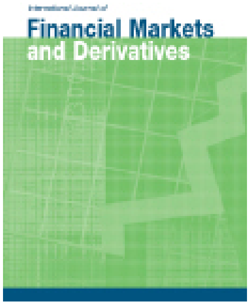


International Journal of Financial Markets and Derivatives

ISSN online: 1756-7149 - ISSN print: 1756-7130
https://www.inderscience.com/ijfmd

The Commitment of Traders report as a trading signal? Short-term price reversals and market efficiency in the US-futures market

Simon Dreesmann, Tim Alexander Herberger, Michel Charifzadeh

DOI: 10.1504/IJFMD.2023.10053051

Article History:

Received:	21 April 2022
Accepted:	20 November 2022
Published online:	17 February 2023

The Commitment of Traders report as a trading signal? Short-term price reversals and market efficiency in the US-futures market

Simon Dreesmann*

IU International University,
Erfurt, Germany
and
Chair of Business Administration,
especially Entrepreneurship, Finance and Digitalization
Andrássy University,
Budapest, Hungary
Email: simon.dreesmann@gmail.com
*Corresponding author

Tim Alexander Herberger

Chair of Business Administration,
especially Entrepreneurship, Finance and Digitalization
Andrássy University,
Budapest, Hungary
Email: tim.herberger@andrassyuni.hu

Michel Charifzadeh

ESB Business School,
Reutlingen University,
Reutlingen, Germany
Email: michel.charifzadeh@reutlingen-university.de

Abstract: The Commitment of Traders report (CoT) has been around for over 30 years, consistently revealing the futures positions of key market players. This study's primary aim is to use the comprehensive data from the Commitment of Traders reports to develop a short-term reversal trading strategy. Against the benchmark, a S&P 500 buy-and-hold approach with a Sharpe ratio of 1.07, the CoT long only strategy generated significant results in six individual markets. Extending the strategy to long-and-short, two markets outperformed the benchmark significantly. However, a scenario analysis indicated underperformance of the CoT strategy when traded in a portfolio, confirming that the chosen strategy parameters could not generate excess Sharpe ratios. Our results indicate that the Commodity Futures Trading Commission, more specifically the CoT report, contributed to efficient derivatives market.

Keywords: futures short-term reversal trading strategy; Commitment of Traders report; portfolio optimisation; Monte Carlo simulation; efficient derivatives market.

JEL codes: G13; G15; G17.

Reference to this paper should be made as follows: Dreesmann, S., Herberger, T.A. and Charifzadeh, M. (2023) 'The Commitment of Traders report as a trading signal? Short-term price reversals and market efficiency in the US-futures market', *Int. J. Financial Markets and Derivatives*, Vol. 9, Nos. 1/2, pp.76–113.

Biographical notes: Simon Dreesmann is Adjunct Researcher at the Chair of Business Administration, especially Entrepreneurship, Finance and Digitalization at Andrássy University, Budapest. Professionally, he has been working as a trader in derivatives and spot markets since 2018.

Tim Alexander Herberger is an Associate Professor and holds the Chair of Business Administration, specifically Entrepreneurship, Finance and Digitalization at Andrássy University Budapest, Hungary. Furthermore, he is a Research Fellow at Salzburg University of Applied Sciences in Austria, University of Bamberg and University of Erlangen-Nuremberg in Germany as well as lecturer at IU International University, Erfurt, in Germany. His major fields of research are behavioural and empirical finance as well as innovations and sustainability efforts in the financial sector.

Michel Charifzadeh is a Professor of Accounting and teaches at ESB Business School, Reutlingen University, Reutlingen, Germany. He has several years of professional experience as a project manager in corporate finance. Professor Charifzadeh has published articles, monographs, and textbooks on topics related to finance, management accounting, financial accounting, and sustainability reporting.

1 Introduction

While full transparency of traders' open interest may be difficult to find for most markets, this situation is a reality in the US futures market as the Commodity Futures Trading Commission (CFTC) is devoted precisely to this mission (CFTC, 2021a). To achieve market transparency, the CFTC regularly releases the Commitment of Traders (CoT) report since 1986. This report categorises the participants of futures markets and presents this data by measuring the short and long positions currently held in the market. It provides vital information about commercial traders, speculators, and retail traders in US futures markets to the public. Thus, examining CoT report data could help to explain moves and reversals in US futures markets. This unparalleled transparency can be used to analyse the efficiency of trading strategies and, beyond that, may even shed light on the efficiency of capital markets. Research on the success of trading strategies in futures markets and their information efficiency and transparency is a well-established research area in the market for derivatives (Tessmann et al., 2021; Bosch, 2017; Gogas and Serletis, 2010; Lien and Xiang, 2010; Avgouleas and Degiannakis, 2009). Previous research, such as Bhardwaj et al.'s (2015) study, covered 27 commodity futures markets during the period 1959–2014 and found that futures markets generated risk premiums of 3.7% per annum as compared to the stock market. This finding entails that holding commodity futures in a portfolio improves diversification and returns at the same time.

Moreover, when comparing the returns generated under a buy-and-hold strategy in the stock and futures markets, the Sharpe ratios of the futures markets were nearly identical to those of the stock market, providing further proof that a diversified futures trading strategy can be a good investment vehicle (Bhardwaj et al., 2015). Initially, Gorton and Rouwenhorst (2006) conducted an identical study using a shorter time horizon and obtained the same findings, confirming that the US stock and futures markets during the 1959–2004 period demonstrate nearly identical risk-return profiles (Sharpe ratios) when portfolio rebalancing is applied.

This study was particulary motivated by Upperman's (2012), Briese's (2008) and Williams' (2005) books on the usage of CoT data in discretionary futures trading. They explored a short-term futures trading strategy which has been examined in different yet similar ways since the 20th century. From the books, a curiosity over the trading performance of such a strategy arose and whether it can be used to beat the markets. Williams' (2005) strategy involved operationalising the CoT reports' data into a simple index. By setting the positioning of today's commercial traders relative to the past 26 weeks' positioning's highs and lows, he created a trading indicator known as the 'Williams Commercial Index' (Williams, 2005). Williams (2005) stated that historical extremes in the index can accurately forecast a market reversal. Similar results were obtained by Jiler (1985) in the 1990s. His study, named 'The Forecasting Methodology', utilised commercial traders' net positioning as a forecasting tool for price changes. In his analysis, he observed the superiority of commercial traders over other market participants (Jiler, 1985). Briese (1990) extended Jiler's (1985) research, focusing on the relative rather than absolute net positioning of commercial traders (Briese, 1990).

Subsequently, two questions arise of whether a systematic interpretation of the CoT data can forecast market moves and whether the returns generated under this kind of trading strategy could exceed annual returns of the S&P 500.

While the current study has multiple goals, its main objective is to empirically backtest a trading strategy that primarily utilises CoT position data to predict market movements. The backtest includes all North American futures markets listed in the CoT report between 1986 and 2020. Upperman (2012) made a similar attempt, testing 38 different futures markets over a substantially lesser period.

To gain a complete picture of the CoT strategy, each futures market should be run through a backtest, including a stress-test. The CoT data helps to determine the investment universe. Since single-market testing does not meet the S&P 500 index's diversification levels, the CoT strategy should also consider trade across the defined investment universe in a portfolio scenario. Furthermore, this scenario will be subjected to trade rebalancing to emulate active portfolio management. Consequently, the goals of this research can be summarised into the following five action items:

- Use CoT data to perform a backtest of a short-term reversal strategy that utilises the CoT data as its foundation for all reportable CFTC futures markets for the period 1986–2020.
- Perform a backtest of the reversal strategy for all futures markets traded simultaneously in a portfolio and compare the gross and net returns with the S&P 500 returns.
- Statistically validate the annual returns of the portfolio.

- Measure portfolio performance in isolation during times of economic distress (1987, 2000–2002, 2008, 2020).
- Answer the following question based on collected data: can a futures market or a portfolio generate more risk-adjusted excess returns under a CoT strategy than under an S&P 500 buy-and-hold strategy?

Our results show that no risk premiums can be earned in the futures markets using CoT data in the analysed period and that returns are comparable to traditional stock markets. Overall, the portfolios did not outperform the reference index, indicating that CoT reports contribute to efficient derivatives markets. Our research provides novel evidence on the effect of the CoT reports on the transparency and informational efficiency of the derivatives market. Our study is the first research using a trading strategy backtest based on the CoT data for all monitored US futures markets for the period 1986–2020. Our results allow the derivation of important implications for regulators, exchange providers, and market participants.

The remainder of this paper is structured as follows. Section 2 presents the theoretical framework for this study and a review of related literature conducted to develop the research questions and research hypotheses. Section 3 describes the trading strategy dataset and models. In Section 4, all the backtest results of the CoT strategy are critically examined. This reflection begins with examining the long-only and long-and-short trading results per futures market. Next, the portfolio results for two different scenarios are inspected. Additionally, portfolio performance is viewed in a Monte Carlo simulation to validate the results. This section concludes by analysing portfolio performance during periods of crises. In Section 5, the findings are summarised and research gaps are addressed. Furthermore, we provide limitations of the backtest and suggestions for future research.

2 Theoretical framework, related literature, and research questions

2.1 Theoretical framework

To identify empirical evidence of the value of CoT data and subsequently build a trading strategy foundation, previous research and the CoT report must be analysed in terms of the three main market participants. This examination should include research regarding commercial market participants as well as that of non-commercials and non-reportables. All three groups of traders were rated based on their forecasting ability to obtain proof that the CoT data is useful for developing a trading strategy. Scientific research regarding this topic spans nearly a century, beginning with Keynes' theory of normal backwardation in the 1930s and leading to Merkoulova's study (2020) most recently. One of the essential questions that arose in most of these studies was whether risk premiums could be earned in the futures markets using CoT data and how they compared to returns generated in traditional stock and bond markets (Briese, 2008).

From 1986 to 2000, the US Commodity Futures Trading Commission (CFTC) published the CoT report biweekly, making the data freely available in 1995 (CFTC, 2021d). After 2000, the frequency of the CoT report changed to a weekly release (CFTC, 2021d). The legacy report distinguished between three groups of market participants:

commercials, non-commercials, and non-reportables. These reports included a detailed breakdown of futures contracts generated in a week and any changes relative to the previous week (Bernstein, 2012; CFTC, 2021b). From a regulatory perspective, the CFTC issues the CoT and other reports to "protect investors against manipulation, abusive trade practices, and fraud" (Logan and Scott, 2021). Therefore, one of the main goals of the CoT report can be understood to be the elimination of asymmetries such that each market participant can access the same information, ultimately promoting market efficiency.

While a single report does not provide much value, opinions on the impact of the CoT report have ranged from 'useless' to 'the holy grail' (Bernstein, 2012). Although, the new CoT reports further disaggregated market participants, most studies focused on the three main groups and examined them in terms of their forecasting ability, excess returns, and price pressure effects (CFTC, 2021c).

In previous studies, 'non-commercials' are often referred to as large traders, large speculators, or speculative funds (Briese, 2008; Bhardwaj et al., 2015; Bernstein, 2012). The disaggregated CoT report concluded that non-commercials are composed of commodity trading advisors (CTAs), commodity pool operators, hedge funds, commodity index traders (CITs), and other reportable speculators (CFTC, 2021c). The combined market shares of commodity futures owned by non-commercial participants ranged from 15% to 30% over a 20-year lookback period (Bhardwaj et al., 2015).

The category 'commercials' includes producers, consumers, processors, and merchants (CFTC, 2021c). Therefore, they can be defined as commercially involved traders buying and selling commodities or financial futures daily (CFTC, 2021c). Since the commercials' day-to-day business involves trading in spot and futures markets, they account for the most significant part of the market (Bhardwaj et al., 2015). While Briese (2008) argued that the percentage of the commercials' market share was around 73%, Bhardwaj et al. (2015) found that it was, on average, closer to 50% after analysing the positioning data of 27 commodity futures markets.

Lastly, the 'non-reportables' category comprises market participants that hold positions that fall below the CFTC's reportable requirements. The positions of these market participants are calculated by subtracting the commercial and non-commercial total positions from the total open interest (CFTC, 2021a, 2021c). Several authors referred to this group as small traders, retail traders, or small speculators (Briese, 2008; Bhardwaj et al., 2015; Bernstein, 2012). They hold the smallest market share ranging between 10% and 25% per futures market (Bhardwaj et al., 2015).

An essential concept in this study is the term hedging pressure. The influence of hedging pressure on market prices arises from commercials buying and selling futures contracts to hedge their price risk (Basu and Stremme, 2009). As part of the Keynesian backwardation theory, the hedging pressure hypothesis assumes that commercials trade in markets to transfer their price risk to non-commercials (Keynes et al., 2012; Basu and Stremme, 2009). Regarding this hypothesis, De Roon et al.'s (2000) study found empirical evidence that hedging affected the risk premiums and price changes in 20 futures markets. Additionally, Cheng and Xiong (2013) examined the relationship between the positioning of hedgers and price pressure. They found a positive correlation between commercials' short positioning and price changes, along with evidence that commercials do not solely rely on futures markets to hedge their risk in the spot market but speculation as well (Cheng and Xiong, 2013). Their results highlighted the following inverse correlation: when the short positioning of commercials increases, prices rise.

When the long positioning of commercials increases, prices decline. This discovery is consistent with hedging pressure theory, leading to the assumption that hedgers initiate short positions when prices increase and long positions when prices decrease, therefore acting inversely to non-commercials. Bosch's (2017) study, which examined three wheat futures markets, confirmed these findings.

In contrast, Kang et al. (2014) presented a new view on the hedging pressure theory. Considering the liquidity provision offered by commercials to non-commercials, they found empirical evidence that "liquidity provision allows commercials to recapture a large portion of the premium paid to non-commercials for price insurance in commodity markets" (Kang et al., 2014). These findings confirm that commercials are not always on the wrong side of the market but follow a different strategy than other participants.

2.2 Related literature

According to research based on the CoT data of corn, wheat, and soybean futures markets, large traders were consistent winners for the period 1951–1980 (Chang, 1985). Furthermore, Chang (1985) provided statistical evidence that large speculators have the excellent ability to forecast futures prices, supporting the theory of normal backwardation while Moran et al. (2020) found the opposite to be true when they looked at agricultural futures, highlighting there's little evidence of Keynesian risk premium. Chatrath et al. (1997) concluded that large speculators are generally profitable but criticised the forecasting ability of the CoT report. The backwardation theory can be confirmed by delivering proof of the inverse relationship between commercial and large speculator positioning. Generally, commercials take the position opposite to that of large speculators, leading to a net short position for commercials and a net long position for large speculators, or vice versa (Chatrath et al., 1997). Chatrath et al.'s (1997) study was recently extended to the futures energy markets in Merkoulova's (2020) article 'Predictive abilities of speculators in energy markets'. The CoT report data for the period 1986–2017 provides a more general application of Chang's (1985) original study, confirming that the consistent profitability of large traders and losses on the side of commercials can also be applied to energy futures (Merkoulova, 2020). Merkoulova's (2020) empirical tests further support Keynes' theory of normal backwardation and the role of commercials as hedgers. Smales (2022) also exhibited commercials to have "contrarian behaviour consistent with hedging as they are selling as prices rise" (p.16). Humpe and Zakrewski (2015) examined CoT data from a different perspective and attempted to identify stock market reversals based on futures data for non-commercials and non-reportables. To identify value in the CoT report, they utilised regression analysis using data for the period 1993–2014. They found that price changes in the S&P stock index were positively correlated to large speculator positions and negatively correlated to small speculator positions (Humpe and Zakrewski, 2015). Humpe and Zakrewski (2015) provided further empirical evidence of the forecasting ability of large speculators and highlighted that the extreme positioning of retail traders (small traders) should be reduced.

In contrast, a look into a subset of the non-commercials, i.e., the track record of CTAs, showed an alpha close to null relative to US Treasury bills for the period 1994–2012 (Bhardwaj et al., 2014). Since CTAs account for a significant portion of non-commercials, Chang's (1985) and Chatrath et al.'s (1997) studies and subsequently, the forecasting ability of non-commercials should be questioned. A more recent analysis of the most prominent participant among non-commercials, the CITs, showed that CITs significantly impact futures commodity prices (Frenk and Turbeville, 2011). Counterarguments against the impact of speculators positioning on futures prices were observed in the wheat, corn, and soybean futures markets (Maul et al., 2015). Maul et al. (2015) employed the statistical Johansen test along with vector autoregression (VAR) and vector error correction models. They found that short-term price fluctuations occur before open interest changes, and thus, speculators did not exhibit price pressure effects in the explored futures markets. This study does not examine data regarding commercials; however, it can be assumed that these participants are the source of price pressure, given the empirical evidence against large speculators. Recently, Hayward (2018) found that extreme speculative positions in the FX futures markets had no effect on prices, rendering non-commercial data unsuitable to predict future price moves.

Sanders et al. (2009) further supported Maul et al.'s (2015) argument against the influence of speculators on price and analysed commercials' data. Conducting a Granger causality test for ten agricultural futures markets, they concluded that speculators cannot forecast prices effectively (Sanders et al., 2009) and that CoT data should not be a stand-alone tool in a trading strategy (Sanders et al., 2009). Furthermore, they observed that commercials can forecast prices with a higher probability of success (Sanders et al., 2009) than that of their non-commercial counterparts. Sanders et al.'s (2009) research points out that other influences, such as price variables, must flow into a trading strategy. They found commercials to have superior forecasting abilities compared with non-commercials. During an analysis of the trading behaviours of all three participant groups for the period 2003–2012 in the S&P 500 futures market, the usage of extreme commercial positions was criticised, once again highlighting the need for combining it with price data to create a complete trading strategy (Smales, 2013). Despite Smales' (2013) criticism of using only CoT data, he concluded that only commercials could generate an abnormal risk-return. Chatrath and Song's (1999) study also investigated the relationship between non-reportable and non-commercial positioning and price changes in the wheat, oat, soybean, corn, and cotton futures markets and found that neither of the two participant groups caused price volatility. As Sanders et al. (2009) and Smales (2013) showed in their studies, this finding can be similarly interpreted as follows: commercial positions could cause price volatility and big market moves. Gao's (2017) findings confirmed this as he demonstrated that the hedging demand of commercials causes volatility and directional price moves.

The most relevant scientific research for this study were the works of Basu et al. (2006) and Basu and Stremme (2009). Their findings confirmed a significant alpha of 15% with a Sharpe ratio of 1.0 relative to the S&P 500 stock index performance (Basu et al., 2006). The long-only trading strategy used S&P 500, copper, and oil futures for the period 2000–2006. Trading signals were generated on a weekly basis and the portfolio was rebalanced at the same frequency (Basu et al., 2006). The hedging pressure indicator revealed the position from 0 to 1 of 'extreme' commercials and non-reportables relative to the past year and created the basis for the study's timing strategy (Basu et al., 2006). This research was extended three years later by Basu and Stremme (2009). They utilised the same trading strategy based on commercial positioning to forecast market reversals, confirming significant annual excess returns of 7.9% (relative to the S&P 500 returns) over the period 1999–2007 (Basu and Stremme, 2009). Moreover, non-reportable data was examined, tested, and combined with the commercial data, leading to an even higher excess return of 8.66% (Basu and Stremme, 2009). Basu and Stremme (2009) showed the value of commercials data in a trading strategy and confirmed that merging commercial and non-reportable data has its advantages.

Additionally, Briese (2008) utilised position extremes to develop and backtest a trading strategy, like the works of Bernstein (2012), Upperman (2012), Basu and Stremme (2009), De Roon et al. (2000) and Cheng and Xiong (2013). He backtested a trading strategy in 38 futures markets based on commercial and non-commercial data for the period 2000–2007 (Briese, 2008). Unlike Upperman (2012), Briese (2008) solely utilised trading signals based on the CoT data but provided a new angle on the data with his application of moving averages to signal position changes. His results provide a general idea of the profitability of the trading strategy. With over 3,502 trades in 35 futures markets, the profits per trade ranged between \$12 and \$2,345 leading to a total profit of 1.7 million dollars (Briese, 2008). On the flip side, Briese (2008) provided no information regarding the risks taken during the trades or the portfolio's value, making it difficult to rate the strategy's risk-return profile.

Based on the vast literature on this topic, commercials index data was identified to generate a Sharpe ratio that exceeds that of the S&P 500 when long signals are generated at an index value of 0.8 and short signals at an index value of 0.1 (Basu and Stremme, 2009). Williams (2005) and Briese (2008) provided a variety of these signal values. For long-term trading, Briese (1990) utilised a lookback period of 24 months and waited for an index value of 0.25 (25%) to signal a short trading opportunity and an index value of 0.75 (75%) to signal a long trading opportunity. Since the focus was a short-term reversal strategy, Williams (2005) utilised a more suitable lookback period of 26 weeks using the same index values as trade signals.

The non-reportables data was identified as a contrarian indicator that can be used as an indicator to exit positions or confirm commercial index signals (Basu et al., 2006; Maul et al., 2015; Sanders et al., 2009). The signals generated by the non-reportable index can be combined with those of the commercial index to improve returns (Basu and Stremme, 2009). The data for the backtest was interpreted as per Briese's (2008) description. Based on the commercial and non-reportable data in an index from zero to one, the following assumptions were extracted from previous literature (Basu and Stremme, 2009; Briese, 2008; Humpe and Zakrewski, 2015; Williams, 2005):

1 A commercial CoT index value range of 0.7–0.9 suggests a commercial buying climax, leading to a potential upward trend change and equates to a long signal.

2 A commercial CoT index value range of 0.1–0.3 suggests a commercial selling climax, leading to a potential downward trend change and equates to a short signal.

3 A non-reportable CoT index value of 0.7 suggests a non-reportable buying climax, predicting a potential downward trend change and equates to a short signal.

4 A non-reportable CoT index value range of 0.2–0.3 suggests a non-reportable selling climax, predicting a potential upward trend change and equates to a long signal.

The indices above did not consider the price. However, research suggests that the hedging pressure indices should not be employed as a stand-alone tool and price related measures must be considered (Sanders et al., 2009; Hammerschmid, 2018). Therefore, the assumptions for trading with the CoT data were extended to avoid large drawdowns due to premature CoT index signals (Upperman, 2012; Sanders et al., 2009; Williams, 2005). Williams (2005) utilised a 52-week moving average for validation, while Upperman (2012) utilised a 10-day moving average. Due to the CoT data's weekly resolution, the parameters explored in this study were fitted to a 10-week (50-day) time horizon to suit a short-term reversal strategy. Therefore, a fifth assumption was added:

5 A 10-week moving average on the closing price validates the short-term trend. CoT signals must be validated by a rising (long) or declining (short) moving average.

Several studies, such as Basu and Stremme (2009), utilised empirical backtests; however, no studies utilised all futures markets listed in the historical CoT reports. Additionally, Basu and Stremme (2009) only tested commercial data in a long-only scenario without considering short-selling. Moreover, the scope of previous papers was limited to
38 futures markets; however, most authors remain focused on the specific asset classes (Basu et al., 2006; Chang, 1985; Chatrath and Song, 1999; Merkoulova, 2020). The most comprehensive test in previous literature covered 38 markets (Briese, 2008). Another study examined 27 futures markets, providing proof of the risk premiums and compared them with those of the stock market (Bhardwaj et al., 2015). However, the investment universe in studies that examined more than 30 assets did not follow a defined pattern. Not all futures markets monitored by the CFTC were covered. The period observed in prior literature provides an incomplete picture since only Merkoulova (2020) inspected data from the inception of the CoT report in 1986 onwards.

A trading strategy backtest based on the CoT data for all monitored US futures markets for the period 1986–2020 has not yet been conducted. This time gap raises questions about the strategy's stability and reliability over time and its performance during periods of economic distress. Furthermore, portfolio scenarios considering certain futures markets were explored by three authors (Basu et al., 2006; Bhardwaj et al., 2015; Gorton and Rouwenhorst, 2006). Limited research was conducted regarding CoT portfolio scenarios, providing an opportunity for modelling, and executing a trading strategy within a portfolio. This scenario could improve the strategy's risk-return profile (annual return, annual volatility, Sharpe ratio) and highlight the diversification effects relative to an S&P 500 buy-and-hold strategy.

Basu and Stremme (2009) combined the commercial and the non-reportable indices to generate more reliable trading signals. However, their focus remained only on commercial and non-reportable data, neglecting non-commercial CoT data (Basu and Stremme, 2009). Basu and Stremme (2009) emphasised that the research regarding the use of non-commercial data for trading signals was not consistent. However, since the commercial index remains the primary indicator used to generate trading signals, it can be assumed that non-commercial data can also be used as a contrarian indicator since this data has an inverse correlation to the position of commercial data (Chatrath et al., 1997). Therefore, the assumptions from prior literature were extended to incorporate
non-commercial data in the trading strategy as follows:

6 A non-commercial CoT index value of 0.7 suggests a non-commercial buying climax, predicting a potential downward trend change and equates to a short signal.

7 A non-commercial CoT index value of 0.2–0.3 suggests a non-commercial selling climax, predicting a potential upward trend change and equates to a long signal.

Additional validation of trading signals was added in the form of the commercials' relative open interest. The research findings confirmed that the second biggest market participant, i.e., the large traders, accounts for a maximum of 30% of the futures markets on average (Bhardwaj et al., 2015). For confirmation that commercials are the leading force in the market, the relative open interest of commercials should exceed this figure as follows:

8 The commercial long open interest must make up at least 30% of the total open interest to validate a long signal.

9 The commercial short open interest must make up at least 30% of the total open interest to validate a short signal.

2.3 Research questions

The research system allows modelling a trading strategy and creating a normalised dataset. For this purpose, Jansen's (2020) workflow was adapted and combined with the knowledge discovery in databases (KDD) process (Fayyad et al., 1996) to conduct systematic data mining. The systematic trading strategy was thus created based on the following six steps:

Step 1 Obtain predictive data.

Step 2 Define the investment universe.

Step 3 Design trading signals.

Step 4 Combine trading signals.

Step 5 Execute trades for each market.

Step 6 Execute trades in portfolio and optimise.

Basu and Stremme (2009) found evidence of a significant performance of the CoT strategy relative to that of the S&P 500. Therefore, in this study, the CoT strategy was examined based on a similar hedging pressure index. The same indices with various lookback periods were utilised by De Roon et al. (2000), Williams (2005), Briese (2008), Gao (2017) and Upperman (2012). Basu and Stremme (2009) customised the reversal strategy further and integrated the relative positioning data into indices similarly to the way they were utilised in the above studies. In Williams' (2005) reversal strategy open interest in commercials and price variables were also considered. In Williams' strategy open interest in commercials and price variables were considered (Basu and Stremme, 2009; Williams, 2005). Furthermore, exploring the data's full potential requires that the trading strategy be tested in both a long-only and a long-and-short environment. To extend Gurrib's (2009) research, the CoT strategy's portfolio performance was analysed for 1987, 2000–2002, 2008, and 2020 to compare it to the market proxy's performance. Incorporating this analysis of the crisis performance, the following list of research questions was developed:

- How will futures markets perform under the CoT trading strategy relative to the risk-adjusted return of the S&P 500?
- Can a short-term reversal strategy in a futures portfolio outperform a simple buy-and-hold S&P 500 strategy (before or after transaction costs)?
  a compare portfolio returns against the S&P 500 annual return and Sharpe ratio
  b examine how the risk-return profile changes when the portfolio is optimised with rebalancing
  c analyse the portfolio's annual return distribution in a Monte Carlo simulation and how it compares to the S&P 500.
- How does the portfolio performance change during times of crises (1987, 2000–2002, 2008, 2020)?

By answering these questions, this study also aims to provide evidence on whether excess premiums can be earned in futures markets applying a reversal strategy using CoT data and whether risk-adjusted performance is comparable to traditional stock markets. From this, indications can be derived whether CoT reports contribute to efficient derivatives market and, ultimately, whether derivatives market was efficient during the analysed time period.

3 Data and methodology

The primary dataset was generated from the yearly data available on the CFTC website (CFTC, 2021e). The combined CoT data comprised 25,074,000 data points, creating the dataset's basis. Hence, the KDD process was employed to extract meaningful information from the large dataset (Fayyad et al., 1996).

Strongly correlated markets representing the same underlying asset were avoided in this study by only choosing one futures market per CFTC commodity code based on the CFTC contract market code. The chosen market was the one with the highest trading activity measured by open interest. For example, the CFTC commodity code '001' refers to several wheat futures contracts differentiated by their CFTC contract market codes. There are wheat futures contracts traded at the Chicago Board of Trade (CBT), Kansas City Board of Trade (KCBT), MidAmerica Commodity Exchange (MCE), and Minneapolis Grain Exchange (MGE). Only the wheat contract market code with the highest open interest was retained for the backtest. In this case, the CBT contract was used for backtesting. The first process of data selection and cleaning of the CoT data generated a market list of 71 unique futures markets. These futures contracts were split into several asset classes. Henceforth, the data was categorised into currencies, bonds and interest rates, stock indices, animal products, softs (agricultural products), grains, metals, and energy markets, as shown in Table 1.

Table 1 Futures markets covered by the CoT report

Futures market	*From year*	*To year*
Currencies		
AUSTRALIAN DOLLAR	1987	2020
BITCOIN	2018	2020
BRAZILIAN REAL	1995	2020
BRITISH POUND STERLING	1986	2020
CANADIAN DOLLAR	1986	2020
EURO FX	1998	2020
MEXICAN PESO	1995	2020
RUSSIAN RUBLE	2009	2020
SWISS FRANC	1986	2020
US DOLLAR INDEX	1986	2020
SOUTH AFRICAN RAND	1998	2020
NEW ZEALAND DOLLAR	1999	2020
JAPANESE YEN	1986	2020
Bonds and interest rates		
1-MONTH SOFR	2018	2020
2-YEAR US TREASURY NOTES (ZT)	1990	2020
3-MO. EUROYEN	2001	2009
5-YEAR US TREASURY NOTES (ZF)	1989	2020
6.5–10-YEAR US TREASURY NOTE (ZN)	1986	2020
EURODOLLARS	1986	1992
30-DAY INTEREST RATES (ZQ)	1988	2020
LONG-TERM US TREASURY BONDS (ZB)	1986	2020
US TREASURY BILLS	1995	2000
Stock indices		
DOW JONES INDUSTRIAL	2002	2020
NASDAQ 100	1999	2020
RUSSEL 2000	2001	2020
S&P 400 MIDCAP	1992	2020
S&P 500 ANNUAL DIVIDEND	2017	2020
EUROTOP 100	1998	2000
S&P 500	1986	2020
VALUE LINE COMPOSITE	1986	1996
E-MINI MSCI EAFE	2008	2020
NEW YORK HDD	2008	2020
NIKKEI 225	1990	2020

Table 1 Futures markets covered by the CoT report (continued)

Futures market	*From year*	*To year*
Animal products		
FROZEN PORK BELLIES	1986	2020
FEEDER CATTLE	1986	2020
LIVE CATTLE	1986	2020
LIVE HOGS	1986	2020
BUTTER	2006	2020
CHEESE	2012	2020
MILK	1997	2020
Softs		
CANOLA	2018	2020
ETHANOL	2009	2020
COCOA	1986	2020
COFFEE C	1986	2020
ORANGE JUICE	1986	2020
LUMBER	1986	2020
COTTON NO. 2	1986	2020
SUGAR NO. 11	1986	2020
COAL	2019	2020
Grains		
OATS	1986	2020
CORN	1986	2020
ROUGH RICE	1994	2020
SOYBEAN MEAL	1986	2020
SOYBEAN OIL	1986	2020
SOYBEANS	1986	2020
WHEAT	1986	2020
Metals		
ALUMINIUM	2014	2020
IRON ORE	2014	2018
COPPER	1986	2020
GOLD	1986	2020
STEEL	2013	2020
PALLADIUM	1986	2020
SILVER	1986	2020
PLATINUM	1986	2020

Table 1 Futures markets covered by the CoT report (continued)

Futures market	*From year*	*To year*
Energy		
CRUDE OIL	1986	2020
NATURAL GAS	1990	2020
NO. 2 HEATING OIL, N.Y. HARBOR	1986	2020
RBOB	1986	2020
PROPANE GAS	2020	2020
Other		
S&P GSCI COMMODITY INDEX	1994	2020
VIX FUTURES	2005	2020

To have consecutive historical price data in the dataset, the front-month was switched when the current contract month's open interest fell below a more distant contract month (Briese, 2008). At the time of contract rollover, the price was back-adjusted automatically by the data vendor (Barchart, 2021; Briese, 2008).

A realistic execution logic of the trading strategy required the current week's closing price mapped to the nearest CoT datapoint in the past. With CoT data being reported biweekly and weekly, the report data must lay in the past relative to the price data to avoid look-ahead bias. This dataset mapping procedure means that, in practice, e.g., wheat futures were priced at 317.75 on 20 January 1986 and matched with CoT report data from 15 January 1986. This logic was applied to the entire dataset. The merge of the CoT and price data concludes the steps of data selection and pre-processing (Fayyad et al., 1996).

The operationalisation of the CoT data began with creating a hedging pressure index for the three market participants. The selected indices are the 'commercial index', the 'non-commercial index', and the 'non-reportable index'. Additional indices for the 'commercial long open interest' and 'commercial short open interest' strategies were calculated to check the percentage of the market attributed to commercials. The indices were estimated using formulas suggested in previous research (Basu and Stremme, 2009; Briese, 2008; Williams, 2005). The last 26 weeks (n = 26) were chosen as the lookback period to represent the participants' half-yearly position climaxes (Williams, 2005). A higher lookback period would provide an unrealistic view for a short-term strategy. Considering the 26-week lookback period the following formulas were used to calculate the hedging pressure indices:

- *Commercial index* (to calculate signal values between 0 and 1) from total commercial positions:

$$\frac{(\text{Current weeks non-commercial position} - \text{Lowest non-commercial position of } n)}{(\text{Highest non-commercial position of } n - \text{Lowest non-commercial position of } n)}$$

- *Non-commercial index* (to calculate signal values between 0 and 1) from total non-commercial positions:

$$\frac{(Current\ weeks\ non\text{-}commercial\ position - Lowest\ non\text{-}commercial\ position\ of\ n)}{(Highest\ non\text{-}commercial\ position\ of\ n - Lowest\ non\text{-}commercial\ position\ of\ n)}$$

- *Non-reportable index* (to calculate signal values between 0 and 1) from total non-reportable positions:

$$\frac{(Current\ weeks\ non\text{-}reportable\ position - Lowest\ non\text{-}reportable\ position\ of\ n)}{(Highest\ non\text{-}reportable\ position\ of\ n - Lowest\ non\text{-}reportable\ position\ of\ n)}$$

- *Commercial long open interest* (to calculate signal values between 0 and 1) from commercial long positions and open interest:

$$\frac{(Current\ weeks\ absolute\ commercial\ long\ contracts\ )}{(Total\ open\ interest)}$$

- *Commercial short open interest* (to calculate signal values between 0 and 1) from commercial short positions and open interest:

$$\frac{(Current\ weeks\ absolute\ commercial\ short\ contracts)}{(Total\ open\ interest)}$$

A simple 10-period moving average based on closing prices represents the trading strategy's price variable. These ten periods equate to a lookback period of 50 days, considering five days per week. This time frame was chosen based on its prominence since the 50-day moving average is one of the most utilised trading tools (Maverick, 2020). To obtain the simple moving average, the mean of the last ten weeks' closing prices was calculated.

- *Simple moving average* (10-week period) from closing prices:

$$(P1 + P2 + \ldots + P10)/n$$

where

P closing price in period i

n total number of weeks in lookback period ($n = 10$).

The trading strategy was modelled in a standardised architecture as shown in Figure 1 (Jansen, 2020; Rodriguez, 2020). This architecture allowed for the simulation of the market environment, including the broker, the commission structure, order execution logic, and detailed strategy parameters (Rodriguez, 2021a, 2021b, 2021c, 2021d).

This strategy considers commercial, non-commercial, and non-reportable data. Assumptions 1 and 2 (Section 2) suggest that the commercials have the highest market power, which is why a check of commercial hedging pressure is a primary signal provider. Additionally, since empirical evidence was found that non-reportables tend to be on the wrong side of the market most of the time, Assumptions 3 and 4 (Section 2) were selected as the second signal provider. For signal validation, the moving average and open interest indices were used. The signals were designed based on assumptions 5, 8 and 9 (Section 2). The data feeds had to meet the exact criteria listed to trigger a trading signal and subsequent order execution. In the first stage, two backtests were executed for each futures market. One backtest allowed the model to only go long, while the second permitted both long and short orders (Rodriguez, 2021c). The same approach was extended to the portfolio scenarios. For a trade exit, the reversed price variable, i.e., the moving average, and an extreme non-commercial index was integrated based on Assumptions 5, 6, and 7 (Section 2). A combination of all the assumptions led to the subsequent trade logic conditions:

1 For a long entry, the following requirements had to be met:
   a commercial index above 0.7 with a 26-week lookback period
   b commercial long positions to be at least 30% of the total open interest
   c non-reportable index below 0.3 with a 26-week lookback period
   c current 10-period moving average is higher than last week's moving average.
2 To trigger a long exit, the following conditions had to be met:
   a non-commercial index above 0.7 with a 26-week lookback period
   b current 10-period moving average is lower than last week's moving average.
3 To trigger a short entry, the following conditions had to be met:
   a commercial index below 0.3 with a 26-week lookback period
   b commercial short positions to be at least 30% of the total open interest
   c non-reportable index above 0.7 with a 26-week lookback period
   d current 10-period moving average was lower than the last week's moving average.
4. To trigger a short exit, the following conditions had to be met:
   a non-commercial index below 0.3 with a 26-week lookback period
   b current 10-period moving average was higher than last week's moving average.

According to Jansen (2020), transaction cost must be included in backtests to avoid biased results. Boussema et al. (2002) studied the average trading costs in developed and emerging markets and found that the average transaction cost percentage per trade was between 0.15% and 0.23% in developed markets (Boussema et al., 2002). For the backtest, the average of the transaction costs obtained in the study was taken and rounded. Therefore, the model assumed a fixed rate of 0.2% transaction costs relative to the order value per trade. Transaction costs were subtracted from the portfolio value once during order entry and then again when the trade was closed (Rodriguez, 2021b).

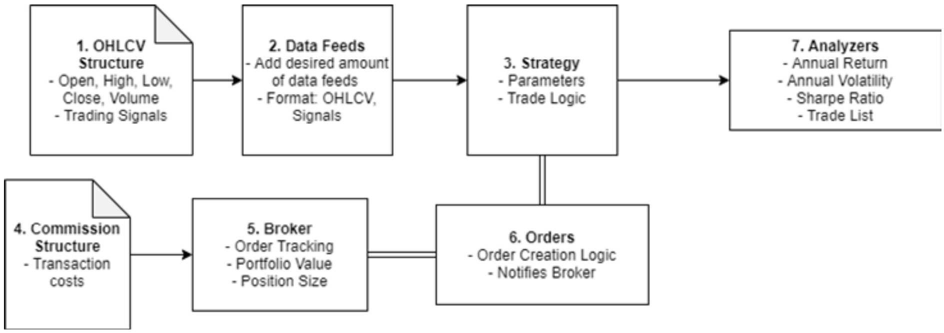
The broker simulation allowed for position sizing and portfolio and order tracking to create a natural market environment (Rodriguez, 2021a).

- portfolio starting value: $1,000,000
- position value per trade: 50% of the portfolio value, allowing only one trade at a time.

The static portfolio scenario allowed multiple futures markets to trade simultaneously, leading to parameter changes in Section 4:

- portfolio starting value: $1,000,000
- position value per futures market: 1% of the portfolio value, allowing a maximum of 71 trades at a time.

Figure 1 The backtest model



Notes: This figure presents the seven components of the backtest, which was modelled in Python. First, for each futures market, the input dataset was in the open, high, low, close, volume (OHLCV) format, including the custom indices that were created to provide trading signals. Second, the data feeds component allows single or multiple data feeds to be read simultaneously. Third, the data feeds were read by the strategy component to scan the data for the programmed parameters. Fourth, to simulate real market conditions, each executed trade generated transaction costs based on their position size. Fifth, a broker was simulated to track open and closed orders, the running portfolio's value, and position size. In the sixth step, the order creation module simulated the order being sent to the exchange. Finally, trades were monitored and summarised by analysers to generate statistics regarding annual return, volatility, and the Sharpe ratio.

Source: Adapted from Jansen (2020)

The rebalanced portfolio scenario in Section 4 introduced dynamic position sizing, rebalancing the portfolio with each trade:

- portfolio starting value: $1,000,000
- position value per futures market: the maximum was 1.3% of the portfolio value, allowing each futures market to execute trade signals in the same direction until the target percentage of 1.3% was reached, or the trade was closed (Rodriguez, 2021e).

The buy and sell orders were created based on the strategy's trading signals after passing through the broker. Once a signal was generated based on the above strategy parameters, a buy or sell market order was created using the current week's closing price and executed using the following week's opening price (Rodriguez, 2021e). The trading model behaved in the same way for exit signals. After each order's execution, the broker was notified. The broker then updates the portfolio and position value accordingly (Rodriguez, 2021d).

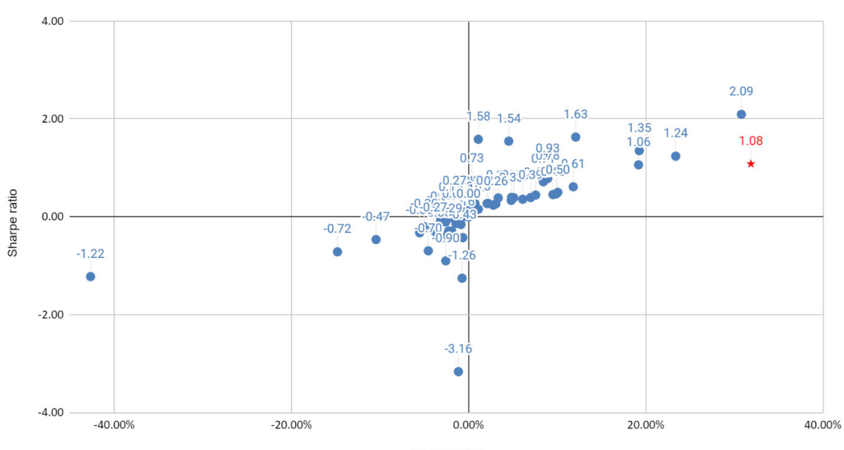
The backtest results included statistics such as relative and absolute profit/loss figures and the dates for each trade. The yearly performance outlined in the next part was measured based on the percentage profit and loss. For statistical significance, additional measures related to the annual return are calculated. Three metrics are used for the standardised backtest to evaluate the risk-return profile per market and portfolio (Python Software Foundation, 2019; Quantopian Inc., 2020). Although an unconventional assumption, the risk-free rate for calculating the Sharpe ratio was assumed to be zero to obtain unskewed and comparable annual results for the S&P 500 buy-and-hold and the CoT strategies. The risk-return metrics are discussed in Section 4.

The Monte Carlo simulation was adapted from Sharma's (2019) research and was performed to generate the expectations for future returns. With this simulation, an attempt was made to create a high number of randomised annual returns, given the portfolio's annual volatility, and provide a probabilistic view of the same (Sharma, 2019). The portfolio's annual return and annual return volatility derived from the backtest were used to simulate 100,000 randomised portfolios over the same backtest period (1986–2020). The annual returns assumed a Gaussian distribution (Harris et al., 2020; SciPy Community, 2021; Sharma, 2019). For the portfolio generation, annual returns were taken as the mean, while annual return volatility was taken as the standard deviation of the distribution (Harris et al., 2020; SciPy Community, 2021). The samples allowed for the creation of a probability distribution of the annual returns to examine the statistical significance of portfolio returns at the 1%, 5%, and 10% levels, both numerically and graphically.

4 Results and discussion

An S&P 500 buy-and-hold strategy was used as a benchmark for evaluating whether a CoT strategy could be successful in the futures market. The S&P 500 returns were generated through a buy-and-hold backtest to generate performance measures for the period 1986–2020. The same backtest engine, portfolio size, and commission scheme used in the previous section was adopted. This backtest differed from the other backtests in that only one trade was executed in January 1986, which remained open until December 2020. For the market proxy (benchmark), the strategy generated a 32% annual return and 30% standard deviation with a Sharpe ratio of 1.07. To beat the proxy, a futures market or portfolio must attain a Sharpe ratio higher than 1.07. The CoT strategy was backtested for the periods shown in Table 1.

Figure 2 Sharpe ratio and annual return of the long-only CoT strategy (see online version for colours)



Notes: The blue dots represent the Sharpe ratios of the 71 tested futures markets. The red star represents the Sharpe ratio of the S&P 500 buy-and-hold strategy.

Source: Own results

Table 2 Long-only backtest results for the stock indices and bonds and interest rates

Market names	*P in % for $\mu 1 > \mu 0$*	*Significance*	*Annual return (μ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
S&P500 buy-and-hold	H0		32.00% (µ0)	30.00%	–0.78	8.32	1.07
Bonds and interest rates			µ1				
1-MONTH SOFR	0.38%	***	1.10%	0.69%	7.17	64.20	1.58
2-YEAR US TREASURY NOTES (ZT)	0.00%	***	–0.64%	1.49%	–4.61	85.03	–0.43
3-MO. EUROYEN	0.00%	***	–1.14%	0.36%	–2.08	15.87	–3.16
5-YEAR US TREASURY NOTES (ZF)	0.00%	***	0.73%	2.81%	–0.40	9.43	0.27
6.5–10-YEAR US TREASURY NOTE (ZN)	0.00%	***	–0.85%	4.71%	–5.13	98.22	–0.16
EURODOLLARS	0.00%	***	–0.73%	0.58%	–0.28	14.70	–1.26
30-DAY INTEREST RATES (ZQ)	0.00%	***	0.37%	0.51%	2.77	25.93	0.73
LONG-TERM U.S. TREASURY BONDS (ZB)	0.00%	***	0.53%	7.65%	–0.10	6.46	0.11
U.S. TREASURY BILLS	1.08%	**	–2.84%	15.82%	–0.45	18.26	–0.10
Stock indices							
DOW JONES INDUSTRIAL	0.00%	***	8.44%	12.45%	–1.26	20.44	0.71
NASDAQ 100	0.41%	***	19.19%	18.13%	–1.05	17.27	1.06
RUSSEL 2000	0.00%	***	3.35%	9.86%	–0.35	11.45	0.38
S&P 400 MIDCAP	0.00%	***	8.93%	11.86%	–0.54	13.19	0.78
S&P 500 ANNUAL DIVIDEND	2.35%	**	4.54%	2.90%	9.89	123.08	1.54
EUROTOP 100	45.95%		30.80%	13.26%	–0.66	3.92	2.09
S&P 500	0.00%	***	19.27%	13.78%	–0.63	6.72	1.35
VALUE LINE COMPOSITE	0.02%	***	12.10%	7.18%	–0.59	11.60	1.63
E-MINI MSCI EAFE	31.32%		23.39%	18.31%	4.04	52.07	1.24
NEW YORK HDD	-	-	0.00%	0.00%	0.00	0.00	0
NIKKEI 225	0.00%	***	4.84%	14.80%	–0.35	6.28	0.39

Notes: The P-values in Table 2 were generated using a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: $\mu 0 > \mu 1$), was tested against each market (H1: $\mu 1 > \mu 0$).
The p-values reflect the probabilities of each market outperforming the benchmark with *** ($p < 0.01$), signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

Source: Own results

Table 3 Long-only backtest results for currencies

Market names	*P in % for $\mu 1 > \mu 0$*	*Significance*	*Annual return (μ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
S&P500 buy-and-hold	H0		32% (µ0)	30.00%	–0.78	8.32	1.07
Currencies			µ1				
AUSTRALIAN DOLLAR	0.00%	***	–0.27%	8.08%	–0.64	9.85	0.01
BITCOIN	-	-	0.00%	0.00%	0.00	0.00	0
BRAZILIAN REAL	0.43%	***	–3.90%	11.62%	–0.86	7.10	–0.28
BRITISH POUND STERLING	0.00%	***	–1.42%	8.06%	–0.55	6.21	–0.14
CANADIAN DOLLAR	0.00%	***	–1.90%	5.94%	–1.07	20.29	–0.29
EURO FX	0.00%	***	–4.54%	6.38%	–0.95	10.36	–0.70
MEXICAN PESO	0.00%	***	–0.10%	5.77%	–1.64	25.52	0.01
RUSSIAN RUBLE	3.88%	**	–2.42%	5.98%	–3.81	44.51	–0.38
SWISS FRANC	0.00%	***	–3.49%	8.36%	–0.67	11.70	–0.38
US DOLLAR INDEX	0.00%	***	–2.57%	2.84%	–0.92	23.29	–0.90
SOUTH AFRICAN RAND	2.71%	**	–2.58%	14.54%	–0.45	3.99	–0.11
NEW ZEALAND DOLLAR	0.00%	***	–1.06%	8.30%	–1.39	16.13	–0.09
JAPANESE YEN	0.00%	***	2.09%	9.23%	1.57	18.37	0.27

Notes: The P-values in Table 3 were generated using a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: $\mu 0 > \mu 1$), was tested against each market (H1: $\mu 1 > \mu 0$). The p values reflect the probabilities of each market outperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

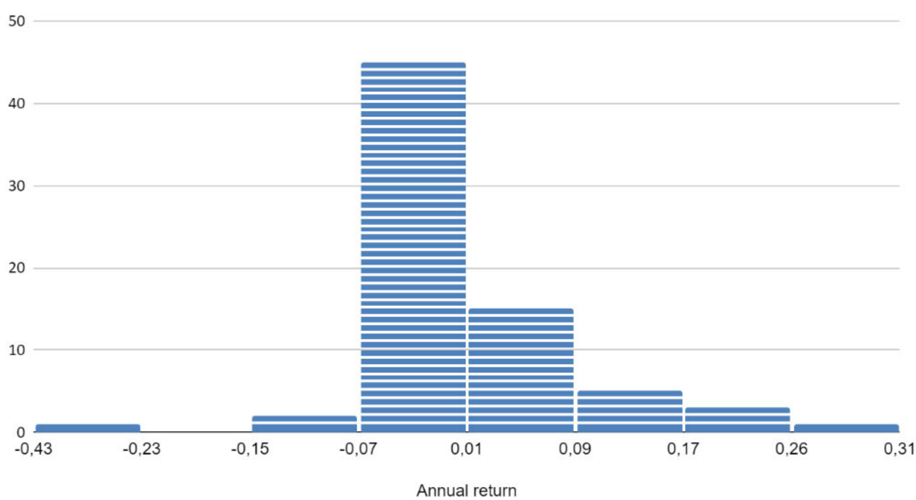
Source: Own results

The application of the long-only strategy in 71 futures markets generated the Sharpe ratios and annual return distribution displayed in Figure 2. Figure 2 shows that a majority of the markets had Sharpe ratios ranging between –2 and 2, with outliers ranging from –3.16 to 2.09. The market proxy was displayed using a star in Figure 2.

The highlights of this backtest are presented in Tables 2 and 3. As shown in Table 2, the 1-MONTH SOFR, EUROTOP 100, S&P 500, S&P 500 ANNUAL DIVIDEND, VALUE LINE COMPOSITE, and E-MINI MSCI EAFE markets showed superior risk-adjusted returns with the CoT strategy than with the benchmark. These superior Sharpe ratios can be attributed to a lower annual return volatility than that generated by the S&P 500 buy-and-hold strategy. A distinctive positive performance spreads across all markets in the category stock indices. The Sharpe ratios of the outperforming markets ranged from 1.24 to 2.09, exceeding the market proxy's Sharpe ratio of 1.07. Additionally, the category with the most future markets outperformed the market proxy. Five out of ten stock indices had a higher Sharpe ratio than the market proxy, while all ten stock indices showed a positive Sharpe ratio. Furthermore, the stock indices displayed less than 50% of the S&P 500's drawdown, confirming that the long-only CoT strategy has a superior risk-adjusted return profile among stock indices.

Another asset class highlighted for its negative performance were the currencies. With 11 out of 12 currencies showing a negative return over the backtest period. The negative performance of the currencies indicates that CoT data may not provide profitable trading signals for this asset class. The results for the currencies are presented in Table 3.

Figure 3 Long-only annual return distribution (see online version for colours)



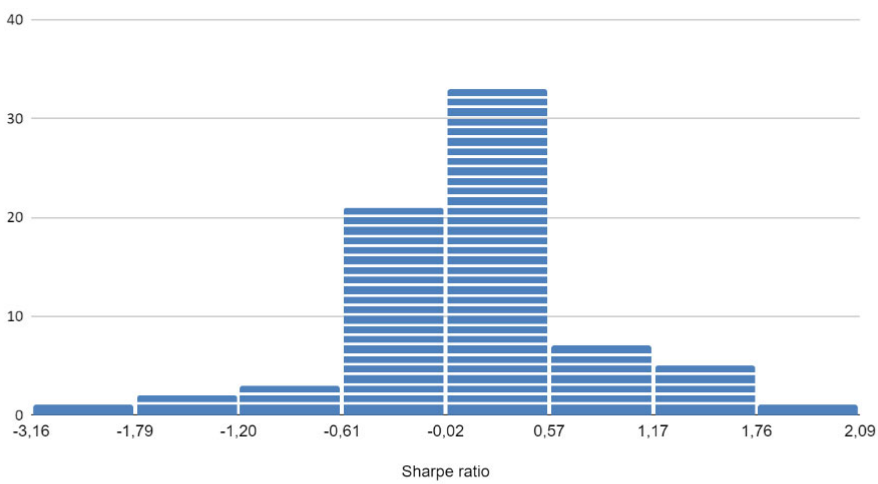
Notes: The x-axis represents the annual return in percent. Therefore, the 0.01 displayed on the distribution is equal to 1% in percentage terms. For the y-axis, the number of markets in each return band was counted, e.g., more than 40 markets fell into the –7%–1% band.

Source: Own results

Taking a statistical view of the annual return and Sharpe ratio, both distributions required further examination (Figures 3 and 4). For the annual return, the mean (excluding the market proxy) across the 71 markets was 1.32%. Since the markets were tested over different periods, this figure in itself is not significant. However, the annual returns in the distribution (Figure 3) shows that 15 markets fell into the return band of 1% to 9% and 45 markets in the –7% to 1% band. This distribution (Figure 3) displayed a standard deviation of 9%. Since most markets fall into negative return distributions, outliers can cause a positive skew. The performance of EUROTOP 100 market was the closest to the market proxy's annual return. With a mean at 1.3% and a high annual return of 30.8%, the CoT long-only strategy could not outperform the annual return of the benchmark.

The Sharpe ratio distribution (Figure 4) was more symmetric than the annual return distribution, with the outliers nearly equally balanced between the positive and negative territories. A majority of the markets had Sharpe ratios ranging from –0.02 to 0.57, with a mean value of 0.12 and a standard deviation of 0.72. The two bands from 1.17 to 2.09 represent the six markets outperforming the market proxy after transaction costs. However, there were six markets on the opposing end of the distribution with significant negative Sharpe ratios ranging between –3.16 and –0.61. While these six markets were able to achieve excess Sharpe ratios, the distribution of the 65 other markets shows that the norm was risk-adjusted underperformance. This finding was further validated by the p-values of the annual return shown in Tables 2 and 3, indicating significant underperformance relative to the benchmark for every market.

Figure 4 Long-only Sharpe ratio distribution (see online version for colours)



Notes: The x-axis represents the risk-adjusted return measured by the Sharpe ratio. Therefore, the –0.02 displayed on the distribution equals a Sharpe ratio of –0.02. For the y-axis, the number of markets in each Sharpe ratio band was counted, e.g., over 30 markets fell into the –0.02–0.57 band.

Source: Own results

The long-and-short strategy caused significant changes in the Sharpe ratio and annual return matrix shown in Figure 5. The Sharpe ratios were distributed more widely than the long-only results, with most values ranging between –2 and 2. However, unlike the long-only strategy, the outliers Sharpe ratios of the long-and-short strategy ranged between –5.63 and 2.99. While fewer markets outperformed the risk-adjusted return of the market proxy (displayed as a star), the range of negative Sharpe ratios was extended.

Figure 5 Sharpe ratio and annual return of the long-and-short CoT strategy (see online version for colours)

Notes: The blue dots represent the Sharpe ratios for the 71 tested futures markets. The red star represents the Sharpe ratio of the S&P 500 buy-and-hold strategy.

Source: Own results

Thus, none of the markets could outperform the market proxy's annual return under the long-and-short strategy measured by their p-values (Table 4). From a risk-adjusted perspective, the long-and-short backtest showed that COAL and EUROTOP 100 outperformed the benchmark with their Sharpe ratios close to 3 and 1.7. One explanation for these Sharpe ratios is the length of the trading periods. COAL traded for two years while EUROTOP 100 traded for three years, only generating a fraction of the trading signals generated by other indices. However, FEEDER CATTLE, SOYBEAN OIL, and CANOLA with their Sharpe ratios of 1.06, 1.00, and 0.96 followed the performance of the proxy performance closely. Based on the two leading markets, it could also be interpreted that the long-and-short strategy performs better over short timeframes.

The long-and-short CoT strategy's return distribution (Figure 6) was slightly skewed negatively, with 29 markets generating an annual return ranging from –8% to –1%, explaining the mean of –2.3%. The second-highest number of futures fell into the –1% to 5% band. The outliers with positive performance achieved annual returns ranging from 12% to 23%. However, the negative outliers reached annual returns ranging from –33% to –14%. Thus, Figure 6 shows that the long-and-short CoT strategy produced less-volatile results compared to the long-only results. However, with a mean of –2.27% and a standard deviation of 7%, the annual return across markets was dominantly negative under this strategy. The long-and-short annual return distribution also confirmed that the absolute benchmark returns could not be achieved. The EUROTOP 100 stock index recorded the highest annual returns again at 22.91%.

Table 4 Long-and-short backtest highlights

Market names	*P in % for µ1 > µ0*		*Annual return (µ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
S&P500 buy-and-hold	H0		32.00% (µ0)	30.00%	–0.78	8.32	1.07
			µ1				
DOW JONES INDUSTRIAL	0.00%	***	–10.57%	9.61%	–1.00	10.20	–1.11
NASDAQ 100	0.00%	***	7.34%	13.04%	0.55	10.05	0.61
RUSSEL 2000	0.00%	***	–6.96%	13.10%	2.19	31.46	–0.49
S&P 400 MIDCAP	0.00%	***	–0.28%	10.68%	0.84	20.03	0.03
S&P 500 ANNUAL DIVIDEND	50.00%		7.30%	10.00%	12.59	164.63	0.75
EUROTOP 100	24.68%		22.91%	12.60%	–0.96	4.46	1.7
S&P 500	0.00%	***	–5.48%	11.27%	–0.59	10.06	–0.44
VALUE LINE COMPOSITE	0.00%	***	1.24%	4.73%	0.88	27.94	0.28
E-MINI MSCI EAFE	0.00%	***	0.75%	1.43%	0.81	51.07	0.53
NEW YORK HDD	-	-	0.00%	0.00%	0.00	0.00	0
NIKKEI 225	0.00%	***	–2.82%	12.43%	0.05	10.45	–0.17
CANOLA	0.26%	***	3.63%	3.80%	6.20	46.76	0.96
ETHANOL	0.00%	***	–7.41%	10.22%	–4.28	53.08	–0.70
COCOA	0.00%	***	–6.73%	19.19%	–0.85	18.26	–0.27
COFFEE C	0.00%	***	5.88%	16.98%	0.20	14.77	0.42
ORANGE JUICE	0.00%	***	–5.11%	20.79%	–0.40	25.85	–0.15

Notes: The P-values for Table 4 were generated in a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: µ0 > µ1), was tested against each market (H1: µ1 > µ0).
The p-values reflect the probabilities of each market outperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

Source: Own results

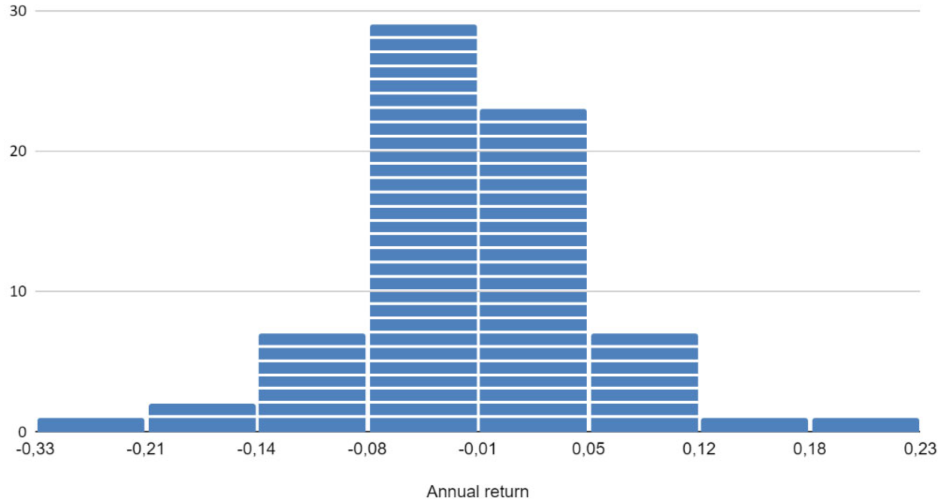
Table 4 Long-and-short backtest highlights (continued)

Market names	*P in % for µ1 > µ0*		*Annual return (µ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
RANDOM LENGTH LUMBER	0.00%	***	–2.50%	12.51%	–3.24	62.10	–0.14
COTTON NO. 2	0.00%	***	2.34%	17.88%	0.50	16.34	0.22
SUGAR NO. 11	0.00%	***	0.35%	19.48%	0.42	19.81	0.12
COAL	0.00%	***	5.98%	1.95%	3.79	18.51	2.99
OATS	0.00%	***	–6.83%	21.08%	–0.43	11.88	–0.23
CORN	0.00%	***	6.58%	18.76%	–0.83	29.02	0.43
ROUGH RICE	0.00%	***	6.54%	15.33%	1.27	27.95	0.49
SOYBEAN MEAL	0.00%	***	–1.37%	24.00%	2.29	46.61	0.06
SOYBEAN OIL	0.00%	***	16.15%	16.30%	0.11	10.96	1
SOYBEANS	0.00%	***	–0.52%	11.94%	–0.18	9.60	0.02
WHEAT	0.00%	***	–0.87%	17.93%	–0.41	22.30	0.04
FROZEN PORK BELLIES	0.00%	***	–2.54%	14.76%	–0.42	105.13	–0.10
FEEDER CATTLE	0.00%	***	7.23%	6.80%	1.91	24.68	1.06
LIVE CATTLE	0.00%	***	–3.81%	9.93%	0.71	14.54	–0.34
LIVE HOGS	0.00%	***	1.86%	15.95%	-2.11	135.98	0.2
BUTTER	0.00%	***	2.45%	12.34%	1.50	51.99	0.26
CHEESE	0.00%	***	–4.44%	15.03%	7.81	145.60	–0.23
MILK	0.00%	***	–11.14%	16.31%	-2.65	67.40	–0.64

Notes: The P-values for Table 4 were generated in a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: $\mu 0 > \mu 1$), was tested against each market (H1: $\mu 1 > \mu 0$).
The p-values reflect the probabilities of each market outperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

Source: Own results

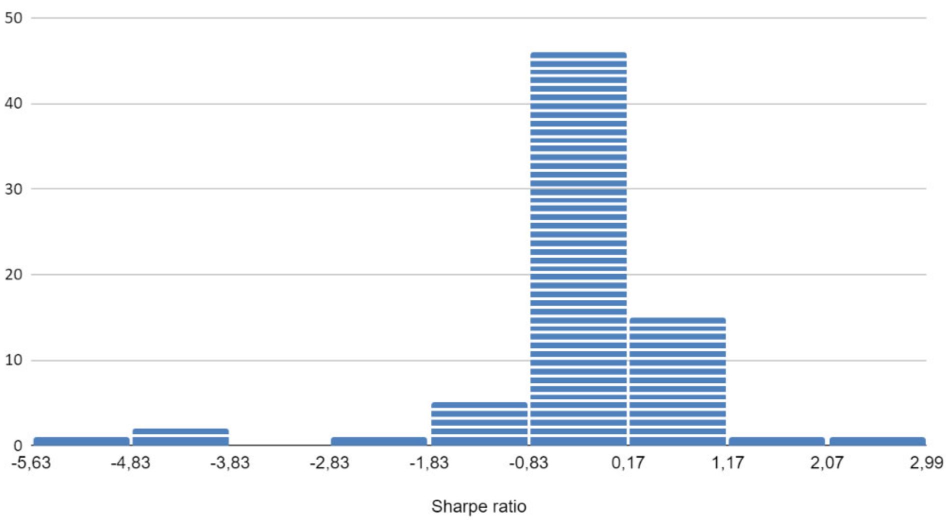
Figure 6 Long-and-short annual return distribution (see online version for colours)



Notes: The x-axis represents the annual return in percent. Therefore, the –0.01 displayed on the distribution is equal to –1% in percentage terms. For the y-axis, the number of markets belonging to each return band was counted, e.g., nearly 30 markets fell into the –8% to –1% band.

Source: Own results

Figure 7 Long-and-short Sharpe ratio distribution (see online version for colours)



Notes: The x-axis represents the risk-adjusted return measured by the Sharpe ratio. Therefore, the 0.17 displayed on the distribution equals a Sharpe ratio of 0.17. On the y-axis, the number of markets that in each Sharpe ratio band were counted, e.g., over 40 markets fell into the –0.83–0.17 band.

Source: Own results

The risk-adjusted returns were more apparent in the Sharpe ratio distribution (Figure 7). The two futures markets that outperformed the market proxy were clearly isolated between 1.17 and 2.99. Comparing this distribution with the long-only performance shows that the distribution is negatively skewed, with negative Sharpe ratios reaching –5.63. The 71 markets reach a combined mean of –0.33, with 46 markets in the –0.83–0.17 band. Similar to the long-only results, the risk-adjusted norm (69 of 71 markets) was verified as being below the benchmark.

In the portfolio scenario, the conditions had to be slightly changed from individual futures markets since the model can trade 71 markets simultaneously. The static portfolio without rebalancing utilised a position size of 1% of the portfolio value for each futures market. The position sizing logic indicates that up to 71% of the portfolio value can be invested at any given time. The results for utilising the CoT strategy in a trading portfolio were evaluated in a long-only and long-and-short environments.

The long-only CoT strategy generated an annual return of 9.92%, with a Sharpe ratio of 0.76 before transaction costs. The net figures for the long-only portfolio were an annual return of 8.88% and a Sharpe ratio of 0.69. For the long-and-short portfolio, the results showed a negative return before and after transaction costs. The figures before transaction costs were a return of -4.77% with a Sharpe ratio of –0.4. The deduction of transaction costs led to an annual return of –5.89% with a Sharpe ratio of –0.51. These results confirm that the static portfolio could not generate excess annual returns or risk-adjusted returns relative to the S&P 500 for the period 1986–2020. The figures discussed are presented in Table 5. The only improvements relative to the benchmark measured were the annual volatility and the maximum drawdown.

Since the portfolio comprised futures markets based on trading signals and not pre-determined weights, it was crucial to identify a suitable portfolio rebalancing strategy. A simple portfolio rebalancing performed at the end of the month or year would be suitable for a buy-and-hold strategy but not a dynamic short-term strategy that uses trading signals. The trades of a particular asset can run into the following year until the exit signal is triggered. In the portfolio rebalancing scenario, the portfolio was rebalanced with every trade, only allowing the CoT strategy to invest 1.3% of the portfolio value into a single asset. For the rebalanced portfolio, annual returns are higher, but Sharpe ratios lower due to higher annual volatility in the rebalanced portfolio. In a long-only portfolio with per-trade rebalancing, the annual return was 12.41% and 10.43% before and after transaction costs, respectively, with Sharpe ratios of 0.64 and 0.59, respectively. A long-and-short approach led to a –3.86% annual return before costs and –5.37% net returns. The Sharpe ratios changed from –0.15 to –0.25 after costs were subtracted. With the annual returns rising slightly and Sharpe ratios declining due to higher annual volatility, neither rebalanced portfolio could outperform the return figures of the S&P 500 for the period 1986–2020. The evaluated data is displayed in Table 6. Similar to the static portfolio scenario, the only enhancements measured relative to the market proxy were the annual volatility and maximum drawdown.

Table 5 Static portfolio backtest results for the CoT strategy

Market names	*P in % for $\mu 1 > \mu 0$*	*Significance*	*Annual return (μ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
S&P 500 buy-and-hold	H0		32.00% (μ0)	30.00%	–0.78	8.32	1.07
Portfolio performance			μ1				
Before transaction costs							
Long-only	0.00%	***	9.92%	13.63%	17.93	467.94	0.76
Long-and-short	0.00%	***	–4.77%	10.68%	1.69	393.07	–0.40
After transaction costs (net)							
Long-only	0.00%	***	8.88%	13.46%	16.4	412.37	0.69
Long-and-short	0.00%	***	–5.89%	10.70%	1.63	388.19	–0.51

Notes: The P-values in Table 5 were generated in a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: μ0 > μ1), was tested against each market (H1: μ1 > μ0). The p-values indicate the probabilities of each market outperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

Source: Own results

Table 6 Rebalanced portfolio backtest results for the CoT strategy

Market names	*P in % for $\mu 1 > \mu 0$*	*Significance*	*Annual return (μ)*	*SD*	*Skewness*	*Kurtosis*	*Sharpe ratio*
S&P500 buy-and-hold	H0		32% (μ0)	30.00%	–0.78	8.32	1.07
Portfolio performance:			μ1				
Before transaction costs							
Long-only	0.00%	***	12.41%	21.38%	19.53	525.23	0.64
Long-and-short	0.00%	***	–3.86%	16.81%	3.85	476.1	–0.15
After transaction costs (net)							
Long-only	0.00%	***	10.43%	19.82%	17.98	502.33	0.59
Long-and-short	0.00%	***	–5.37%	16.80%	4.03	476.68	–0.25

Notes: The P-values in Table 6 were generated in a one-tailed Welch t-test with unequal variances. In the t-tests, the S&P 500 benchmark values were compared to each market. The hypothesis that the S&P benchmark has leading returns, represented by H0 (H0: μ0 > μ1), was tested against each market (H1: μ1 > μ0). The p-values reflect the probabilities of each market outperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively. The skewness and kurtosis refer to the distribution of annual returns over the backtest period. All results are rounded to the second decimal place.

Source: Own results

Over the 34-year period, which matches the backtest period, 100,000 portfolios were simulated. For the market proxy, an annual return of 32% with an SD of 30% was obtained. Both the long-only and the long-and-short portfolios had substantially lower mean values.

The benchmark achieved annual returns ranging between 2% and 62% with a 68% confidence level. There was 95% probability that the annual returns would stay between –28% and 92%. Lastly, it can be estimated that. With a 99.7% probability, the benchmark's annual returns did not decline below –58% or rise above 122%.

The mean of the long-only distribution was close to the former backtest results. The long-only portfolio was estimated to generate a mean annual return of 10% and an SD of 20%. With a CI of 68%, it can be assumed that the portfolio returns remained in the -10% to 30% range. The long-only simulation showed that there was a 0% probability ($p = 0.00$) that the S&P 500 annual return could be outperformed, confirming statistically significant underperformance.

Based on the annual volatility and return of the long-and-short CoT-strategy portfolio, the mean annual return reached –5% with an SD value of 17%. The long-and-short portfolio's annual return distribution showed an even more significant distribution compared with that of the S&P 500. Additionally, since the data confirmed the 0% ($p = 0.00\%$) probability of reaching or exceeding the annual return of the market proxy, it can be concluded that the null hypothesis (H0: $\mu 1$, $\mu 2 < \mu 0$) is correct.

In the analysis of returns obtained during economic distress, the annual returns for 1987, 2000, 2002, 2008, and 2020 were considered in isolation. From Tables 7, 8, and 9 it is evident that trading in a portfolio under the CoT strategy, whether long-only or long-and-short, significantly decreases annual return volatility during periods of economic distress. The market proxy displayed a return low of –34.27% during the great financial crisis and a return high of 14.83% in the year of the COVID pandemic. In contrast, a long-only CoT strategy traded in a portfolio raised the annual return from –1.1% to 6.48% of the total portfolio value. For the long-and-short data, the range increased from –5.81% to 1.6%.

Table 7 S&P 500 buy-and-hold performance during economic crisis

	S&P 500		
	Annual return (µ0)	*SD*	*Sharpe ratio*
1987	–4.91%	26.79%	–0.18
2000	–8.59%	21.80%	–0.39
2001	–9.90%	22.00%	–0.45
2002	–20.73%	20.54%	–1.01
2008	–34.27%	35.25%	–0.97
2020	14.83%	31.54%	0.47

Notes: This table presents the performance of the S&P 500 index futures during the years of market distress. The annual return figures were calculated during the period from 1 January to 31 December of the respective year. The SD values represent the volatility that occurred during the year in terms of the standard deviation. For the Sharpe ratio, the annual returns were divided by the standard deviation values to obtain the risk-adjusted return per unit of risk.

Source: Own results

Table 8 CoT long-only portfolio performance during economic crisis

	Long-only					
	Annual return (µ1)	*SD*	*Sharpe ratio*	*Significance*	*P in % for µ1 > µ0*	*Performance*
1987	0.58%	1.10%	0.53	*	92.48%	Outperforming
2000	2.59%	1.95%	1.33	***	99.97%	Outperforming
2001	–0.44%	1.91%	–0.23	***	99.84%	Outperforming
2002	0.31%	1.45%	0.21	***	100.00%	Outperforming
2008	–1.10%	2.29%	–0.48	***	100.00%	Outperforming
2020	6.48%	17.66%	0.37	**	4.91%	Underperforming

Notes: This table presents the performance of the long-only CoT strategy during the years of market distress. The annual return figures were calculated from the closed trades during the period from 1 January to 31 December of the respective year. The SD values represent the volatility that occurred during the year in terms of the standard deviation. For the Sharpe ratio, the annual returns were divided by the standard deviation values to obtain the risk-adjusted return per unit of risk. The results were generated using a one-tailed Welch t-test with unequal variances. The hypothesis 'P in % for µ1 > µ0' measures the probability of outperformance or underperformance relative to the S&P 500. The p-values in percentage reflect the probabilities of each market outperforming or underperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively.

Source: Own results

Although both portfolios did not beat the S&P 500's average annual return, they showed significant outperformance during stock market crashes as seen in Tables 8 and 9. Based on the annual returns and Sharpe ratios, the results indicate the significant outperformance of the long-only portfolio relative to the S&P 500. Apart from 2020, the long-only portfolio recorded superior returns for every year, obtaining Sharpe ratios of –1.48–1.33. Simultaneously, the S&P 500 had negative Sharpe ratios ranging from –1.01 in 2002 to the highest result of 0.47 during the COVID-19 pandemic year. Although the long-and-short portfolio significantly underperformed in 1987 and 2020 with Sharpe ratios of –0.91 and –0.9, respectively, it exceeded the Sharpe ratios of the S&P 500 during the technology bubble, providing a maximum risk-adjusted return of 0.76 for every unit of risk. While the Sharpe ratio of –1.67 in 2008 suggests underperformance relative to the S&P 500, the p-values for the annual returns confirmed that the CoT long-and-short strategy outperformed the benchmark returns.

Excluding the 2020 figures, for the S&P 500, the negative annual returns increased over time from single digits in 1987, 2000, and 2001 to double digits in 2002 and 2008. Previous studies presented different findings regarding the causes of this increasing volatility across the five market crashes. Zhang (2010) investigated the impact of high-frequency trading (HFT) on the US stock markets. After examining a sample for the period 1985–2009, he found that regulators like the CFTC and the SEC became increasingly worried about the footprint of HFT over time due to the rapid growth rate. They estimate HFT to account for up to 78% of the dollar trading volume (Zhang, 2010). This indicated a 78% growth in HFT activity between 1995 and 2009 (Zhang, 2010). Zhang (2010) confirmed that there is a strong positive correlation between HFT and stock price volatility, especially for the top 3,000 US stocks examined in his study (measured by market capitalisation). Elevated levels of market uncertainty can further amplify the volatility caused by HFT (Zhang, 2010). Based on the above findings and Zhang's (2010) research, it can be suggested that the magnitude of stock market crashes (measured by negative annual return) has increased with the rise of HFT. From no HFT activity in 1987 to around 40% during the tech bubble and almost 80% in 2008, the correlation seems evident (Zhang, 2010). However, to find any significant information in this data, the current research would need to be extended.

Table 9 CoT long-and-short portfolio performance during an economic crisis

	Long-and-short					
	Annual return (μ2)	*SD*	*Sharpe ratio*	*Significance*	*P in % for μ2 > μ0*	*Performance*
1987	–1.30%	1.43%	–0.91	-	82.94%	-
2000	1.60%	2.10%	0.76	***	99.92%	Outperforming
2001	–0.20%	2.06%	–0.10	***	99.87%	Outperforming
2002	0.80%	1.72%	0.46	***	100.00%	Outperforming
2008	–5.56%	3.32%	–1.67	***	100.00%	Outperforming
2020	–5.81%	6.46%	–0.9	***	0.00%	Underperforming

Notes: This table presents the performance of the long-and-short CoT strategy during the years of market distress. The annual return figures were calculated based on closed trades during the period from 1 January to 31 December of the respective year. The SD values represent the volatility that occurred during the year in terms of the standard deviation. For the Sharpe ratio, the annual returns were divided by the standard deviations to obtain the risk-adjusted return per unit of risk. The results were generated using a one-tailed Welch t-test with unequal variances. The hypothesis 'P in % for μ2 > μ0' measures the probability of outperformance or underperformance relative to the S&P 500. The p-values in percentage reflect the probabilities of each market outperforming or underperforming the benchmark with *** ($p < 0.01$) signalling significance at the 1% level relative to the S&P 500. ** ($p < 0.05$) and * ($p < 0.1$) represent significance at the 5% and 10% levels, respectively.

Source: Own results

147 out of the 162 t-tests showed significant underperformance, confirming the null hypothesis (13 t-tests produced statistically insignificant results). From a regulatory perspective, this research confirmed that no excess alpha can be achieved with the CoT data (under the chosen strategy parameters). This indicates that the CoT report promotes market transparency, thus fulfilling the CFTC's goal of achieving fair and efficient derivative markets (Logan and Scott, 2021). Thus, our research shows the effectiveness of the CoT in ensuring efficient markets and that reversal strategies cannot be economically exploited.

Not only the CFTC but other regulatory bodies like European futures exchanges (Eurex, ICE Europe, LME) that produce similar reports could benefit from the conclusions of this study. In contrast, institutional investors and traders (e.g., banks, hedge funds, asset managers) can interpret this data in a different way, wherein CoT reports data alone should not be used to develop live trading strategies.

5 Conclusions

Our research provides novel evidence on the effect of the CoT reports on the transparency and informational efficiency of derivative financial markets. Our study is the first research using a trading strategy backtest based on the CoT data for all monitored US futures markets for the period 1986–2020.

In this study, the CoT reports from inception in 1986–2020 were used to develop a short-term reversal strategy. None of the selected markets or portfolios were able to outperform the S&P 500 benchmark's annual return over the entire backtest period. However, the risk-adjusted returns exceeded the S&P 500 buy-and-hold strategy's Sharpe ratio in various futures markets (after transaction costs). Among the 71 tested markets, six generated superior risk-adjusted returns in a long-only environment, while two generated superior risk-adjusted returns in a long-and-short environment. The long-only CoT strategy had Sharpe ratios ranging from 1.24 to 2.09 as compared with the market proxy, which had a Sharpe ratio of 1.07. Only one stock index outperformed the proxy. However, it should be noted that the EUROTOP 100 and COAL futures were examined over periods of three and two years, respectively, generating a relatively limited number of trades compared to markets that were examined for a period of over 34 years.

The portfolio strategies had a significantly lower annual return volatility than that of the market proxies. Annual return volatility (after transaction costs) dropped to 13.46% and 10.7% under the long-only strategy and long-and-short strategy, respectively, for the portfolio scenario with static 1% asset weights. For the optimised portfolio scenario, annual return volatility (after transaction costs) was slightly higher at 19.82% and 16.8% for the long-only strategy and long-and-short strategy, respectively. However, even with a lower annual volatility, the annual returns and Sharpe ratios of these portfolio scenarios could not outperform the benchmark. The long-only portfolio achieved a highest Sharpe ratio of 0.69 after transaction costs with a net return of 8.88%. The long-and-short portfolio reached a highest Sharpe ratio of –0.25 with an annual return of –5.37%. Rebalancing slightly improved the Sharpe ratio of the long-and-short portfolio from –0.35 to –0.25; however, the same measure in the long-only portfolio reduced its Sharpe ratio from 0.69 to 0.59. The underperformance of these portfolios relative to the S&P 500 buy-and-hold strategy was further validated in a Monte Carlo simulation. These findings confirm that, at $p < 0.01$, both the long-only and the long-and-short portfolios performed below the market proxy's mean. These results amplify the findings of Shpak et al. (2017) as their long-short strategy also turned out to be inferior to a long-only portfolio.

Although the portfolios did not exceed the market proxy in the long-term, they significantly outperformed the S&P 500 benchmark during periods of economic distress, confirming Gurrib's (2009) findings about the objectivity and consistency of market participants as portrayed in the CoT report. In 2000, the long-only portfolio achieved one of the highest Sharpe ratios of 1.33 in this research. Despite the long-and-short portfolio's consistent negative returns over the 34-year research period, its risk-adjusted returns and annual returns significantly outperformed the benchmark in 2000 and 2002. These results lead to the assumption that a short-term reversal strategy using CoT data would perform better than the overall stock market (represented by the S&P 500) during periods of economic turmoil.

This backtest partially confirmed Basu et al.'s (2006) findings. For copper (Sharpe ratio: 0.1) and crude oil (Sharpe ratio: 0.39), no superior returns were observed under the S&P 500 buy-and-hold approach (Sharpe of 1.07), while the S&P 500 CoT strategy's performance (Sharpe of 1.35) confirmed parts of this research (Basu et al., 2006). Excess risk-adjusted returns were found in five additional futures contracts, which were untested by Basu et al. (2006). Additionally, two markets showed superior risk-adjusted returns in a long-and-short scenario. Furthermore, Basu and Stremme (2009) excluded non-commercial data from their research, utilising a non-commercial index to exit trading positions based on the inverse relationship with commercial data (Chatrath et al., 1997). However, the non-commercial index cannot be assumed to be a superior exit signal since the empirical results showed the underperformance of copper and oil relative to Basu and Stremme's (2009) research findings.

In previous literature, the most comprehensive backtest conducted for over 38 markets was extended by 33 markets in this study based on a systematic selection process to obtain insights into all CFTC reportable US futures markets (Briese, 2008). This extension to 71 markets allowed for the addition of discontinued markets to the backtest and the ruling out of the survivorship bias problem present in literature (Briese, 2008; Williams, 2005). Additionally, not only the scope of the investment universe, but also the investment horizon increased. Despite the underperformance of both portfolios under the CoT strategy, the backtest period of 34 years, along with a Monte Carlo simulation, allowed for the estimation of standard deviations in annual returns, further validating the backtest results. In contrast, Merkoulova's research (2020) only explored the period 1986–2017.

While limited assets were observed in the portfolio scenarios studied by Basu et al. (2006), Bhardwaj et al. (2015) and Gorton and Rouwenhorst (2006), extending the investment universe to 71 markets shifted their performance below the selected benchmark. According to Gorton and Rouwenhorst's (2006) diversification study of 27 markets in a portfolio, it can be confirmed that the CoT strategy reduces annual return volatility in a diversified futures portfolio. However, the current study observed no excess risk-adjusted return in portfolios relative to the S&P 500.

In the crises analysis, Gurrib's (2008, 2009) findings were confirmed. Their hypothesis about commercial objectivity and stability was extended from a portfolio of 29 futures markets to 71 futures markets. In line with Gurrib (2008, 2009), we find that portfolio returns during crises show a significantly better risk-adjusted return compared to the stock index.

This study also aimed to provide evidence on whether excess returns in comparison to traditional stock markets can be earned in futures markets adopting a reversal strategy using CoT data. The findings demonstrate that no excess alpha can be achieved in futures markets applying a reversal strategy using CoT data. Thus, our research shows the effectiveness of the CoT report in ensuring efficient markets, fulfilling the CFTC's goal of achieving transparent and efficient derivative trading.

Our findings hold implications for regulators, exchange providers, and market participants. Regulators receive confirmation that excess returns were not achieved during the study period and that transparency through CoT reports played a key role in this. The CFTC should find confirmation that the transaction data they provide makes an important contribution to transparency and thus to fair and efficient trading. The same conclusions can be drawn from other exchange providers who disclose similar trading data. Finally, market participants can draw important conclusions from our results. Trading strategies based on traders' open interest data cannot be used to develop trading strategies for generating excess returns.

Our research is subject to certain limitations that open the opportunity for future research. The backtest only offers results obtained from one specific combination of commercial, non-commercial, and non-reportable index signals. Further research of the utilised hedging pressure indices could be conducted under a trading strategy optimisation scenario. The commercial index values of 0.7 for long signals and 0.3 for short signals could be varied between 0.7 and 1.0 for long signals and 0.0–0.3 for short signals to examine how different hedging pressure levels perform in futures markets or portfolios (vice versa for non-reportable and non-commercial index values). The 50-day or 10-week moving average for price confirmation could be varied to increase or decrease timeframes and analyse changes in the return profile. Furthermore, this backtest was limited to the utilised framework and its functionality (Rodriguez, 2020), and the technologies used were limited in their functionality. Utilising the modern portfolio theory for a Sharpe ratio optimisation based on portfolio asset weights proved to be unfeasible within the research framework. Instead, a simplified rebalancing model had to be employed for portfolio optimisation. This research could be extended to different backtesting frameworks (Python Software Foundation, 2016, 2019; Quantopian Inc., 2020). Moreover, a study of the variations in transaction costs and implied performance could also provide interesting insights. Furthermore, an investigation of the CoT report data reported by European exchanges like LME, Eurex, and ICE Europe provides an opportunity for an interesting extension of our work.

References

Avgouleas, E. and Degiannakis, S. (2009) 'Trade transparency and trading volume: the possible impact of the financial instruments markets directive on the trading volume of EU equity markets', *International Journal of Financial Markets and Derivatives*, Vol. 1, No. 1, pp.96–23, https://doi.org/10.1504/IJFMD.2009.028947.

Barchart (2021) *Barchart Excel*, Barchart.com [online] https://www.barchart.com/excel (accessed 9 December 2022).

Basu, D. and Stremme, A. (2009) 'The economic value of linkages between spot and futures market', *SSRN Electronic Journal*, p.1, https://doi.org/10.2139/ssrn.1340879.

Basu, D., Oomen, R.C.A. and Stremme, A. (2006) 'How to time the commodity market', *SSRN Electronic Journal*, p.1, https://doi.org/10.2139/ssrn.910907.

Bernstein, J. (2012) 'Understanding and using the commitment of traders report: part I and part II', *30 Days to Market Mastery*, pp.133–149. https://doi.org/10.1002/9781119198147.ch16.

Bhardwaj, G., Gorton, G.B. and Rouwenhorst, K.G. (2014) 'Fooling some of the people all of the time: the inefficient performance and persistence of commodity trading advisors', *Review of Financial Studies*, Vol. 27, No. 11, pp.3099–3132, https://doi.org/10.1093/rfs/hhu040.

Bhardwaj, G., Gorton, G.B. and Rouwenhorst, K.G. (2015) 'Facts and fantasies about commodity futures ten years later', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2610772.

Bosch, D. (2017) 'The impact of market participants' interaction on futures prices: comparing three US wheat futures markets', *International Journal of Financial Markets and Derivatives*, Vol. 6, No. 2, pp.120–148, https://doi.org/10.1504/IJFMD.2017.087996.

Boussema, M., Bueno, A. and Séquier, P. (2002) *Transaction Costs and Trading Strategies (Best Execution: Executing Transactions in Securities Markets on behalf of Investors)*, June, European Asset Management Association [online] http://www.afg.asso.fr/wp-content/uploads/2002/06/Fichier116.pdf (accessed 9 December 2022).

Briese, S.E. (1990) 'Commitments of traders as a sentiment indicator', *Technical Analysis of Stocks & Commodities Magazine* [online] http://commitmentsoftraders.org/wp-content/uploads/Static/Perm/cotsentimentindicator.pdf (accessed 9 December 2022).

Briese, S.E. (2008) *The Commitments of Traders Bible: How To Profit from Insider Market Intelligence*, 1st ed., Wiley, Hoboken, New Jersey, USA.

CFTC (2021a) *Explanatory Notes* [online] https://www.cftc.gov/MarketReports/CommitmentsofTraders/ExplanatoryNotes/index.htm (accessed 9 December 2022).

CFTC (2021b) *Commitments of Traders* [online] https://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm (accessed 9 December 2022).

CFTC (2021c) *Disaggregated Explanatory Notes* [online] https://www.cftc.gov/MarketReports/CommitmentsofTraders/DisaggregatedExplanatoryNotes/index.htm (accessed 9 December 2022).

CFTC (2021d) *About the COT Reports* [online] https://www.cftc.gov/MarketReports/CommitmentsofTraders/AbouttheCOTReports/index.htm (accessed 9 December 2022).

CFTC (2021e) *Historical Compressed* [online] https://www.cftc.gov/MarketReports/CommitmentsofTraders/HistoricalCompressed/index.htm (accessed 9 December 2022).

Chang, E.C. (1985) 'Returns to speculators and the theory of normal backwardation', *The Journal of Finance*, Vol. 40, No. 1, pp.193–208, https://doi.org/10.1111/j.1540-6261.1985.tb04944.x,.

Chatrath, A. and Song, F. (1999) 'Futures commitments and commodity price jumps', *The Financial Review*, Vol. 34, No. 3, pp.95–111, https://doi.org/10.1111/j.1540-6288.1999.tb00465.x.

Chatrath, A., Liang, Y. and Song, F. (1997) 'Commitment of traders, basis behavior, and the issue of risk premia in futures markets', *Journal of Futures Markets*, Vol. 17, No. 6, pp.707–731, https://doi.org/10.1002/(sici)1096-9934(199709)17:6.

Cheng, I-H. and Xiong, W. (2013) 'Why do hedgers trade so much?', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2353218.

De Roon, F.A., Nijman, T.E. and Veld, C. (2000) 'Hedging pressure effects in futures markets', *The Journal of Finance*, Vol. 55, No. 3, pp.1437–1456, https://doi.org/10.1111/0022-1082.00253.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'Knowledge discovery and data mining: towards a unifying framework', *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp.82–88, https://dl.acm.org/doi/10.5555/3001460.3001477.

Frenk, D. and Turbeville, W. (2011) 'Commodity index traders and the boom/bust cycle in commodities prices', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.1945570.

Gao, L. (2017) *Commodity Option Implied Volatilities and the Expected Futures Returns*, 12 November, SSRN, https://ssrn.com/abstract=2939649; http://dx.doi.org/10.2139/ssrn.2939649.

Gogas, P. and Serletis, A. (2010) 'Forecast evaluation in daily commodities futures markets', *International Journal of Financial Markets and Derivatives*, Vol. 1, No. 2, pp.155–168, https://doi.org/10.1504/IJFMD.2010.032466.

Gorton, G. and Rouwenhorst, K.G. (2006) 'Facts and fantasies about commodity futures', *Financial Analysts Journal*, Vol. 62, No. 2, pp.47–68, https://doi.org/10.2469/faj.v62.n2.4083.

Gurrib, I. (2008) 'Do large hedgers and speculators react to events? A stability and events analysis', *Applied Financial Economics Letters*, Vol. 4, No. 4, pp.259–267, https://doi.org/10.1080/17446540701720501.

Gurrib, I. (2009) 'Measuring risk for large hedgers and large speculators in major US futures markets', *The Journal of Risk*, Vol. 12, No. 2, pp.79–103, https://doi.org/10.21314/jor.2009.252.

Hammerschmid, R. (2018) *Commodity Return Predictability*, 26 June, SSRN, https://ssrn.com/abstract=2909209; http://dx.doi.org/10.2139/ssrn.2909209.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., … Liphant, T.E. (2020) 'Array programming with NumPy', *Nature*, Vol. 585, No. 7825, pp.357–362, https://doi.org/10.1038/s41586-020-2649-2.

Hayward, R. (2018) 'Foreign exchange speculation: an event study', *International Journal of Financial Studies*, Vol. 6, No. 1, pp.22–35, https://doi.org/10.3390/ijfs6010022.

Humpe, A. and Zakrewski, M. (2015) 'Der CoT-Report als Indikator für Wendepunkte am Aktienmarkt', *Zeitschrift Für Das Gesamte Kreditwesen*, February, Vol. 68, No. 4 [online] https://www.kreditwesen.de/kreditwesen/themenschwerpunkte/aufsaetze/cot-report-indikator-fuer-wendepunkte-aktienmarkt-id4997.html (accessed 9 December 2022).

Jansen, S. (2020) *Machine Learning for Algorithmic Trading: Predictive Models to Extract Signals from Market and Alternative Data for Systematic Trading Strategies with Python*, 2nd ed., Packt Publishing, Birmingham, UK.

Jiler, W.L. (1985) *The Forecasting Methodology*, Commodity Research Bureau Yearbook.

Kang, W., Rouwenhorst, K.G. and Tang, K. (2014) 'The role of hedgers and speculators in liquidity provision to commodity futures markets', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2449315.

Keynes, J.M., Johnson, E. and Moggridge, D. (2012) *The Collected Writings of John Maynard Keynes (Volume 5)*, 2nd ed., Cambridge University Press, Cambridge, UK.

Lien, D. and Xiang, J. (2010) 'Convergence to efficiency in FTSE-100 futures market', *International Journal of Financial Markets and Derivatives*, Vol. 1, No. 3, pp.243–257, https://doi.org/10.1504/IJFMD.2010.034237.

Logan, M. and Scott, M.P. (2021) *Commodity Futures Trading Commission (CFTC)*, September, Investopedia [online] https://www.investopedia.com/terms/c/cftc.asp (accessed 9 December 2022).

Maul, D., Fischer, M. and Schiereck, D. (2015) 'Spekulation am Terminmarkt und die Preisentwicklung von Agrarrohstoffen am Kassamarkt: Eine Zeitreihenanalyse der CFTC Berichte für Weizen, Mais und Sojabohnen' [Speculation in futures markets and the impact on agricultural commodity prices: a time series analysis of the CFTC reports for wheat, corn and soybeans], *Jahrbücher für Nationalökonomie und Statistik*, Vol. 235, No. 6, pp.608–629, https://doi.org/10.1515/jbnst-2015-0606.

Maverick, J.B. (2020) *What to Know About the 50-Day Simple Moving Average*, 4 May, Investopedia [online] https://www.investopedia.com/ask/answers/012815/why-50-simple-moving-average-sma-so-common-traders-and-analysts.asp (accessed 9 December 2022).

Merkoulova, Y. (2020) 'Predictive abilities of speculators in energy markets', *Journal of Futures Markets*, Vol. 40, No. 5, pp.804–815, https://doi.org/10.1002/fut.22058.

Moran, N.M., Irwin, S.H. and Garcia, P. (2020) 'Who wins and who loses? Trader returns and risk premiums in agricultural futures markets', *Applied Economic Perspectives and Policy*, Vol. 42, No. 4, pp.611–652, https://doi.org/10.1002/aepp.13048.

Python Software Foundation (2016) *zipline*, 25 February, PyPI [online] https://pypi.org/project/zipline/0.8.4/ (accessed 9 December 2022).

Python Software Foundation (2019) *pyfolio*, 15 April, PyPI [online] https://pypi.org/project/pyfolio/ (accessed 9 December 2022).

Quantopian Inc. (2020) *quantopian/pyfolio*, 15 July, Quantopian.Github.Io/Pyfolio [online] https://github.com/quantopian/pyfolio (accessed 9 December 2022).

Rodriguez, D. (2020) *Cerebro – Backtrader* [online] https://www.backtrader.com/docu/cerebro/ (accessed 9 December 2022).

Rodriguez, D. (2021a) *Broker – Backtrader* [online] https://www.backtrader.com/docu/broker/ (accessed 9 December 2022).

Rodriguez, D. (2021b) *Commission Schemes – Backtrader* [online] https://www.backtrader.com/docu/commission-schemes/commission-schemes/ (accessed 9 December 2022).

Rodriguez, D. (2021c) *Introduction – Backtrader* [online] https://www.backtrader.com/docu/ (accessed 9 December 2022).

Rodriguez, D. (2021d) *Orders – General – Backtrader* [online] https://www.backtrader.com/docu/order/ (accessed 9 December 2022).

Rodriguez, D. (2021e) *Orders – Target Orders – Backtrader* [online] https://www.backtrader.com/docu/order_target/order_target/ (accessed 9 December 2022).

Sanders, D.R., Irwin, S.H. and Merrin, R.P. (2009) 'Smart money: the forecasting ability of CFTC large traders in agricultural futures markets', *Journal of Agricultural and Resource Economics*, Vol. 34, No. 2, pp.276–296, https://doi.org/10.22004/ag.econ.54547.

SciPy Community (2021) *numpy.random.normal – NumPy v1.20 Manual* [online] https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html (accessed 9 December 2022).

Sharma, A. (2019). *Monte Carlo Simulation in Finance Python Part-2*, 25 September, Medium [online] https://medium.com/@dstrader/monte-carlo-simulation-in-finance-python-part-2-13eb7bd89220 (accessed 9 December 2022).

Shpak, I., Human, B. and Nardon, A. (2017) *Idiosyncratic Momentum in Commodity Futures*, 11 September, SSRN, https://ssrn.com/abstract=3035397; http://dx.doi.org/10.2139/ssrn.3035397.

Smales, L. A. (2022) 'Trading behavior in agricultural commodity futures around the 52-week high', *Commodities*, Vol. 1, pp.3–17, https://doi.org/10.3390/commodities1010002.

Smales, L.A. (2013) 'Trading behaviour in closely related markets for S&P 500 index futures', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2292256.

Tessmann, M.S., Ely, R.A. and Canever, M.D. (2021) 'Volatility transmissions between commodity futures contracts in short, medium and long term', *International Journal of Financial Markets and Derivatives*, Vol. 8, No. 1, pp.79–99, https://doi.org/10.1504/IJFMD.2021.113870.

Upperman, F. (2012) *Commitments of Traders: Strategies for Tracking the Market and Trading Profitably*, 1st ed., Wiley, https://doi.org/10.1002/9781119201328.

Williams, L.R. (2005) *Trade Stocks and Commodities with the Insiders*, 1st ed., Wiley.

Zhang, F. (2010) 'High-frequency trading, stock volatility, and price discovery', *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.1691679.