest our indicators on their entry signals.

Last, 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 to do so at a high and constant velocity. However, 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 the 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 data

3.1. Futures data

We test DeMark indicators over twenty-one commodity futures markets and ten years of data (January 2004 - January 2014). Some idiosyncratic properties of these markets can be found in Appendix B. In this paper, we limit our analysis to outright positions. The performance is computed on daily settlement prices, on which we calculate daily returns, while the entry signals are generated using daily price bars which contain intraday information such as daily opening, high and low prices. Complete daily price bars were not available for London Metal Exchange contracts. Some other contracts had to be excluded as well, for example, CBOT soybean oil, soybean meal and NYMEX Gasoline US could not cover the whole tested period. Table 1 lists all futures contracts chosen for the backtesting. There is not much flexibility in the choice of the data period, because we want as many years as possible to maximize the number of entry signals, which are sparse (on average, between two to five per year); however, it is not possible to include older data without reducing the number of commodities. Meanwhile, if a new price history builds up, we should keep it aside for out-of-sample testing. Anyway, the data period is still quite heterogeneous and includes uptrends and downtrends before and after the great recession in 2007-2008 which is our structural break. A detailed analysis of each commodity price would be informative for the reader, but it is not essential for performance testing because the methodology described later self-corrects the results based on the uptrends and downtrends in the markets.

Table 1: 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

This paper essentially deals with daily data. Commodity futures can be traded intraday, but this is as far as we can get if we want to generalize this study to most commodity markets, which are traded over-the-counter (OTC) and do not have sufficient liquidity (also compared to all the other asset classes) to be part of a high-frequency setting. For most products, it is difficult to enter and exit positions on an intraday basis, and even getting daily highs and lows from a broker might be very challenging. Front energy futures contracts traded on NYMEX or

ICE, like those in this study, have an open interest (OI) of a few hundreds thousands contracts, while the most liquid OTC energy products have an open interest between 1000 and 10 000 front-month contracts. Even if the focus is just on futures, the importance of high-frequency data in commodities would still be debatable because of the role played by physical hedgers. The overall impact of these players is clear only at the end of the day. In fact, physical markets are based on end-of-day assessments by pricing agencies such as Platts, Argus and Icis and when physical deals are pricing, hedging orders are sent to the exchange to trade at the daily settlement price. During the day the impact of hedgers generates noise: new exchange-for-physical deals (EFPs) or, perhaps, contractual triggers due to physical liftings, which have no direct relationship with the current flat price moves, force the hedgers to reposition immediately in market. In addition, a sudden increase in expected forward oil-production due to an expected increase in long-term demand (e.g. equivalent to ten Aframax cargoes or, equivalently, 10 000 crude futures contracts) could immediately generate strong selling pressure on crude futures due to the need to hedge by immediately shorting the futures.

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 only be traded 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 show abnormal volatility in the final weeks of their lives (Samuelson, 1965). Carchano and Pardo (2009) and Ma et al. (1992) suggest rolling the front contract one to two weeks before maturity or on the first day of the expiring month. As an alternative, they also recommend staying 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.

No strategy is permitted to roll back to the previous contract, i.e. from $M+1$ to M.

The 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.

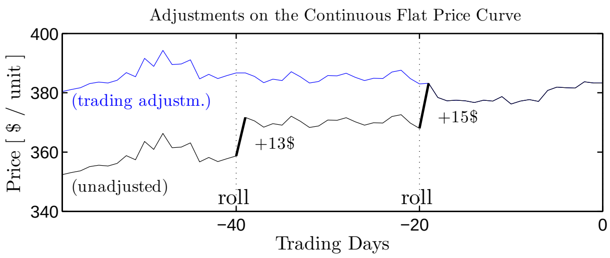


Figure 6: The black (unadjusted) curve is a generic continuous flat price curve, generated by one of the rolling strategies described in the text. Given the price level and the market's contango, it could also represent a very steep front contango market on ICE Gasoil after a flat price crash. Trading 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.

There 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 (see the black curve in Fig. 6). 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 that fit better with our backtests. The black unadjusted curve in Fig. 6 shows a steep front contango market; in fact, the front contract M is rolled twice to a higher valued M+1 contract. Assuming 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 and loss (P&L) 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 a 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. However, this backward-adjustment is more computationally intensive than a forward-adjustment, because a change to a single roll-over requires additional changes to all previous trading days.

Fig. 6 uses backward-adjustments and employs a discrete translation of the price by the adjustment Δ on the day of the roll such that, on day d:

$$\Delta := P_o(d) - P_c(d-1). \tag{9}$$

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:

$$\rho := {}^{P_o(d)}/_{P_c(d-1)}. \tag{10}$$

Using such ratios has the advantage that negative prices are not possible by construction. However, they may create large reconstructed price fluctuations. This unwanted feature has 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:

$$r_d := \frac{P_c(d) - P_c(d-1)}{P_c(d-1)} \; . \tag{11}$$

The blue curve in Fig. 6 cannot be used directly to compute daily returns; this is because settlement prices (P_c) can be negative or close to zero values, and this would distort daily returns. We propose using the blue curve to compute the numerator 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 to 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 toward it. Profitability requires a much broader discussion that we can only partially undertake 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 (OriginalTurtles.org, 2003; Covel, 2009). 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. From this perspective, the role of a predictive entry signal is to give an edge to the trader.

One of the simplest ways to test the predictive power of entry signals is to compare their total cumulated performance, e.g. the total cumulated returns or the net asset value (NAV), versus a buy-and-hold strategy over the whole data period. Nevertheless, this approach might not be the best choice even if we were willing to accept that buy-and-hold is a reference strategy in commodity trading. A general limitation of this setup is that each trade carries a different money allocation depending on the results of the previous trades (Siligardos, 2014) and this distorts the measurement of predictive power. In addition, the results would be very sensitive to the data period, especially because the signals that we are testing are sparse and mostly out of the market. We would be comparing different things, and further investigations would be needed. We would have to convince the reader about the choice of the data period, perhaps through a robustness analysis that describes the general statistical properties of each commodity and the related structural breaks. For example, we would have to comment on the uptrends and downtrends before and after the great recession in 2007-2008 and how these might affect the results. The analysis should be also strengthened by out-of-sample and in-sample data-snooping tests Kuang et al. (2014); Diebold (2015). To some extent, by comparing the distribution of (daily) returns conditional on entry signals to their unconditional counterpart, one could attempt to derive conclusions regarding the aforementioned comparison between trading strategies and the buy-and-hold (Lo et al., 2000). The conditional versus unconditional returns approach weights every entry signal equally, is easy to implement and offers graphical intuition, although it cannot combine longs and shorts on the same graph. For example, Fig. 7 suggests that long Sequential entry signals might overperform the cocoa futures market when the holding period reaches thirteen or fourteen days.

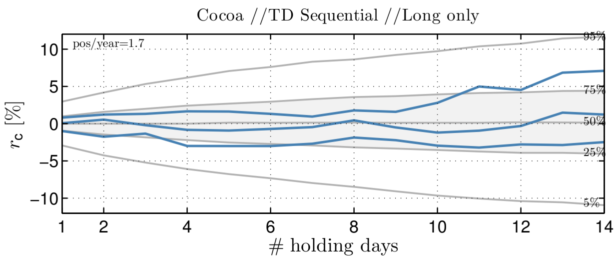


Figure 7: The conditional distribution of returns on cocoa for long Sequential entry signals represented as 25%, 50% and 75% quantiles (blue curves) compared with the market's unconditional returns distribution (in grey, representing buy-and-hold). This is based on fourteen holding days: the average increases from 0.0% to +2.7%, standard deviation goes slightly up from 7.4% to 7.7%, kurtosis is lowered from 4.2 to 3.8 and skewness is pushed in positive territories from 0.0 to +1.1.

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 fourteen. A limit of this test is that it only compares 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 (idd) 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, which, in this case, suggests that the long Sequential entry signals do not over-perform the cocoa futures market, may just be due to its lack of power.

Unsurprisingly, the Monte-Carlo permutation tests (Aronson, 2007; Masters, 2010) in Fig. 8-10 provide a conclusion that is more in line with the previous graphical intuition. We will be explain this later. For now, we will stick to this methodology, because, by design, it can handle sparse entry signals that are also held for short holding periods. In addition, the results are not biased by the trends and the structural breaks of the underlying commodities, and this makes the choice of the data period less critical. Finally, this testing keeps the returns in the same order as they appeared historically.

According to the null hypothesis H_0 the long, short and neutral positions of the signal 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 an equal chance of appearing in real life and in the randomization process. For example, the total number of trades, longs, shorts 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.

Compared with the bootstrap method, permutation tests have more requirements to fulfill. However, 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. Further, bootstrap tests re-sample returns with replacement while the trading signal is kept unchanged. This is how the null hypothesis distribution is generated. However, 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, entered trades and holding days for each trade. Let us assume that there are a total of thirty 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 three holding days, then there are only twenty-eight possible permutations. If, instead, there are five 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 the true and the observed quantile, respectively. 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%), the true quantile q has the confidence interval $\hat{q} - \varepsilon \leq q \leq \hat{q} + \varepsilon$, with

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

where $n_0 = 400$, 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, using the "exact" Clopper-Pearson (1934) confidence interval is generally recommended; this 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. 2.

Table 2: Quantile transformations in the Monte-Carlo permutation test. The null hypothesis distributions are approximated; therefore, observed quantiles need to be conservatively adjusted according to confidence intervals. q and $\hat{q}$ are the true and the observed quantiles, respectively.

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

Each trading signal uses an aggregate measure to determine its performance. In our case, this measure can be a mean value over single positions, e.g., a mean return over all the single trade returns, which we will denote $\text{Profit}_{\text{trade}}$. A further step in the analysis could be to substitute returns with a return-to-risk ratio, e.g., the Risk-Return-Ratio (RRR) (Johnsson, 2010), which is defined for a single trade as

$$RRR := \frac{r}{DD_{max}}, \quad (13)$$

with r and DD_{max} being the (conditional) return and the maximum drawdown of the trade, respectively. 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 (Chekhlov et al., 2005; Eling and Schuhmacher, 2007). The overall RRR value will be computed as the mean value of the RRR values of each trade and we will refer to it as RRR_{trade}. The set of performance metrics is completed by a measure that can be computed only over multiple positions, such as the profit factor. This is a profit-to-loss ratio which is defined for a basket of trades as

$$P_f := \frac{\sum \text{Profits}}{\sum \text{Losses}}. \quad (14)$$

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:

$$P_f = \frac{\sum r^+}{\sum r^-} = \frac{N^+ \overline{r}^+}{N^- \overline{r}^-}, \quad (15)$$

where N^+ and $\overline{r}^+$ are the number of winning trades and their mean return, respectively. 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 following definitions of payoff ratio and win ratio,

$$P_r := \frac{\overline{r}^+}{\overline{r}^-} \quad \text{and} \quad W := \frac{N^+}{N^+ + N^-}, \quad (16)$$

the profit factor can be then written to obtain

$$P_f = \frac{W}{1-W} P_r. \quad (17)$$

A trading signal generates profits when $P_f > 1$ and, by definition, the breakeven is reached when $P_f = 1$. The desired condition is often $P_f > 2$ (Harris, 2009). Eq. 17 highlights how signals with low P_r must have a high W to generate profits, as is the case for intraday 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 breakeven ($P_f = 1$) and the random distribution is performing significantly better, 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 a symmetrical interpretation to this measure when returns per trade are, for example, all negative.

The results for long Sequential tests on cocoa are shown in Fig. 8-10. In the current backtesting, a p-value of 10% translates into a possibly significant rejection of H_0; a value of 5% translates into a statistically significant rejection. The lightly shaded area represents possible statistical significance at the 90% level, and the darkly shaded area represents statistical significance at the 95% level. Both use conservatively adjusted quantiles by taking into consideration confidence intervals. Turning our attention to 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%, $\hat{q_1}$=3% and $\hat{q_2}$=97%. This means that the candidate indicator (in Fig. 8-10, and in Tab. 3) is never statistically significant, because the use of conservative quantile transformations on all performance measures is always decisive to downgrading possibly significant results to nonsignificant ones when each position is being held for thirteen or fourteen days.

Table 3: Measured quantiles for long Sequential on cocoa.

	$\textbf{Profit}_{\textbf{trade}}$	P_f	$RRR_{\textbf{trade}}$
13 holding days:	95%	95%	87%
14 holding days:	95%	92%	90%

4.2. Impact of the rolling method on performance results

Fig. 11 examines the impact of different rolling strategies for long ST on soybean meal, short Sequential on cocoa and long and short Sequential on natural gas ICE. When the roll is done on the expiry day, the number of statistically significant holding days decreases compared with rolling 3, and in two examples there is no statistical significance left. In

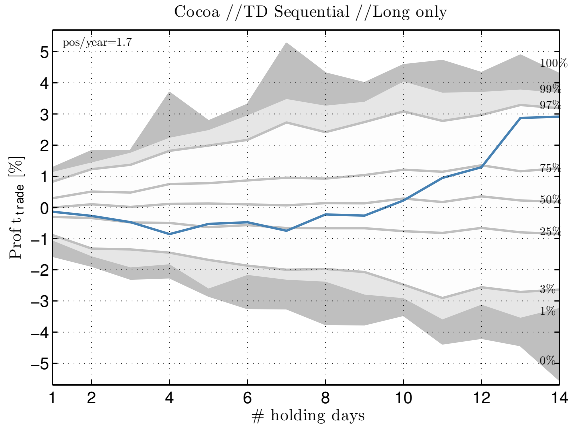


Figure 8: $\text{Profit}_{\text{trade}}$ permutation tests for long Sequential on cocoa.

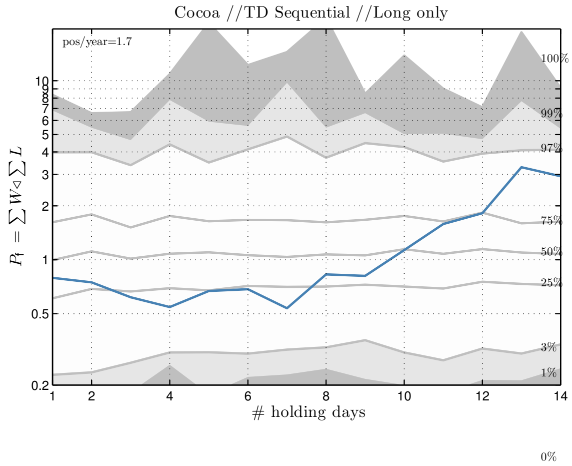


Figure 9: P_f permutation tests for long Sequential on cocoa.

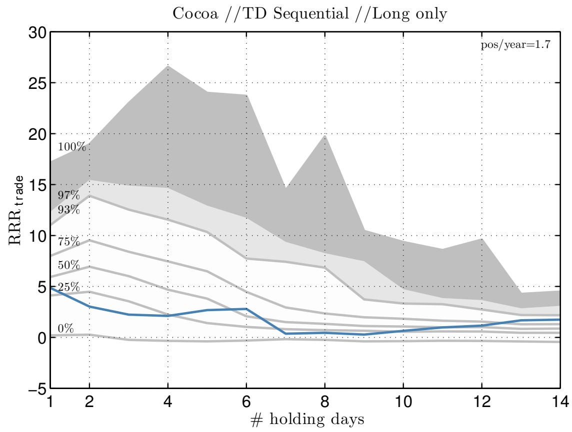
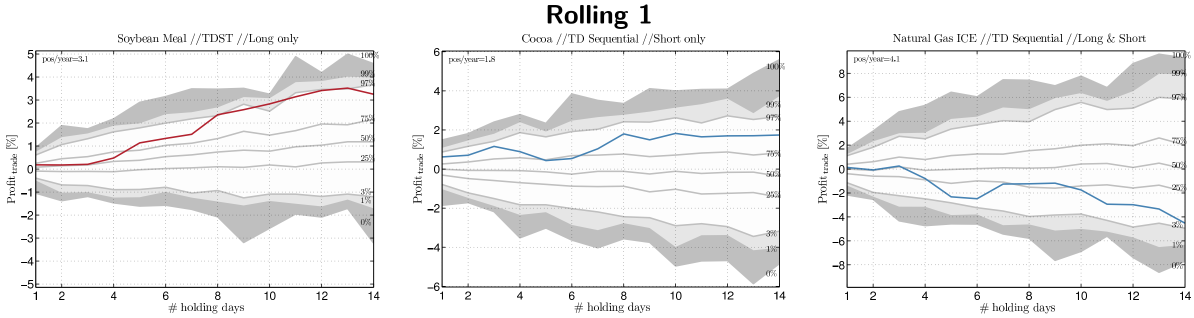


Figure 10: RRR_{trade} permutation tests for long Sequential on cocoa.

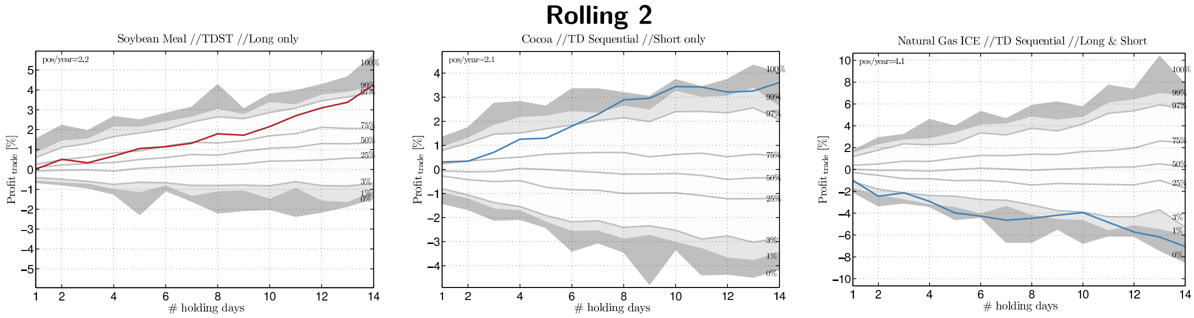
Figure 11: Impact of rolling strategies on the indicator's $\text{Profit}_{\text{trade}}$ performance and statistical significance.

Long ST (Setup Trend) on Soybean Meal	Short Sequential on Cocoa	Short Sequential on Nat. Gas

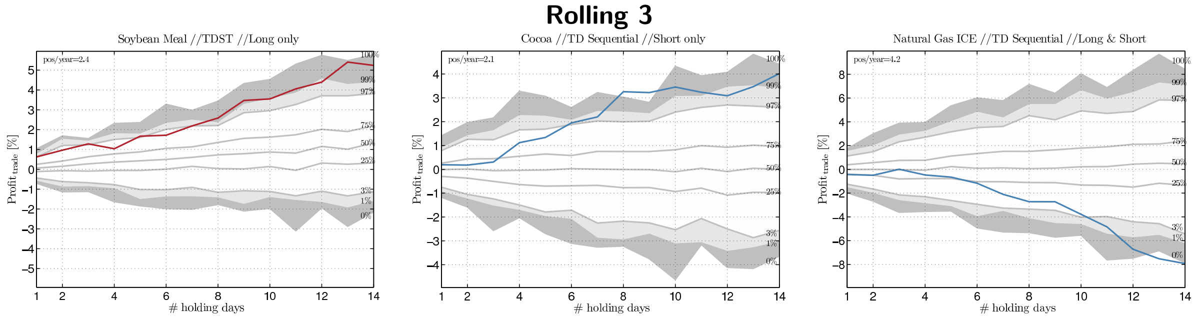
Rolling 1



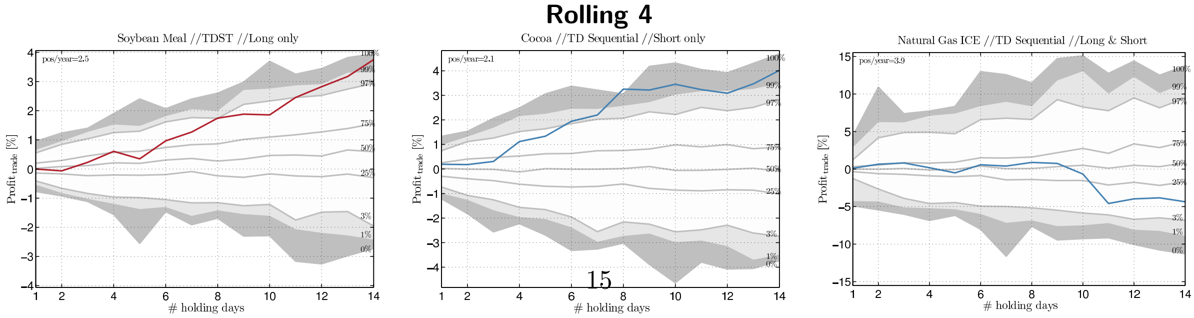
Rolling 2



Rolling 3



Rolling 4



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 may underestimate the predictive power of an indicator. Further, rolling 1 always has the lowest $\text{Profit}_{\text{trade}}$ values (in magnitude) among the four rolling strategies. Rollings 2 and 3 have a similar number of trades and 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 rolling 4. The reason for this is that the front contract is the most liquid until twenty to thirty days before its expiry; therefore, the two strategies roll very close in time. In contrast, rollings 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 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. The number of holding days is swept, which allows us to study trade performances during the days that follow the entry signals, as we have seen in Fig. 8-10.

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 (14); and (iii) Risk-Return-Ratio RRR_{trade}. If such a performance metric is measured to go significantly beyond what can be expected from randomness by having a value within the shaded area of previous figures, then the indicator is predictive. Ideally, measured performance should be within the darkly 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. They can also be closed whenever the trader feels more comfortable.

However, in most cases, there is a limited range of holding days for which the indicator has predictive power. This may sound counterintuitive, but it means that a significantly exceptional price move is yet to come, as the entry signal suggests; however, it may come with a delay and only last for a few days. If decision makers are aware of these signal-market behaviours, they can have a better idea of how to manage the initial phase of a trade, for example, by delaying the market entry or by holding to a temporary loss which is expected to pay back shortly.

By design, the 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, for instance, Fig. 8-10, 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. The most-suited indicator for a given market should jointly maximize σ (this is the priority) and the performance measure, which we also call 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. This seems a simple but effective way of limiting the data-mining effect that would occur if the maximum $\text{Profit}_{\text{trade}}$ value within the fourteen holding days was chosen to represent the profit potential of an indicator.

This two-dimensional evaluation framework with σ on the horizontal axis (i.e. percentage of significant holding days) and profit potential on the vertical axis (i.e average $\text{Profit}_{\text{trade}}$) has been applied to the DeMark indicators over the twenty-one commodity futures markets listed in Tab. 1. The indicators are evaluated only for their long entry signals in Fig. 12(a), only for short entry signals in Fig. 12(b) and for both longs and shorts in Fig. 12(c). The most easily manageable signal-market combinations are represented in the shaded areas of fig. 12.

DeMark indicators have sparse entry signals. Across all commodities, each indicator enters the markets with the frequencies shown in Tab. 4.

Table 4: Across all commodities, each indicator enters the markets with the following frequencies (i.e., the average number of positions per year).

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 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 than ST. As explained in Section 2, Sequential and Combo always need 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 the fact that these two indicators need to fulfil additional conditions during the Countdown or Combo phase, then it (ideally) takes 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 a forthcoming price move in an unknown direction can still be profitable, e.g. by using options-based strategies. Looking at fig. 12(a), a long Sequential on energy products is more of a trend detector that a turning point detector, and the market is more likely to continue its negative trend than 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. The good news is that signal-market combinations, which in Fig. 12 appear in a stable but unprofitable area, can still be informative if interpreted correctly.

So far, Fig. 12 has provided a high-level overview of the behaviour of the indicators on several markets.

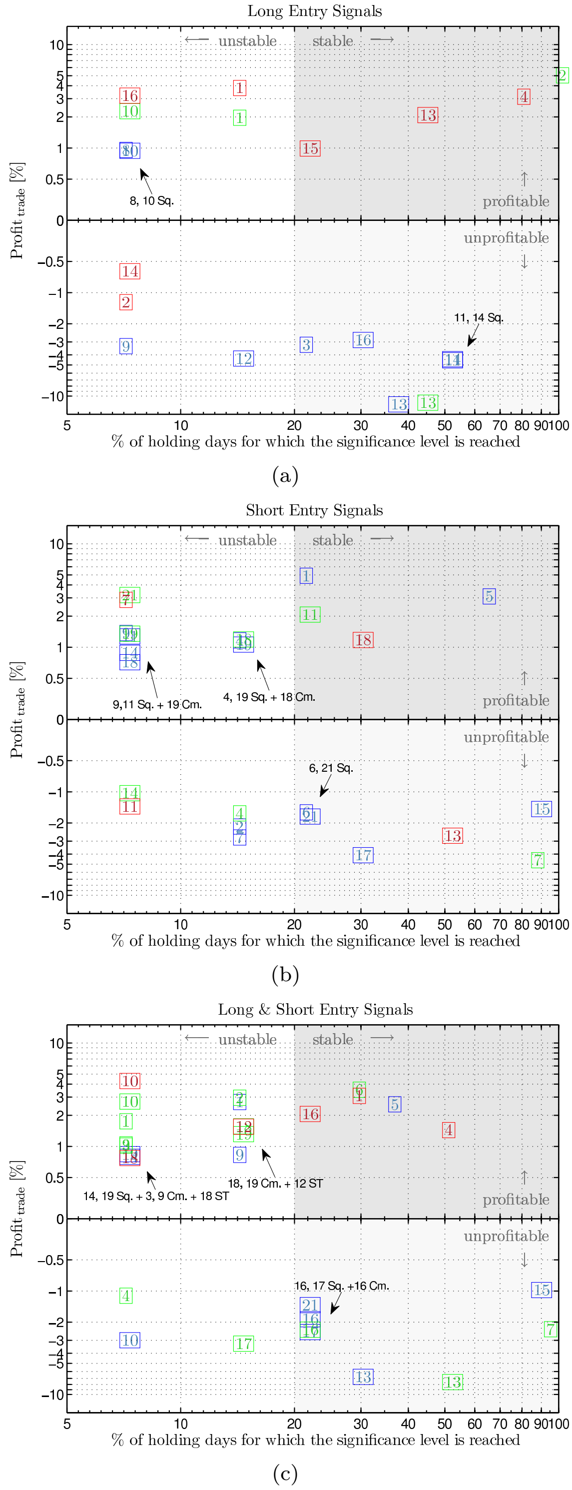


Figure 12: Evaluation framework that shows for each indicator its σ on the horizontal axis (% of significant holding days) and the corresponding profit potential on the vertical axis (average $\mathrm{Profit}_{\mathrm{trade}}$). The commodity market is specified by a number which refers to Tab. 1. Blue colour represents Sequential (Sq.), green is for Combo (Cm.) and red is for ST. Combinations with no predictive power are not shown here.

The framework can be studied in more detail by comparing results from different performance measures within the permutation test setup, as in Fig. 13. Conditional distributions of returns versus unconditional returns can be used as an additional countercheck.

For example, Fig. 12(a) highlights a good performance for long ST entry signals on soybean meal. A more in-depth analysis based on Fig. 13 shows that $\text{Profit}_{\text{trade}}$ grows regularly over the holding period; at the same time, it stays steadily in the shaded area, which represents a possible or statistically significant result. Further, 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 positions. If we also recall the interesting results for long positions previously described in Section 4, then Sequential is also a predictive indicator when both long and short positions are considered; this is confirmed by Fig. 12(c). The last example in Fig. 13 exhibits statistical significance for long and short ST entry signals on wheat and shows how results may depend on the holding period. There is good predictive power across all metrics, but this only appears eleven to twelve days after the position has been entered.

5. Conclusion

Overall, this paper has taken a step away from the conventional backtesting approaches. The most common way of studying the weak form of market efficiency is to test a wide range of technical indicators over many markets by using simple buy-and-hold-based metrics, and to legitimize the best results by using standard data-snooping tests, such as the Diebold-Mariano test. In our case, we tailor a general interdisciplinary tool (a permutation test) to the needs of a few specific indicators that generate sparse entry signals. We also analyse the impact of different commodity futures' rolling strategies and performance metrics, with a focus on testing what happens immediately after a new entry signal is generated.

The specific indicators we study are DeMark indicators. They are intriguing because their layout has been described in multiple books (DeMark, 1997; Perl, 2008) and they are often used by practitioners to time the market, but the rationale for these indicators is not entirely clear, and the choice of their parameters is not really motivated and/or supported by data.

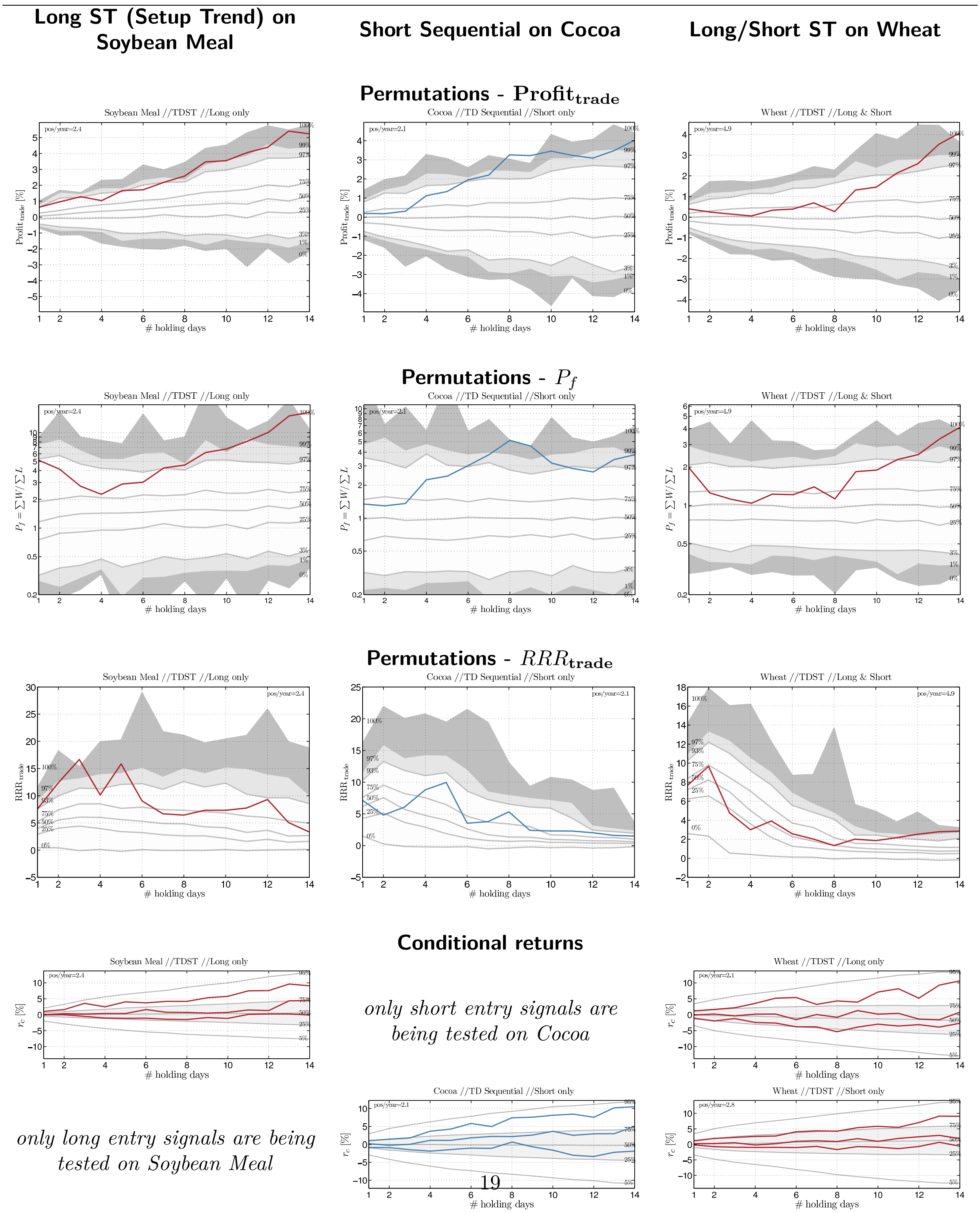
This paper accepts the DeMark indicators as they are and focuses on providing a better understanding of how they might perform as generators of market entry signals. More specifically, we set out to test if there is some predictive power in them: whether they can time exceptional price moves and suggest the right side of a trade. As decision makers, we would like advice on how to manage the initial phase of a trade, e.g., by delaying the market entry, or by holding to a temporary initial loss that is expected to pay back shortly. Profitability measurement is out of scope, since it also depends also on exposure management techniques that lock-in profits and cut losses.

The first step is to describe the signal generation process of these indicators. In this paper we study three specific indicators: *Sequential*, *Combo* and *Setup Trend* (ST). DeMark indicators (especially Sequential) are not so easy to replicate because they involve multiple conditions that overlap in time. Therefore, we introduce a pseudocode to facilitate the understanding and reproducibility of these signals.

The testing is based on daily prices. Although commodity futures can be traded intraday, daily data is the most natural way to start looking at commodities. In fact, daily futures settlement prices are the best reflection of the effects of all market players, including the hedgers, who typically send orders to the exchange to trade at the daily settlement price. Further, most commodity markets are traded OTC, so they do not have sufficient liquidity (compared with all other asset classes) to be part of a high-frequency setting.

Our first conclusion is that these entry signals are sparse. New signals materialize mostly between one to five times per year. This observation affects the way in which the backtesting is designed. The data

Figure 13: Permutation tests and conditional versus unconditional returns are being applied to three signal-market examples.



period has to be as long as possible so that we are able to compute some statistics on the entry signals. The starting date is January 2004 (giving us ten years of data); we prefer not to go further back because DeMark signals need also daily highs and lows, and having older price history would mean that we would have to exclude other futures contracts from our study. Nevertheless, the data period is still quite heterogeneous and includes uptrends and downtrends before and after the great recession in 2007-8, which is our structural break.

A performance comparison with a buy-and-hold strategy over the whole data period is one of the simplest performance tests. The comparison can be done by calculating a metric like the Net Asset Value, or by looking at the distribution of (daily) returns conditional on entry signals versus their unconditional counterpart to test if average returns are increasing, if standard deviation or kurtosis are decreasing and if skewness is being pushing into a positive territory. If we compared strategies that are mostly out-of-the-market (few entry signals that are held for short holding periods) with a buy-and-hold that is always in-the-market, the analysis would fall short of convincing the reader about the performance of the indicators; data-snooping tests would become necessary. The comparisons would depend on the trends and the structural breaks of the underlying data periods and would be sensitive to the performance of a limited number of entry signals. On top of that, a strategy such as buy-and-hold, which is in the market for more days, would also carry more risk. We opt for permutation tests, which bypass several problems, including the ones just mentioned, and are suited to study what happens immediately after the completion of these sparse entry signals over different trade holding horizons. In this setup, all possible permutations, including the candidate signals, must have both equal chances of appearing in real life and in the randomization process. Initially, the tests are based on three performance metrics. Mean return and profit factor always provide very similar results, while a risk-adjusted metric, such as the risk-return ratio, may lead to slightly different conclusions, but with no meaningful differences. For this reason, a high level investigation is only shown for the mean return in Fig. 12.

Multiple rolling strategies have been investigated; however, only one is used to test DeMark signals, and this rolls from the front contract to the next when its open interest is lower than that of the nearby contract. Other rolling strategies that slightly anticipate the expiry may be used as well, with similar results (see Fig. 11). In particular, our paper suggests avoiding a roll on the expiry because it might underestimate the predictive power of these signals.

Our results show that, in most cases, there is a limited range of holding days for which the indicators have predictive power. This may sound counterintuitive, but it means that a significantly exceptional price move is yet to come, as the entry signal suggests; however, it may come with a delay and only last for a few days before becoming a nonexceptional move again. Armed with this information, a decision maker can better decide when and how to enter the markets. Sequential is an interesting case, because there are statistically significant price moves following either long or short entry signals on all twenty-one commodities save one. Although this signal is designed to time trend reversals, in several cases the exceptional price move maintains the direction of the trend. This happens for expected trend reversals on energy futures: negative trends would rather continue after the signal instead of reverting back. These results contradict the design of the indicator and make it difficult to grasp the economic rationale behind it, which is assumed to be obvious. As a consequence, it becomes challenging to use Sequential, as it is designed, to also offer some insights into the data-generating process for commodity futures prices and their times series properties. Nevertheless, we could still try to do so by exploring different values for the parameters, perhaps by sweeping Sequential parameters p, q and k; although we would really have to pay attention to the data snooping bias or selection bias. The higher the number of indicators being tested on the same historical data-set, or the higher the number of data-sets tested on the same indicator, the higher the probability that luck will have an impact on the most statistically interesting results. In our tests, the results do not seem to be driven by luck due to a relatively low number of combinations (three trading rules over twenty-one data sets). In order to approach the problem gradually, without creating de facto new versions of the same indicator, we could first leave the parameters untouched and use the methodology in this paper to look at what happens before the completion of a new signal, to test if the signals are maybe arriving too late. The same methodology can be extend to other entry and exposure management signals.

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. doi:10.2307/2685469.

Andersen, J. V., Gluzman, S., Sornette, D., 2000. Fundamental framework for technical analysis. European Physical Journal Vol . B 14, 579–601. doi:10.1007/s100510051067.

Ariely, D., 2010. Pedictably irrational. New York and London: Harper Perennial. doi:10.1002/bdm.673.

Aronson, D. R., 2007. Evidence-Based Techical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals. Wiley. doi:10.1002/9781118268315.

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. doi:10.2307/3867650.

Billingsley, R., Chance, D., 1996. Benefits and limitations of diversification among commodity trading advisors. Journal of Portfolio Management Vol. 23, 65–80. doi:10.3905/jpm.1996.409581.

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. doi:10.1111/j.1540–6261.1992.tb04681.x.

Carchano, O., Pardo, A., 2009. Rolling over stock index futures contracts. The Journal of Futures Markets Vol. 29:7, 684–694. doi: 10.1002/fut.20373.

Chan, E. P., 2013. Algorithmic trading: winning strategies and their rationale. John Wiley & Sons. doi:10.1002/9781118676998.

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

Chekhlov, A., Uryasev, S., Zabarankin, M., 2005. Drawdown measure in portfolio optimization. International Journal of Theoretical and Applied Finance Vol. 8:1, 13–58. doi:10.1142/S0219024905002767.

Chinn, M. D., Coibion, O., 2014. The predictive content of commodity futures. Journal of Futures Markets Vol. 34:7, 607–636. doi:10.1002/fut.21615.

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. doi:10.3905/JPM.2009.36.1.12.

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

Devenow, A., Welch, I., 1996. Rational herding in financial economics. European Economic Review Vol. 40:3-5, 603–615. doi:10.1016/0014–2921(95)00073–9.

Diebold, F. X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. Journal of Business & Economic Statistics Vol. 33:1, doi:10.1080/07350015.2014.983236.

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

Eling, M., Schuhmacher, F., 2007. Does the choice of performance measure influence the evaluation of hedge funds? Journal of Banking & Finance Vol. 31, 2632–2647. doi:10.1016/j.jbankfin.2006.09.015.

Fama, E., 1995. Random walks in stock market prices. Financial Analysts Journal Vol. 51:1, 75–80. doi:10.2469/faj.v51.n1.1861.

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. doi:10.1016/j.jimonfin.2013.08.010.

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. doi:10.1038/srep04487.

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. doi:10.1002/ijfe.301.

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. doi:10.1016/j.jfineco.2004.10.006.

Gromb, D., Vayanos, D., 2010. Limits of arbitrage. Annual Review of Financial Economics Vol. 2:1, 251–275. doi:10.1146/annurev–financial–073009–104107.

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. doi:10.1002/fut.1905.

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. doi:10.1111/1468–036X.00207.

Irwin, S. H., 2007. What do we know about the profitability of technical analysis? Journal of Economc Surveys Vol. 21, 786–826. doi:10.1111/j.1467–6419.2007.00519.x.

James, F., 1968. Monthly moving averages - an effective investment tool? Journal of Financial and Quantitiative Analysis Vol 3:3, 315–326. doi:10.2307/2329816.

Jensen, M., 1967. Random walks: reality or myth - comment. Financial Analysts Journal Vol. 23, 77–85. doi:10.2469/faj.v23.n6.69.

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. doi:10.1016/j.ijforecast.2013.07.015.

Lo, A. W., Harry, M., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference and empirical implementation. The Journal of Finance Vol. 55:4, 1705–1765. doi:10.1111/0022–1082.00265.

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. doi:10.1111/j.1540–6288.1990.tb01300.x.

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. doi:10.1080/00036848800000113.

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. doi:10.1002/fut.3990120208.

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

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. doi: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. doi:10.1257/jel.45.4.936.

Miffre, J., Rallis, G., 2007. Momentum strategies in commodity futures markets. Journal of Banking & Finance Vol. 31:6, 1863–1886. doi:10.1016/j.jbankfin.2006.12.005.

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. doi:10.1002/9781119196969.

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. doi:10.1111/j.1540–6261.1997.tb03807.x.

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

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. doi:10.3390/risks2020103.

Sornette, D., Cauwels, P., 2015a. Financial bubbles: mechanisms and diagnostics. Review of Behavioral Economics Vol. 2 (3), 279–305. doi:10.1561/105.00000035.

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. doi:10.1016/j.physa.2009.01.011.

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

Szakmary, A. C., Shen, Q., Sharma, S. C., 2010. Trendfollowing trading strategies in commodity futures: A reexamination. Journal of Banking & Finance Vol. 34:2, 409–426. doi:10.1016/j.jbankfin.2009.08.004.

Tang, K., Xiong, W., 2010. Index investment and financialization of commodities. working paper no. 16385. Tech. rep., National Bureau of Economic Research (NBER). doi:10.3386/w16385.

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. doi:10.1111/1468–0262.00152.

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. doi:10.1016/j.econlet.2010.05.008.

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. doi:10.1504/IJPAM.2012.046908.

Appendix A Parameters used during the backtests

Table 5: 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. 1	
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)	

Appendix B Idiosyncratic properties of commodity futures markets

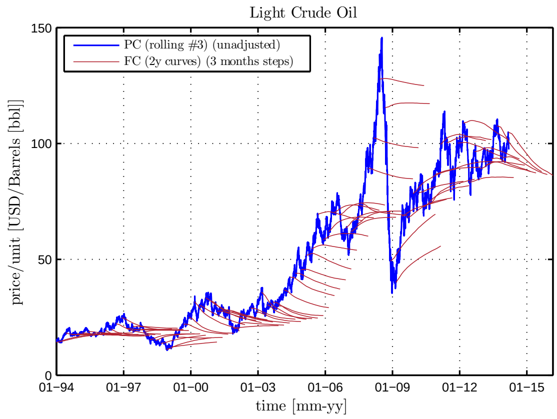
A commodity market is a market that trades raw materials. The market is physical when the product is delivered to the buyer, otherwise it is purely financial, for hedging, exchange for physical (EFP), investment or speculative purposes. Physical traders place themselves in between producers and consumers. 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 being used since the 18th century (Hamori et al., 2001) 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.

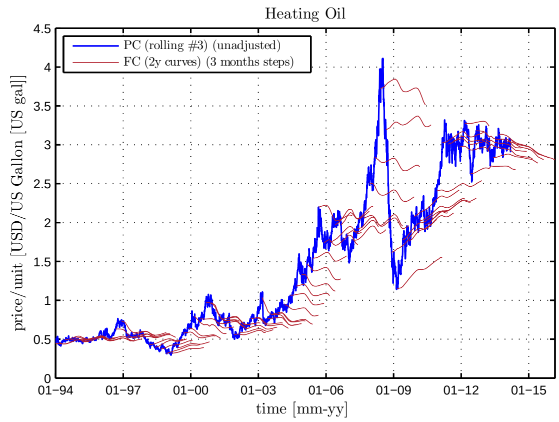
Futures contracts on commodities constitute the basic type of exchange traded financial instrument. Currently, the most popular contracts cover soft agriculturals, grains, energy products, industrial and precious metals. A commodity futures contract is an agreement to buy or sell a standardized quantity 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 6: 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	$\cdots$



(a)



(b)

Figure 14: 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.

Table 6 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.

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 Tab. 6, for example on the front contracts to maximize volatility. This trading style is the most common for technical traders. Blue curves in Fig. 14 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* [2]. 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.

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.

[2] this definition is the one used by traders. It is worth noting that an academic definition of (normal) backwardation entails a comparison of the expected spot price at maturity with the forward price at the same maturity. The two definitions are not equivalent.