Wisdom of Crowds and Commodity Pricing

John Hua FAN[a*], Sebastian BINNEWIES[b] and Sanuri De SILVA[a]

[a] *Department of AFE, Griffith Business School, Griffith University*

[b] *School of Information and Communication Technology, Griffith University*

August 2022

Abstract

We extract commodity-level sentiment from the Twittersphere in 2009-2020. A long-short systematic strategy based on sentiment shifts more than doubles the Sharpe ratio of extant commodity factors. The sentiment premium is unrelated to fundamentals but is exposed negatively to basis risk and is more pronounced during periods of macro contraction and deteriorating funding liquidity. Sentiment-induced mispricing is asymmetric, i.e., commodities with low (high) sentiment shifts tend to be overvalued (undervalued) when the aggregate market is in backwardation (contango). Furthermore, the observed premium arises almost entirely from commodities with the most retweet activities, while retweets and likes themselves do not exhibit stronger predictive ability compared to non-influential tweets.

JEL Classification: C55, G13, G14, G41

Keywords: commodity futures; social media; sentiment; textual analysis; behavioral

* Corresponding author. 170 Kessels Road, Nathan, Queensland 4111, Australia. Tel: +61 7 3735 4638; E-mail: j.fan@griffith.edu.au (J. Fan).

This research is supported by Griffith University Research Ethics Clearance (2021/654). We thank seminar participants at Griffith University for comments and suggestions.

1. Introduction

Previous studies have uncovered a large number of factors that influence the pricing of commodity futures.[1] Almost all of these factors arise from the two pillars of commodity pricing literature that pertains to the shape of commodity futures term structure (i.e., backwardation/contango), other factors are anchored on inefficiencies in the way markets incorporate information into prices (Bianchi et al., 2016), and more recently investor's active attention to hazard fear (Fernandez-Perez et al., 2020) and media emotion (Chi et al., 2022). In this paper, we introduce a novel form of systematic risk – *sentiment* and document its influence on commodity prices beyond the extant commodity fundamentals.

The literature on investor sentiment is extensive. In their seminal work, Baker and Wurgler (2006) formally established the role of investor sentiment in explaining the cross-section of stock returns, highlighting two channels (i.e., limits-to-arbitrage and speculative demand) for which sentiment can impact prices. Accordingly, the effect of sentiment on returns is more pronounced among stocks that are more susceptible to speculation and are relatively more difficult to trade. Since sentiment is not directly observable, it must be proxied.[2]

Whilst the *general* market sentiment as measured by Baker and Wurgler can affect returns and volatilities across asset classes including commodities, the literature is silent on whether sentiment is a priced risk within commodity futures markets. How can sentiment impact commodity returns? As one of the few studies that focus exclusively on commodities,

[1] For example, inventory levels (Gorton et al., 2013), roll-yield (Erb & Harvey, 2006), hedging pressure (Basu & Miffre, 2013), past returns (Gorton & Rouwenhorst, 2006; Miffre & Rallis, 2007), skewness (Fernandez-Perez et al., 2018; Fuertes et al., 2022), liquidity (Szymanowska et al., 2014), basis-momentum (Boons & Prado, 2019) and relative-basis (Gu et al., 2021).

[2] For example, value-weighted dividend premium, first-day returns on IPOs, IPO volume, odd-lot transactions, the closed-end fund discount, the equity share in new issues (Baker and Wurgler, 2006), fund flows (Ben-Rephael et al., 2012), survey (Greenwood and Shleifer, 2014), VIX and SKEW indices (Gao and Süss, 2015), market breadth (Zhou, 2018), news (Heston and Sinha, 2017), social media (Giannini et al., 2017; Agrawal et al., 2018), pictures (Obaid and Pukthuanthong, 2022), and music (Edmans et al., 2022).

Gao and Süss (2015) conclude that market sentiment explains the excess-comovement of commodity futures. To avoid capturing dynamics related to a commodity sentiment risk, they deliberately designed a sentiment index to be exogenous to commodity markets using non return based measures in stocks. They also stress the importance of speculative demand and arbitrage constraints as two conditions for sentiment to affect commodity returns beyond fundamentals. Although investing in commodity futures is not subject to short-sell constraints, arbitrage activities are still constrained by the aggregate liquidity in the capital market (Acharya et al., 2013). Meanwhile, increased presence of speculators has had lasting impact on commodity markets (Basak and Pavlova, 2016). Therefore, given that commodity markets are often thought to be less prone to retail trading frenzies, the existence of sentiment-induced mispricing raises an intriguing question concerning the behavior of large commodity traders.[3]

Unlike Gao and Süss (2015), we measure commodity-specific sentiment from user-generated tweets, thus our sentiment measure is endogenous to commodity markets. By harvesting the collective "wisdom of the crowd", we investigate whether investor emotions play a role in the cross-sectional pricing of commodity futures. To this end, the extant literature on social media sentiment lends little help as most studies focus on short-term dynamics. Azar and Lo (2016) find that tweets contain predictive content on FOMC meeting days for the CRSP value-weighted index. Similarly, Kargozoglu and Fabozzi (2017) show that increase in daily social anomaly score predicts an increase in market volatility. More in line with our setup, Chen et al. (2014) extract investor views and opinions from "SeekingApha.com" and find that the fraction of negative words negatively predicts stock returns and earnings surprise over the following three months, but they do not test the investment performance of such information.

[3] Using CFTC's Commitment of Traders report, Bhardwaj et al. (2016 exhibit 3) estimate that the "non-reportables" category of traders (i.e., small/retail) accounts for less than 20% of the total open interest held, whereas producers/merchant/user, money managers, and swap dealers account for more than 70% of the total open interests.

We find that commodities with higher changes in sentiment systematically outperform commodities with lower changes in sentiment. While the *level* of optimism/pessimism is also important, its predictive ability is markedly weaker compared to the shifts in sentiment (as well as sentiment fluctuations). From 2010 to 2020, an equal-weighed and monthly rebalanced long-short portfolio of high versus low sentiment shifts commodities generates statistically significant mean returns of 7.2% per annum. Such a naïve strategy reports a Sharpe ratio of 0.75 and a maximum drawdown of 12.2% which more than doubles the performance while halves the downside risk compared to conventional factors. The sentiment premium is unrelated to momentum, hedging pressure, skewness and basis-momentum, but is exposed negatively to the basis risk. Consistent with time-series spanning tests, cross-sectional pricing tests highlight the significant pricing ability of sentiment beyond commodity fundamentals.

To identify the sources of this predictability, we consider 15 macro conditions both endogenous and exogenous to commodity markets including general market sentiments, commodity market cycles, business cycles and economic conditions, shipping and policy and geopolitical uncertainties. We find that sentiment-induced mispricing is asymmetric across commodity market cycles, i.e., commodities with higher changes in sentiment tend to be oversold when the aggregate market is in contango/less backwarded, whereas commodities with lower changes in sentiment tend to be overbought when the aggregate market is in backwardation/less contangoed. Furthermore, consistent with the literature on limits-to-arbitrage (Stambaugh et al. 2012; Acharya et al., 2013; Chu et al., 2020), we find that sentiment induced mispricing is most pronounced in contractionary macro environments and periods of deteriorating funding liquidity, and predominately comes from the short leg of the portfolio.

Having isolated the drivers of the sentiment factor returns, we examine if the predictive ability of sentiment stems from some but not all tweets. If the sentiment induced mispricing is a result of a sub-group of influential users, we expect to observe a higher (lower) average return when sentiment is measured using only tweets with non-zero (zero) retweets and likes. We failed to find statistically significant differences, suggesting that the predictive ability of sentiment depends on the collective wisdom of the crowd rather than a select group of high attention tweets. Furthermore, if the mispricing is induced by sentiment shifts of the crowd, we should observe a stronger effect in commodities that exhibit higher tweeter activities to begin with. Indeed, we find that the profitability of sentiment factor almost entirely comes from commodities with the most retweet activities, cementing the fact that sentiment does influence the pricing of commodity futures. However, consistent with Ballinari and Behrendt (2021), the choice of lexica for sentiment extraction is crucial for the detection of mispricing.

Our approach is informative to both the commodity futures and sentiment literature. First, we proxy commodity-level sentiment through activities in the Twittersphere. Unlike other proxies, tweets are a more direct measure of sentiment because they reflect opinions, moods, and emotions of the user, rather than an "attention" proxied through google search volume (Da et al., 2011; Han et al., 2017; Fernandez-Perez et al., 2020) or news (Smales, 2014; Omura and Todorova, 2019) which may be difficult to classify or interpret. Indeed, Chi et al. (2022) find that the emotion intensity of news articles conveys additional information beyond news sentiment and coverage. Unlike these studies, we derive sentiment by determining the semantic orientation of *each tweet* relevant to each commodity, without having to infer from search results or news. By "polling" the opinions of the crowd, our approach mitigates the potential biases and ambiguities faced by other proxies, resulting in a "cleaner" channel for sentiment to impact returns. Second, unlike the stock market where retail participation is at all-time highs, it is relatively more difficult to speculate in the futures market.[4] Large commodity

[4] See CNBC's recap of GameStop, Reddit and Robinhood at https://www.cnbc.com/2021/01/30/gamestop-reddit-and-robinhood-a-full-recap-of-the-historic-retail-trading-mania-on-wall-street.html. Although ETFs on crude oil and gold are readily available in retail trading platforms, access to direct futures trading requires higher

hedgers or speculators who are presumably more sophisticated may not be searching for answers but instead be expressing their views via twitter. These tweets encompass comments on news articles related to a particular commodity/sector as well as predictions on future price movements. Therefore, our twitter-based commodity sentiment is potentially less "noisy" compared to those extracted for individual stocks.

Contrary to the belief that the predictive ability of social media sentiment is only non-negligible over the short-term (1 to 2 days), our findings suggest that social media sentiment can have a lasting impact on commodity returns. The literature predominately assumes commodity markets behave according to rational expectations. However, even though commodity markets are less affected by retail trading noise, evidence on the sentiment premium suggests that large commodity hedgers and speculators who are presumably more sophisticated than retail investors also exhibit speculative behaviors that deviate from rational expectations. This is remarkably consistent with the recent findings of DeVault et al. (2019) that sentiment captures institutional rather than individual investors demand shocks in stocks. The behavioral inefficiencies of commodity traders are under appreciated in the literature.

The sentiment mispricing can be grounded in the appraisal-tendency framework (ATF) in psychology (Lerner et al., 2015). The ATF predicts that once an emotion is generated, it might lead to a cognitive propensity to evaluate future occurrences based on the primary appraisal qualities that produced the feeling. ATF posits that emotions predispose people to evaluate the world in certain ways that lead to comparable functional outcomes. Therefore, when markets are in contango, traders are trigged by the crowd to hold biased views about future price depreciation, leading them to over-sell commodities that exhibit higher changes in

prerequisites (i.e. minimum capital requirements and experience). Furthermore, Mao et al. (2012) estimate an average of 9,434 S&P500 related tweets per day, which is approximately 37.9 million in our sample period, compared to a mere 3.5 million unique tweets for *all* major commodities.

sentiment. Similarly, when markets are in backwardation, traders are trigged to hold biased views about future price appreciation, leading them to over-buy commodities that exhibit lower changes in sentiment. Subsequently, the correction of this systematic mispricing results in a significant return spread between high versus low sentiment shift commodities.

Our findings are robust to transactions costs, alternative sentiment signals, seasonality, and a suite of implementation concerns. We demonstrate that excluding the 20% least liquid commodities, employing a rank-weight scheme in portfolio construction, tilting more aggressively, and focusing on the performance over the last 5 years do not significantly impact the risk adjusted performance and downside risk profile of the sentiment strategy. As a placebo test, we find that google search volume index does not contain information content in the same setting. Taken together, the sentiment factor is readily implementable and should be of immediate interests to hedge funds and CTAs either as an overlay or as a standalone absolute return strategy.

The remainder of the paper proceeds as follows. Section 2 describes the collection and processing of commodity futures and twitter data. Section 3 outlines the main results, asset pricing tests and sources of sentiment profits. In Section 4, we examine whether the wisdom is coming from the crowd of twitter users or a select group of influential tweets, followed by a suite of robustness tests in Section 5. The paper concludes in Section 6.

2. Commodity sentiment

2.1. Futures data

For ease of replication, we employ commodity futures data from the Blomberg Commodities Index (BCOM) family. From January 2009 to December 2020, we download daily open, high, low, and settlement prices on a broad sample of 28 commodities across six sectors. Namely industrial metals (aluminum; copper; lead; nickel; tin; zinc); energy (Brent crude; gas oil; ULS Diesel; natural gas; unleaded gas; WTI crude oil); grains (corn; soybean meal; soybean oil; soybeans; wheat); livestock (feed cattle; lean hogs; live cattle); precious metals (gold; platinum; silver); and softs (cocoa; coffee; cotton; orange juice; sugar). To ensure roll-yields are correctly accounted for, we compute commodity returns using the excess return index values.

Formerly known as the DJ-UBS index, BCOM adopts a "gradual rolling" approach in which the weights of positions in the expiring contact are gradually increased to the next futures contract from the 6th to 10th day of the preceding month to maturity. This may be preferred by large investors to cushion the adverse price impact from rolling large futures positions (see Bianchi et al. 2015 for other advantages). To supplement the returns data, we also obtain settlement prices on the first three nearest contracts. The selection of commodities is broadly consistent with the extant literature (e.g. Basu & Miffre, 2013; Boons & Prado, 2019). Our final sample is dictated by the availability of twitter data. Twitter was official launched on 21 March 2006; however, due to limited activities in the earlier years, our final sample covers January 2009-December 2020. All futures data are sourced from the Bloomberg terminal.

Table 1 reports the summary statistics. Over the sample period, the mean excess returns vary widely across commodities and time. The highest mean returns are observed in soybean meal and the lowest in diesel. On average, a long position in an equally weighted portfolio of all commodities losses 2% p.a. which is far below the long-term average of 4.5% reported by Bhardwaj et al. (2021). This is due largely to the structural downturns experienced by the asset class post-GFC and the "de-financialization" of commodities post-2014 (see Bianchi et al., 2020), which also partially explains the underperformance of commodity factors.

2.2. Sentiment extraction

From 1 January 2009 to 31 December 2020, we obtain via the Twitter academic API all tweets matching commodity keywords outlined in Table 1, along with user id, likes and retweet counts. Due to privacy reasons, the quality of data limits our ability to meaningfully analyze tweet locations. This results in a total of approximately 416 million tweets and retweets. We use a lexicon-based approach to extract commodity-level sentiment. This method mitigates the overfitting problem encountered in other machine learning methods. Because a sentiment lexicon focuses on lexical features and labels text according to its semantic orientation, the accuracy of a lexicon depends on its ability to define sentiment in specific contexts. For this reason, we anchor our sentiment extraction on the Loughran and McDonald (2011, L&M thereafter) financial lexicon.

Before implementing sentiment analysis, we first take the following procedures to standardize and remove noise from the tweets.[5] We remove Twitter URLs (e.g., t.co/xyz), general URLs (e.g., https://abc.com....), hash tags (e.g., #adani), handle tags (e.g., @realmessi), and replace internet slang with word equivalents (e.g., n00b -> newbie), replace word elongations (e.g., whyyyy -> why), ordinal numbers with words (e.g., 1st -> first), non-ASCII characters with a text representation, contractions with both words (e.g., I'll -> I will), control characters (e.g., tab, line break) with a space, and replace multiple spaces with a single space.

The L&M lexicon assigns a word, *not* a sentence or tweet, to one of the following categories: “positive”, “negative”, “litigious”, “uncertain”, “constraining”, “superfluous” and none. For each word in each tweet, we assign a value of 1 to positive words, -0.15 to negative words and a value of 0 for words classified as neither positive nor negative. Given the

[5] The cleaning and pre-processing packages used are: Rinker, T. W. (2017). qdapRegex: Regular Expression Removal, Extraction, and Replacement Tools. 0.7.2. University of Buffalo. Buffalo, New York. http://github.com/trinker/qdapRegex. Rinker, T. W. (2018). textclean: Text Cleaning Tools version 0.9.3. Buffalo, New York. https://github.com/trinker/textclean

overwhelming number of negative words (2355) compared to positive words (354) in the L&M dictionary, we set the negative value to -0.15 (instead of -1) to obtain a more balanced sentiment measure. More formally, let $S(w)$ be the sentiment score of word w, so that:

$$S(w) = \begin{cases} +1 \; if \; w \; is \; a \; positive \; word \\ 0 \; if \; w \; is \; a \; neutral \; word \\ -0.15 \; if \; w \; is \; a \; negative \; word \end{cases} \tag{1}$$

Let further $W(v, i)$ denotes the set of words contained in tweet v related to commodity i after pre-processing, $V(i, d)$ denotes the set of tweets related to commodity i posted on date d, and $sgn(\cdot)$ is the signum function. Thus, the sentiment for commodity i on day d is given as:

$$Sentiment_{i,d} = \frac{\sum_{v \in V(i,d)} sgn(\sum_{w \in W(v,i)} S(w))}{|V(i,d)|} \tag{2}$$

Lastly, let $n(t)$ be the number of days in a month t. The sentiment for commodity i in month t is given as:

$$Sentiment_{i,t} = \frac{\sum_{d \in t} Sentiment_{i,d}}{n(t)} \tag{3}$$

Table 1 reports the average sentiment for each commodity over the sample period along with the total number of unique tweets, users and total number of retweets and likes. We observe large sentiment variations in the cross-section, with dispersion of 9.8 times the average. Not surprisingly, Feeder cattle, RBOB and diesel report the lowest number of activities whereas crude and gold report the highest activities in terms of users, retweets and likes. Owing to their economic significance, WTI Crude and Gold alone account for nearly half of all tweets, suggesting that commodities receive vastly different attention on social media.

Figure 1 illustrates commodity sentiments across sectors along with the nearest futures contract prices. The plots indicate that sentiments vary considerably from commodity to commodity, though the change in sentiment seemingly tends to lead near term futures price movements. Furthermore, Figure 2 plots the intra-sector correlations of sentiment in comparison to returns. We draw two key observations. First, consistent with Figure 1, on average, sentiment scores do vary from one sector to another. The largest intra-sector sentiment dispersion is observed in energy and the lowest in precious metals. Second, in contrast to returns which often comove considerable within sectors, sentiment does vary from commodity to commodity even within the same sector, confirming the observation in Figure 1. Overall, Figures 1 and 2 suggest that sentiment behaves rather differently to returns. We formally test this relationship in the following section.

2.3. Sentiment and commodity returns

To establish the link between commodity sentiment and futures returns, we estimate the following regression model:

$$r_{i,t} = \beta_i Sentiment_{i,t} + [u_i] + \varepsilon_{i,t} \tag{4}$$

where $r_{i,t}$ and $Sentiment_{i,t}$ are the excess returns and sentiment of commodity i at time *t*, respectively. $[u_i]$ represents a vector of control variables that incorporates trading volume and core term structure fundamentals such as basis, momentum, basis-momentum and relative-basis. We include fixed effects on year, month and commodity to control for the passive predictability across commodity markets and time.

Table 2 reports the results of the contemporaneous regression. We first run the baseline model without controls in models (1) and (2) to examine the relationship of $Sentiment_{i,t}$and $\Delta Sentiment_{i,t}$ with excess returns, where $\Delta Sentiment_{i,t} = Sentiment_{i,t} - Sentiment_{i,t-1}$. We find a strong and positive relationship at the 1% significance level. Previous studies have documented the association of sentiment with trading volume (Duz Tan and Tas, 2021), prompting us to question whether the link between sentiment and returns manifests itself in volumes traded. After controlling for volume, basis and past returns, the results in models (3) and (4) remain largely consistent. In models (5) through (8), we include the fix effects and find that the relationship is significant at the 5% level. Finally in models (9) and (10) we include the recently developed basis-momentum and relative-basis characteristics and find that the results continue to hold.

Overall, the findings presented in Table 2 suggest that commodity sentiments proxied by Twitter activities conveys additional information beyond known commodity fundamentals. Therefore, we move on to test the predictive ability of sentiment on commodity futures returns out-of-sample.

3. Is sentiment priced?

3.1. Performance of the sentiment portfolio

At the last trading day of each month, we sort the cross-section of 28 commodities by $\Delta Sentiment_{i,t}$ into two portfolios i.e., high versus low.[6] As a baseline setup, we do not apply any smoothing, scaling or optimization to modify the signal or the asset weights. We simply take long positions in commodities within the high group and short positions in the low group. The long-short portfolio is weighted equally, rebalanced monthly and 50% collateralized. We follow the same procedure to construct carry, momentum, hedging pressure, skewness, basis-momentum and relative-basis factors (see Appendix I for details).

Table 3 reports the performance of the sentiment strategy along with traditional commodity factor strategies in Panel A and long-only commodity exposures in Panel B. The findings suggest that systematically buying (selling) commodities with improved (deteriorated) sentiment generate statistically significant average returns of 7.2% p.a. with a Sharpe ratio of

[6] Tweeter Academic API records tweets according to Greenwich Mean Time (GMT), which 1) does not match with the closing times for US-based exchanges; and 2) is available each calendar day, thereby creating a look-ahead bias especially when examining sentiment and returns over the short-run. Unlike these studies, we focus on medium to long-term sentiment. Thus, we do not expect this time difference to drive our results after aggregation. Nevertheless, to address the potential concern, no future commodity sentiments are considered when rebalancing takes place. Besides, we take a more aggressive approach by eliminating the sentiment data beyond the 25th calendar day in each month. We find that excluding the last several days do not impact the results.

0.75. This more than doubles the performance of conventional factors which on average reported a Sharpe ratio of 0.30 over the sample period. The sentiment strategy also exhibits the lowest volatility and VaR, the highest percentage of positive months and CER, and a mere 12.2% maximum drawdown (i.e. more than three times lower compared to the average).[7] The strong outperformance of the sentiment strategy stands in sharp contrast to the underperformance of key commodity factors documented in the literature. Not surprisingly, the sentiment strategy dominates all long-only commodity exposures amid the structural downturn in the asset class over the last decade. Overall, if one interprets sentiment as investors' mood in commodity futures markets, the findings in Table 3 suggest that commodities with high mood swings systematically outperforms commodities with lower mood swings.

Figure 3 illustrates the cumulative performance (Panel A) and drawdown (Panel B) of the sentiment strategy along with conventional factors and the average market portfolio (AVG). In contrast to conventional factors, the sentiment strategy has consistently maintained wealth accumulation over the entire sample period with minimal drawdowns. Notably, AVG, carry(basis) and momentum are among the factors which experienced catastrophic drawdowns reminiscent of commodity benchmarks S&P-GSCI and the BCOM in this period.

3.2. Time-series spanning tests

The preceding section reveals that the sentiment strategy captures attractive mean excess returns in commodity futures, in this section we proceed to test whether the sentiment premium is exposed to pervasive commodity risk factors previously documented.

[7] The certainty-equivalent-return (CER) is computed as $CER = \left(\frac{n}{T}\right)\sum_{t=0}^{T-1}\frac{(1+r_{P,t+1})^{1-\gamma}-1}{1-\gamma}$, where $r_{P,t+1}$ represents the return at time $t+1$, n is the total number of trading days in a year and γ, the relative risk aversion parameter, is set to 5. The CER can be interpreted as the return an investor is willing to accept now rather than taking a chance on a higher, but uncertain, return in the future.

Table 4 reports the results of time-series spanning tests. Using the Bakshi et al. (2019) three-factor model, we first examine in model (1) whether returns to the sentiment strategy can be explained by the market (AVG), basis and momentum factors. While the alpha is economically and statistically significant (6.9% p.a. with a t-stats of 2.45), we observe a negative loading on basis (-0.19), suggesting that sentiment's outperformance stems at least partially from betting against the term structure. From models (2) to (5), we augment the benchmark model by adding one additional risk factor at a time, and Model (6) incorporates all factors. We find that the sentiment factor is unrelated to momentum, hedging pressure, skewness, relative-basis and basis-momentum factors, but remains to be negatively exposed to the basis factor.[8] After accounting for commodity-specific risk factors, the sentiment strategy reports an average alpha of 6.8% p.a. with negligible R-squares. We further investigate sentiment's link with the commodity term structure in later sections.

3.3. Cross-sectional pricing tests

The preceding section reveals that sentiment factor cannot be spanned by existing risk factors, in this section we examine whether sentiment commands a risk premium cross-sectionally above extant risk factors.

Following Fernandez-Perez et al. (2020), we first estimate the full-sample betas via OLS time-series regressions:

$$r_{i,t} = \alpha_i + \beta_i F_t + \varepsilon_{i,t}, t = 1, \ldots., T \tag{5}$$

where $r_{i,t}$ is the time *t* excess returns of (*a*) quintile portfolios based on $\Delta Sentiment_{i,t}$, (*b*) the quintile portfolios based on the six characteristics listed in Appendix I, and (c) the equally weighted and monthly rebalanced portfolios from the 6 commodity sectors reported in Table

[8] Since our sample includes LME listed commodities that do not have CFTC positions data, models (2) and (6) may be misspecified. To address this concern, we exclude all commodities traded on the LME, and re-run the same models with newly constructed sentiment strategy and risk factors. The alpha (6.1%) remains significant at 5% and the loadings on hedging pressure remain insignificant, albeit that the loadings on Basis are weakened.

3. Thus, we have N=46 test assets altogether. F_t includes the sentiment factor as well as the 6 systematic risk factors that can potentially price the cross-section of returns. In step two, we estimate the following cross-sectional regression of average excess returns on full-sample betas obtained under step-one

$$\bar{r}_i = \lambda_0 + \lambda\hat{\beta}_i + \varepsilon_i, i = 1, \dots, N \quad (6)$$

where λ is a vector containing the prices of risk associated with each of the factors.

Table 5 reports the results of the cross-sectional pricing tests. Similar to time-series spanning tests, we first employ the benchmark three-factor model then introduce additional factors one at a time. Model (1) first shows that sentiment is priced without the influence of other risk factors, and in model (2) we find that basis and momentum command negative prices of risks, whereas in model (3) sentiment remains to be priced in the presence of AVG, basis and momentum, although none of the extant factors report statistically significant lambdas after Shanken correction. In models (4) to (11), we find that sentiment remains priced in the presence of skewness, relative-basis and basis-momentum factors, but losses its significance when hedging pressure was considered in isolation to other factors. In models (12) and (13), we show that sentiment continues to show significance in the presence of all factors. On average across models, the sentiment strategy commands a risk premium of 9.9% p.a., with *t*-statistics ranging from 1.48 to 2.28. Therefore, the pricing ability of sentiment cannot be fully rationalized by commodity fundamentals. Moreover, across all models, including sentiment improves the model fitness and MAPE monotonically. Overall, the findings presented in Table 5 suggest that sentiment is a non-negligible driver of commodity futures returns.

3.4. Macro drivers

Having established that sentiment shifts are an important factor in commodity markets and that commodity fundamentals cannot fully explain its excess returns, we proceed to investigate the sources of the mispricing. To do so, we consider five classes of macro drivers that can potentially explain the return spreads between high versus low sentiment commodities. These variables are drawn from a large number of studies that advocate the role of market sentiment and states (Copper et al., 2004; Fernandez-Perez et al., 2020), business cycle and macroeconomic conditions (Gorton and Rouwenhorst, 2006; Shang et al., 2016; Levine et al., 2018; Cotter et al., 2020), funding liquidity (Asness et al., 2013), shipping cost (Bakshi et al., 2011), policy uncertainty (Brogaard and Detzel, 2015) and geopolitical risks (Baur and Smales, 2020; Cheng et al., 2022) in understanding commodity and factor returns.

Table 6 reports the returns of long leg, short leg and long-short portfolios with 15 state variables that are either endogenous or exogenous to commodity markets and trade. Panel A focuses on general market sentiments proxied by the VIX index, Baker and Wuglar (2006) sentiment, and the 24-month returns of the S&P500. Panel B considers aggregate commodity market states measured by the average market wide basis and hedging pressure (which both proxy for the shape of the commodity term structure), and the 24-month return of the S&P-GSCI. In Panels C and D, we consider business cycle and economic conditions such as yield curve shifts, term spread, TED spread, G7 GDP and inflation, and the Chicago Fed National Activity. Finally in Panel E, we consider shipping cost, economic policy and geopolitical uncertainties proxied by the Baltic Dry, EPU and GPR indices, respectively.[9] The High versus Low state is divided by the full-sample median value in each respective state variable.

In Panel A, we find the long-short sentiment factor appears to be stronger when the equity market is less volatile, less optimistic, and performing better. However, the average returns of sentiment during these states are statistically indifferent from their respective counter

[9] The S&P500, S&P/GSCI and BDI data are sourced from the Bloomberg terminal. B&W is from Jeffrey Wurgler's NYU site. Data on yield curve changes, term spread, TED spread, GDP, inflation, and CFNAI are downloaded from the St. Louis Fed. EPU is from https://www.policyuncertainty.com/ and GPR is from https://www.matteoiacoviello.com/gpr.htm.

states, suggesting that general market sentiment and states cannot explain the excess return of the commodity sentiment factor. This result holds when examining the long/short legs in isolation. In Panel B, consistent with the results on equity market states, none of the commodity market states can fully account the long-short spread between high vs. low sentiment commodities. Unsurprisingly, the average returns are higher when the aggregate market is more backwardeated/less contangoed. Interestingly however, we find that the short (long) leg of the sentiment strategy is significantly stronger when the aggregate commodity market is more backwarded/less contangoed (contangoed/less backwardeated), suggesting that sentiment is a much stronger predictor of commodity return than basis itself. This finding implies that sentiment-induced mispricing is asymmetric to the commodity term structure. i.e., low sentiment shift commodities tend to be overbought when (and only when) the aggregate market is in backwardation, and high sentiment shifts commodities tend to be oversold when (and only when) the aggregate market is in contango. This also explains why the long-short hedged sentiment portfolio does not vary significantly across commodity term structures. Since sentiment-induced mispricing is more pronounced during specific commodity market states, a double-sorted strategy that is long high sentiment commodities with low basis and short low sentiment commodities with high basis can potentially outperform the baseline sentiment strategy.

Turning to Panel C, we find that market liquidity plays an important role in determining the returns spread between high and low sentiment commodities, as the sentiment factor performs drastically better when the yield curve shifts upwards, and when funding liquidity tightens. Put alternatively, the outperformance of the sentiment strategy predominately originates from periods when the market is more concerned about funding liquidity and economic growth. In line with Fernandez-Perez et al. (2020), these findings suggest that limits-to-arbitrage also induces sentiment-based mispricing in commodity futures markets. In Panel

D, we find that even though global GDP and inflation cannot explain the return spread, the sentiment factor tends to perform better when the global economy has grown stronger, albeit that the differences with the low state are not statistically significant. In contrast, we find that sentiment profits tend to be marginally stronger when a composite of US economic activities is below its long-term average. Finally, we find that shipping cost, economic and geopolitical uncertainties cannot adequately explain the return spread between high vs low sentiment commodities, although in Panel E, sentiment tends to perform better when the EPU and GPR are lower than their long-term averages.

Overall, the findings presented in Table 6 suggest that sentiment-induced mispricing is asymmetric across commodity market cycles, i.e., commodities with higher sentiment tend to be oversold when the aggregate market is in contango/less backwarded, whereas commodities with lower sentiment cross-sectionally tend to be overbought when the aggregate market is in backwardation/less contangoed. Furthermore, limits-to-arbitrage plays a crucial role in determining the return spread between high vs. low sentiment commodities. These findings imply that it may be possible to improve its risk-adjusted performance of the commodity sentiment strategy.

4. Wisdom of the crowd or the few?

4.1. Influential tweets

Having established the potential source of the mispricing documented in the previous section, we now proceed to disentangle the effects within our database of tweets. In our main test, commodity sentiment is measured using all tweets. Some of these tweets may contain likes and/or are retweeted while others will have neither. We believe including retweets are important because each retweet represents an additional expression by the user. Given the large differences between users, retweets and likes in Table 1, we seek to distinguish the information content of tweets with zero versus non-zero retweet and likes. By definition, a tweet with zero retweet and/or likes receives less attention compared to a tweet with one or more retweet and/or likes. In light of recent predatory trading activities observed in nickel and silver markets[10], we are compelled to investigate whether the predictive ability of sentiment shifts truly originates from the tweeter crowd or a select group of influential tweets proxied by likes and retweets.

Table 7 reports the performance of the sentiment strategy using sentiment extracted from tweets with non-zero versus zero tweets. If the sentiment induced mispricing is a result of influential tweets, we expect to observe a stronger (or similar) performance when commodity sentiment is measured using only tweet with non-zero retweets and/or likes compared to using all tweets. Meanwhile, we should observe a weakened average return on the contrary. These strategies report largely insignificant or weakly significant returns when considering retweets, likes, or jointly. Neither zero nor non-zero tweets generate stronger performance compared to our main results in Table 3. Moreover, difference-in-mean tests reveal that whether sentiment is measured based on high or low attention tweets does not differ statistically. Overall, these findings suggest the predictive ability of sentiment depends on the collective wisdom of the crowd rather than a select group of users. This finding also implies, at least in a cross-sectional setting, that the pricing influence of high attention users/tweets is rather limited.[11]

[10] Silver: https://www.cnbc.com/2021/01/31/silver-futures-jump-7percent-as-reddit-traders-try-their-squeeze-play-with-the-metal.html. Nickel: https://www.bloomberg.com/news/articles/2022-03-14/inside-nickel-s-short-squeeze-how-price-surges-halted-lme-trading

[11] We cannot rule out the possibility that influential tweets (i.e., with non-zero retweet or likes) may have "inspired" more tweets that do not show the same amount of attention (i.e. with zero retweet or likes). Nevertheless, the causal relationship between influential and non-influential tweets is beyond the scope of the current work, but it may be of regulatory interests.

4.2. Mispricing intensity

If the mispricing is indeed induced by sentiment shifts of the crowd, we should observe a stronger effect in commodities that exhibit higher tweeter activities to begin with. To test this hypothesis, we divide our sample into two groups based on the *level* of activities each month. Commodities with the 50% most activities are in the high group and commodities with the 50% least activities are in the low group. It is important to note that this test is different from Section 4.1. In Table 7, we re-calculate sentiment scores based on a sub-group of *tweets* (i.e. zero or non-zero retweet/likes), whereas in Table 8 we employ all tweets but reevaluate the performance of sentiment strategy in sub-group of *commodities* (i.e. high or low total activities).

Table 8 reports the performance of the sentiment strategy in sub-group of commodities with the most and least amount of tweeter activities classified by tweets, retweets and likes. The volatility is markedly higher compared to those in Table 7, since each specification only contains half of the cross-section. In line with our conjecture, we find that the sentiment strategy indeed reports higher average returns in the high activities group, suggesting that the mispricing is stronger in commodity markets that attract more attention on Twitter. The strongest effects are observed when the level of activity is proxied by retweets. The Sharpe ratio (0.65) is the closest to the main results 0.75, whereas the least active group reports insignificant returns. This is corroborated by the difference-in-mean tests (p=0.01), suggesting that sentiment-induced mispricing is concentrated in commodities that receive relatively more attention. We observe similar albeit weaker results when the level of activities is proxied by unique tweets and likes. This is somewhat expected because retweets arguably are a superior proxy for emotion because it generally takes more for someone to retweet a post than liking it. Overall, the findings in Table 8 offers reassurance that a sentiment-induced mispricing is prevalent in the commodity futures market.

4.3. Alternative lexica

If the "wisdom" is indeed in the crowd, and that the mispricing is most pronounced in markets with higher level of tweeter activities, we now ask the question whether lexica matter in arriving at our results. Loughran and McDonald (2011) stress that the "right" context has a significant impact on the semantic orientation of texts, and hence the inference. For this reason, we have employed the financial dictionary developed by L&M in the main analysis. In this section, we examine whether the dictionary matters in the context of commodity markets. We consider four alternative lexicons widely used in the computer science literature, including Bing (Hu and Liu, 2004), NRC (Mohammad et al., 2013), TextBlob (textblob.readthedocs.io) and AFINN (Nielsen, 2011). In contrast to the finance-targeted L&M lexicon, these lexicons were constructed through annotations of general purpose (Bing and NRC), movie review, and tweet datasets, respectively.[12]

Table 9 reports the performance of alternative lexicons (specifications (2) to (5)) in comparison to the main results (specification (1)). While all alternative lexicons report average returns indifferent from zero, their return distributions and downside risk profiles vary markedly. Ballinari and Behrendt (2021) find that finance-specific dictionaries perform well compared to alternative approaches in equity pricing. Our findings suggest that the context has a significant impact on the inference. Overall, consistent with Ballinari and Behrendt (2021), the findings presented in Table 9 suggests that the choice of lexica is critical in capturing sentiment-induced mispricing in commodity futures.

[12] Numerous comparisons between these lexicons have been conducted. See for example as part of the SemEval competitions: https://semeval.github.io/

5. Robustness

5.1. Friction

Although sentiment profits are unlikely to be eliminated by transaction costs, because like traditional commodity factors, the sentiment strategy only requires monthly rebalancing, it is still informative to examine to what extend are the profits eroded by transaction costs. If the portfolio turnover is exceedingly high, its practical application is limited even if the paper portfolio outperforms other commodity factors.

Figure 4 illustrates the average and break-even portfolio turnovers (Panel A) and Sharpe ratios before and after trading costs (Panel B). The turnover is calculated by measuring the difference between a commodity's actual weight at the end of preceding month and the target weight in the following month aggregated across the portfolio and time. Consistent with the previous observation that sentiment signals tend to vary across markets and time, the sentiment factor indeed exhibits the highest level of turnover per month (1.3x compared to an average of 0.4x by other factors). This naturally raises the question: is the excess return driven by transactions cost?

To address this concern, two tests are considered. First, we plot the break-even turnover (TO). This is the portfolio TO required to generate a zero net return. Since commodity factor strategies typically rebalances once a month, the maximum turnover in theory should not exceed 2x (assuming no additional leverages). The break-even TO therefore can be interpreted as a capacity gauge for the activeness of the strategy. Clearly, the sentiment factor exhibits the highest break-even TO (2.2x higher than the average TO). In contrast, momentum strategy will report a negative net return after accounting for transaction cost, as the break-even TO is merely 0.1x. Secondly, in Panel B we compare the after-cost Sharpe ratio (assuming a round-trip transactions cost of 0.086% each based on the turnovers computed in Panel A) with the after-cost Sharpe ratios.[13] While the impact of transactions cost on the sentiment factor is non-negligeable, the after-cost Sharpe ratio is still among the highest even compared to the gross Sharpe ratios across factors. Overall, the findings presented in Figure 4 suggest that transaction costs alone cannot explain the outperformance of the sentiment factor.

5.2. Alternative signal

To ensure the observed sentiment effect is not a statistical fluke, we examine alternative sentiment signals. We first employ the level ($Sentiment_{i,t}$) instead of the change ($\Delta Sentiment_{i,t}$) in sentiment and then introduce shocks to the main sentiment signal over a longer look-back window. This test is motivated by the following. First, Basu and Miffre (2013) measure hedging pressure and basis over the past 52 weeks, they do so because a smoothed signal is potentially less noisy compared to a shorter, unsmoothed signal. Second, Kang et al., (2020) find that a longer-term hedging pressure signal in fact conveys a different type of risk premium compared to the short-term hedging pressure.

Table 10 reports the performance of sentiment strategies based on alternative sentiment signals. Panel A reports the results when $Sentiment_{i,t}$ is measured as per Eq(3), whereas Panel B reports the results when $Sentiment_{i,t}$ is computed as the standard deviation within a month. Specifications (1) and (6) are based on the one-month level signals, whereas specifications (2) and (7) are based on the changes. Specifications (3) to (5) and (7) to (10) are based on the change in sentiment demeaned from the last 6- to 12-month average. The demeaned (DMA) $\Delta Sentiment_{i,t}$ signals can be viewed as shocks to sentiment shifts over the medium to long terms. While still significant at the 10% level, the results in Panel A clearly suggest that sentiment itself does not exhibit strong predictive ability for the next period return, rather it is

[13] Refer to Marshall et al. (2012) and Lock and Venkatesh (1997) for transaction costs estimates in commodity futures markets.

the variations in sentiment that contain informational content. Thus, it is natural to ask whether a longer-term sentiment variation leads to improved performance?

The results from specification (3) to (5) monotonically suggest that a longer-term signal can indeed further improve the Sharpe ratio of the sentiment strategy, suggesting that a longer-term window may reduce the prediction error. The best performance is obtained when the look-back window is 10 months, with a Sharpe of 0.93, average return of 8.89% (t=3.6) at the same level of MaxDD with the original signal. Whether the modified signal generates statistically better performance is beyond the scope of the current paper, though the average correlation of the modified strategies with the original is 0.83. Turning to Panel B, we observe a monotonic performance reduction when commodity sentiment is measured by standard deviation instead of the mean, suggesting that the variation in *average* sentiment better captures the mispricing than variations in sentiment *volatility*. This may be explained by the fact that sentiment volatility captures the degree of *bidirectional* fluctuations between optimism and pessimism, whereas the average sentiment measures the level of sentiment. Therefore, a change in average sentiment more accurately conveys a directional shift in market sentiment. Albeit much weaker than in Panel A, the findings presented in Panel B remind us that the return spread between high and low sentiment shift commodities is unlikely because of data mining.

5.3. Seasonality

Commodities are known to exhibit seasonal patterns (Sørensen, 2002; Back et al., 2013; Keloharju et al., 2016). In this section we examine whether the profitability of the sentiment factor can be explained by calendar-based seasonalities.

Figure 5 illustrates the average returns of the sentiment strategy in each calendar month (Panel A) and calendar year (Panel B) compared to the broad market returns (AVG). It can be clearly seen that the average returns tend to fluctuate from one month to another, although the strategy tends to exhibit the highest (lowest) average returns in August (November). Turning to the calendar year in Panel B, the sentiment factor reports positive average monthly returns in every calendar year in the sample period except for 2010. Clearly, the sentiment factor has not only managed to avoid the losses in every instance when the board market is down, but also generated significant positive returns. This outperformance far exceeds the underperformance (4 out of 11 years) when the broad market is up.

Overall, the findings presented in Figure 5 suggests that the average returns of the sentiment factor are not skewed by commodity seasonality effects or dominated by one or more commodity market cycles. In other words, sentiment-induced mispricing is unlikely a product of calendar-based seasonality effect.

5.4. Placebo test

To gain further robustness, we employ a placebo test to ascertain our observed results are not due to random chance. To structure the placebo test, we proxy commodity sentiment by google search volume index.[14] For each commodity keyword, we obtain the monthly search index from Google and interpret the change in search intensity ($\Delta GoogleTrends_{i,t}$) as the sorting signal. We then follow the same portfolio construction rule and take long positions in high and short positions in low groups. We do not expect GoogleTrends-based "sentiment" to exhibit the same level of predictive ability compared to our lexicon-based sentiment extracted via millions of tweets. Because unlike opinion lexicon, GoogleTrends on commodity search terms merely represent "attention", which says nothing about the semantic orientation of a given keyword. Of course, one can combine semantic search terms with commodity names to extract skewed search activities (see e.g., Fernandez-Perez et al., 2020).

[14] We obtain Google Trends data via https://trends.google.com/trends/ using commodity keywords listed in Table 1. The search parameters are set to "Worldwide", "2009-2020", "Finance", and "Web Search".

Table 11 reports the results of the placebo test. In specification (1), we find that the average return of the alternative strategy is largely zero, and negatively skewed with a negative CER. To gain further robustness, we also follow the same procedure to capture short to medium term (3 to 12 months) shocks in $\Delta GoogleTrends_{i,t}$ in specifications (2) to (5). None of the tests report statistically significant average returns. The annualized returns on average across all setups are 1% with a Sharpe ratio of 0.1, suggesting that search attention does not inform the cross-sectional pricing of commodity futures.

Overall, the results in Table 11 suggest that our sentiment signal extracted in this study is non-trivial. i.e., commodity sentiment proxied by Twitter activities contain important information content for the pricing of commodity futures.

5.5. Investability

Having established the robustness of the sentiment signal and strategy, we proceed to consider the investability of the sentiment strategy.

Table 12 reports the results. Since the size of our commodity cross-section is large, one may question whether the average return of the sentiment strategy is primarily earned through relatively illiquid commodities that do not have the same depth as other major commodities. In specification (1), we re-examine the performance of the sentiment strategy by excluding the bottom 20% least liquid commodities. The risk-return profile remains largely consistent, but notably the MaxDD has deteriorated due largely to the diminishing diversification effects. In specification (2), we assess whether the observed performance is sensitive to how the portfolio is weighted. By employing a rank-weight scheme of Koijen et al., (2018) in portfolio construction, the percentage of positive month has improved slightly at the expense of risk-adjusted performance.

In specifications (3) and (4), we widen the spread between long and short positions by focusing on more extreme breakpoints of commodities. Indeed, a more aggressive tilt (e.g., Q3) elevates the average mean return at the expense of portfolio downside risks. Notably, an extreme tilt (Q5) further reduces the risk adjusted performance considerably, suggesting that the success of the sentiment strategy is due at least partially to the diversification effects. This also explains why the rank-weighted portfolio does not improve the risk-adjusted performance of the sentiment factor. Lastly, to mitigate the concern that sentiment mispricing is concentrated in the first half of the sample, we examine its performance over the last 5 years, and found that the Sharpe ratio is highly consistent with the full-sample result, and that the MaxDD occurred in the earlier sample period. Taken together, the findings in Table 12 suggest that the sentiment factor is readily implementable and should be of immediate interests to hedge funds and CTAs either as an overlay or as a standalone absolute return strategy.

6. Conclusion

This paper examined the pricing ability of sentiment in the board commodity futures markets. To extract commodity-level sentiment, we first obtained all tweets relating to each commodity on the Twitter microblogging platform and then applied a lexicon-based approach to gauge the semantic orientation of each tweet. We find that sentiment shifts are priced in both the time-series and cross-section in commodity futures. Despite the structural bear market in commodities as an asset class over the last decade, a systematic strategy that takes long positions in commodities with high sentiment swings and short positions in commodities with low sentiment swings generate statistically significant average returns of 7.2% p.a. with a Sharpe ratio and a maximum drawdown which more than doubles the performance while halves the downside risk to extant commodity risk factors on average. The sentiment premium is unrelated to momentum, hedging pressure, skewness, basis-momentum and relative-basis but is negatively exposed to basis risk.

To rationalize the sentiment premium, we associate its returns to commodity market states, general market sentiment, business cycles, macroeconomic environments, shipping cost and political risks, and found that sentiment-induced mispricing is more pronounced during contractionary environments and periods of deteriorating funding liquidity. Furthermore, we demonstrated that sentiment premium arise from the overvaluation of commodities with low sentiment shifts when the aggregate market is in backwardation, and the undervaluation of commodities with high sentiment shifts when the aggregate market is in contango. To shed further light in our understanding of sentiment, we examined the role of retweets and likes. Influential tweets do not exhibit stronger predictive ability over tweets with zero retweets and likes, suggesting that the wisdom is indeed coming from the crowd. Our findings on sentiment premium implies that large commodity hedgers and speculators who are presumably more sophisticated and rational than retail investors can also exhibit behaviors that deviate from rational expectations. The behavioral inefficiencies of commodity traders are underappreciated in the literature and could present an interesting avenue for future research.

Finally, we demonstrated that sentiment-induced mispricing almost entirely comes from commodities that exhibit the highest amount of retweet activities. Since retweet is a superior emotional proxy compared to tweets and likes, this solidifies the fact that the sentiment premium is indeed driven by the wisdom of the crowd. Nevertheless, we also found that the choice of lexica plays a pivotal role in extracting the information content of tweets. Finally, as placebo test, we failed to find predictive ability using google trends, which does not distinguish the semantic orientation of search terms. Our findings are robust to a battery of tests including turnover and transaction costs, seasonality, alternative formation periods, portfolio construction methods and investability concerns. Taken together, the sentiment factor is readily implementable and should be of immediate interests to hedge funds and CTAs either as an overlay or as a standalone absolute return strategy.

Reference

Acharya, V. V., Lochstoer, L. A., & Ramadorai, T. (2013). Limits to arbitrage and hedging: Evidence from commodity markets. *Journal of Financial Economics*, *109*(2), 441-465.

Agrawal, S., Azar, P. D., Lo, A. W., & Singh, T. (2018). Momentum, mean-reversion, and social media: Evidence from stocktwits and twitter. *The Journal of Portfolio Management*, *44*(7), 85-95.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929-985.

Azar, P. D., & Lo, A. W. (2016). The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds. *The Journal of Portfolio Management*, *42*(5), 123-134.

Back, J., Prokopczuk, M., & Rudolf, M. (2013). Seasonality and the valuation of commodity options. *Journal of Banking & Finance*, *37*(2), 273-290.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645-1680.

Bakshi, G., Gao, X., & Rossi, A. G. (2019). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, *65*(2), 619-641.

Ballinari, D., & Behrendt, S. (2021). How to gauge investor behavior? A comparison of online investor sentiment measures. *Digital Finance*, *3*(2), 169-204.

Basak, S., & Pavlova, A. (2016). A model of financialization of commodities. *The Journal of Finance*, *71*(4), 1511-1556.

Basu, D., & Miffre, J. (2013). Capturing the risk premium of commodity futures: The role of hedging pressure. *Journal of Banking and Finance*, *37*(7), 2652-2664.

Baur, D. G., & Smales, L. A. (2020). Hedging geopolitical risk with precious metals. *Journal of Banking & Finance*, *117*, 105823.

Ben-Rephael, A., Kandel, S., & Wohl, A. (2012). Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics*, *104*(2), 363-382.

Bhardwaj, G., Gorton, G. B., & Rouwenhorst, K. G. (2016). Investor interest and the returns to commodity investing. *The Journal of Portfolio Management*, *42*(3), 44-55.

Bhardwaj, G., Janardanan, R., & Rouwenhorst, K. G. (2021). The first commodity futures index of 1933. *Journal of Commodity Markets*, *23*, 100157.

Bianchi, R. J., Drew, M. E., & Fan, J. H. (2015). Combining momentum with reversal in commodity futures. *Journal of Banking & Finance*, *59*, 423-444.

Bianchi, R. J., Drew, M. E., & Fan, J. H. (2016). Commodities momentum: A behavioral perspective. *Journal of Banking & Finance*, *72*, 133-150.

Bianchi, R. J., Fan, J. H., & Todorova, N. (2020). Financialization and de-financialization of commodity futures: A quantile regression approach. *International Review of Financial Analysis*, *68*, 101451.

Boons, M., & Prado, M. P. (2019). Basis-Momentum. *The Journal of Finance*, *74*(1), 239-279.

Brogaard, J., & Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, *61*(1), 3-18.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367-1403.

Cheng, D., Liao, Y., & Pan, Z. (2022). The Pricing of Geopolitical Risk in Cross-Sectional Commodity Returns. *Working paper Macquarie University*.

Chi, Y., El-Jahel, L., & Vu, T. (2022). Hidden in the spotlight: media emotion intensity and commodity futures pricing. *Working paper University of Auckland.*

Chu, Y., Hirshleifer, D., & Ma, L. (2020). The causal effect of limits to arbitrage on asset pricing anomalies. *The Journal of Finance*, *75*(5), 2631-2672.

Cooper, M. J., Gutierrez Jr, R. C., & Hameed, A. (2004). Market states and momentum. *The Journal of Finance*, *59*(3), 1345-1365.

Cotter, J., Eyiah-Donkor, E., & Potì, V. (2020). Commodity futures return predictability and intertemporal asset pricing. *Michael J. Brennan Irish Finance Working Paper Series Research Paper*, (20-8).

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461-1499.

DeVault, L., Sias, R., & Starks, L. (2019). Sentiment metrics and investor demand. *The Journal of Finance*, *74*(2), 985-1024.

Duz Tan, S., & Tas, O. (2021). Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance*, *22*(2), 221-234.

Edmans, A., Fernandez-Perez, A., Garel, A., & Indriawan, I. (2022). Music sentiment and stock returns around the world. *Journal of Financial Economics*, forthcoming.

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, *62*(2), 69-97.

Fernandez-Perez, A., Frijns, B., Fuertes, A.-M., & Miffre, J. (2018). The skewness of commodity futures returns. *Journal of Banking and Finance, 86*, 143-158.

Fernandez-Perez, A., Fuertes, A. M., Gonzalez-Fernandez, M., & Miffre, J. (2020). Fear of hazards in commodity futures markets. *Journal of Banking & Finance*, *119*, 105902.

Fuertes, A. M., Liu, Z., & Tang, W. (2022). Risk-neutral skewness and commodity futures pricing. *Journal of Futures Markets*, *42*(4), 751-785.

Gao, L., & Süss, S. (2015). Market sentiment in commodity futures returns. *Journal of Empirical Finance*, *33*, 84-103.

Giannini, R., Irvine, P., & Shu, T. (2018). Nonlocal disadvantage: An examination of social media sentiment. *The Review of Asset Pricing Studies*, *8*(2), 293-336.

Gorton, G. B., Hayashi, F., and Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance, 17*(1), 35-105.

Gorton, G., and Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47-68.

Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, *27*(3), 714-746.

Gu, M., Kang, W., Lou, D., and Tang, K. (2021). Relative basis and risk premia in commodity futures markets. *Working paper available at SSRN 3404561*.

Han, L., Li, Z., & Yin, L. (2017). The effects of investor attention on commodity futures markets. *Journal of Futures Markets*, *37*(10), 1031-1049.

Heston, S. L., & Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, *73*(3), 67-83.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).

Kang, W., Rouwenhorst, K. G., & Tang, K. (2020). A tale of two premiums: the role of hedgers and speculators in commodity futures markets. *The Journal of Finance*, *75*(1), 377-417.

Karagozoglu, A. K., & Fabozzi, F. J. (2017). Volatility wisdom of social media crowds. *The Journal of Portfolio Management*, *43*(2), 136-151.

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2016). Return seasonalities. *The Journal of Finance*, *71*(4), 1557-1590.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*(2), 197-225.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799-823.

Levine, A., Ooi, Y. H., Richardson, M., & Sasseville, C. (2018). Commodities for the long run. *Financial Analysts Journal*, *74*(2), 55-68.

Locke, P. R., and Venkatesh, P. (1997). Futures market transaction costs. *Journal of Futures Markets*, *17*(2), 229-245.

Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance*, *66*(1), 35–65.

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. In *Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research* (pp. 69-72).

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2012). Commodity liquidity measurement and transaction costs. *The Review of Financial Studies, 25*(2), 599-638.

Miffre, J., and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking and Finance, 31*(6), 1863-1886.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242.*

Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703-708.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, *144*(1), 273-297.

Omura, A., & Todorova, N. (2019). The quantile dependence of commodity futures markets on news sentiment. *Journal of Futures Markets,* 39(7), 818-837.

Shang, H., Yuan, P., & Huang, L. (2016). Macroeconomic factors and the cross-section of commodity futures returns. *International Review of Economics & Finance, 45,* 316-332.

Shanken, J. (1992). On the estimation of beta-pricing models. *The Review of Financial Studies*, *5*(1), 1-33.

Smales, L. A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, *49*, 275-286.

Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, *22*(5), 393-426.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288-302.

Szymanowska, M., De Roon, F., Nijman, T., & Van Den Goorbergh, R. (2014). An anatomy of commodity futures risk premia. *The Journal of Finance*, *69*(1), 453-482.

Zhou, G. (2018). Measuring investor sentiment. *Annual Review of Financial Economics*, *10*, 239-259.

Appendix I Factor construction

Factor	Commodity-specific signals	Definition at the time of portfolio formation *t*	References
Momentum	$MOM_t = \prod_{s=0}^{11}(1 + r_{t-s}) - 1$	r_t denotes the time *t* monthly excess return of the front-end contract.	Miffre and Rallis (2007)
Carry	$Basis_t = \frac{F_t^1}{F_t^2} - 1$	F_t^1 and F_t^2 denote the prices of the nearest and 2nd nearest contract at time *t*, respectively.	Erb and Harvey (2006); Gorton and Rouwenhorst (2006); Koijen et al. (2018)
Hedging Pressure	$HP_t = \frac{1}{52}\sum_{w=0}^{51}\frac{S_{t-w} - L_{t-w}}{S_{t-w} + L_{t-w}}$	S_t and L_t denote the week *t* short and long positions of a given commodity as held by commercial traders in the CFTC report, respectively.	Bessembinder (1992); De Roon et al. (2000); Basu and Miffre (2013)
Skewness	$SKEW_t = \frac{\frac{1}{D_1}\sum_{d=0}^{D_1-1}(r_{t-d} - \mu_t)^3}{\sigma_t^3}$	r_d denotes the daily excess return of the front-end contract at time *d*, μ_t and σ_t denote mean and standard deviation of daily excess returns as measured at time *t* using daily data over the past year and D_1 is the number of days in the past one year.	Fernandez-Perez et al. (2018)
Basis-Momentum	$BM_t = \prod_{s=0}^{11}(1 + r_{t-s}^1) - \prod_{s=0}^{11}(1 + r_{t-s}^2)$	r_t^1 (r_t^2) represents the time *t* monthly excess return of the front (second-nearest) contract.	Boons and Prado (2019)
Relative-Basis	$RB_t = \ln\left(\frac{F_t^1}{F_t^2}\right) - \ln\left(\frac{F_t^2}{F_t^3}\right)$	F_t^m denotes the time *t* price of the *m*th nearest contract, T_t^m represents the time to maturity of the *m*th nearest contract expressed in number of days at time *t*.	Gu et al. (2022)

Table 1 Summary statistics

This table reports the mean and standard deviation (Std Dev) of commodity returns, sentiment, along with the associated number of tweets, retweets, likes and users. Index ticker represents the Bloomberg Commodity Excess Return Index for the respective commodity whereas commodity ticker represents the commodity ticker on the respective exchange. The sample period covers January 2009-December 2020.

Commodity	Commodity	Index	Commodity		Excess return		Sentiment		Total #	# Unique	Total #	Total #
Sector	Name	Ticker	Ticker	Exchanges	Mean	Std Dev	Mean	Std Dev	Tweets	users	Retweets	Likes
Industrial Metals	Aluminium	BCOMAL	LA	LME	-0.1%	5.6%	-0.023	0.135	18,586	4,612	110,860	8,272
	Copper	BCOMHG	HG	COMEX	0.7%	7.5%	0.014	0.067	87,922	18,882	348,791	22,842
	Lead	BCOMPB	LD	LME	0.9%	8.4%	0.076	0.112	122,668	72,706	23,400,996	127,609
	Nickel	BCOMNI	LN	LME	0.8%	10.2%	0.042	0.149	34,393	25,707	144,216	21,780
	Tin	BCOMSN	LT	LME	0.8%	7.1%	-0.007	0.196	5,845	3,287	36,803	4,975
	Zinc	BCOMZS	LX	LME	0.4%	7.6%	0.040	0.148	20,989	5,703	74,826	14,198
Energy	Brent Crude	BCOMCO	CO	ICE(US)	0.6%	9.3%	-0.002	0.134	135,301	24,896	1,113,543	50,378
	Gas Oil	BCOMGO	QS	ICE(US)	0.5%	9.2%	0.015	0.093	100,432	32,676	1,903,466	57,855
	ULS Diesel	BCOMHO	HO	ICE (US)	-1.5%	13.5%	0.033	0.079	890	288	547	228
	Natural Gas	BCOMNG	NG	NYMEX	0.4%	9.0%	0.037	0.308	125,071	25,707	861,227	37,763
	Unleaded Gas	BCOMRB	RBOB	ICE (US)	0.8%	10.5%	0.048	0.291	368	153	45	77
	WTI Crude Oil	BCOMCL	CL	NYMEX	0.1%	10.4%	-0.020	0.091	58,159	27,305	147,000,000	55,010
Grains	Corn	BCOMCN	C	CBOT	-0.4%	7.9%	-0.022	0.063	189,843	40,263	5,048,327	72,500
	Soybean Meal	BCOMSM	SM	CBOT	1.2%	8.0%	-0.013	0.217	3,530	1,071	7,007	1,363
	Soybean Oil	BCOMBO	BO	CBOT	0.2%	7.3%	-0.223	0.362	10,033	2,105	18,038	3,586
	Soybeans	BCOMSM	S	CBOT	0.7%	7.2%	-0.005	0.086	76,006	17,464	417,091	38,106
	Wheat	BCOMWH	W	CBOT	-0.5%	8.5%	-0.043	0.063	174,693	43,308	15,002,109	77,920
Livestock	Feeder Cattle	BCOMKW	FC	KBOT	0.1%	4.6%	-0.032	0.384	368	263	1,053	361
	Lean Hogs	BCOMFC	LH	CME	-0.8%	7.7%	-0.063	0.152	5,804	1,268	7,162	1,645
	Live Cattle	BCOMLH	LC	CME	-0.1%	4.3%	-0.057	0.127	21,540	3,518	18,367	4,386
Precious Metals	Gold	BCOMLC	GC	CME	0.7%	4.8%	0.016	0.058	1,652,667	229,978	193,700,000	842,490
	Platinum	BCOMGC	PL	COMEX	0.6%	6.6%	0.055	0.104	59,796	19,645	5,748,839	36,908
	Silver	BCOMPL	SI	NYMEX	0.8%	8.9%	0.023	0.075	396,921	66,988	12,915,940	269,739
Softs	Cocoa	BCOMSI	CC	COMEX	0.6%	8.8%	0.003	0.134	19,334	6,220	172,449	6,282
	Coffee	BCOMCC	KC	ICE(US)	-0.7%	8.9%	0.017	0.064	94,705	36,984	6,227,180	86,723
	Cotton	BCOMKC	CT	ICE(US)	-0.2%	8.1%	-0.023	0.120	44,876	8,789	162,533	9,124
	Orange Juice	BCOMCT	JO	ICE(US)	0.0%	8.6%	-0.007	0.150	13,297	8,829	241,705	10,586
	Sugar	BCOMOJ	SB	ICE(US)	0.4%	9.2%	-0.030	0.082	72,973	26,519	1,504,379	62,154

Table 2 Commodity returns and sentiment

This table reports the panel regression results. The dependent variable is a panel of commodity returns and the independent variables include sentiment, monthly changes in sentiment (ΔSentiment) and volume (ΔVolume), and fundamental commodity characteristics encompassing basis, past 12-month returns, basis-momentum and relative-basis (refer to Appendix I). FE indicates fixed effects, NW presents Newey-West standard errors, N denotes the total number of observations, and Adj. R2 denotes the adjusted R-squares. The sample period covers January 2009-December 2020, with varying start dates dictated by the availability of explanatory variables. * denotes significance at 5% or better.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sentiment	0.037*		0.031*		0.033*		0.026*		0.031*	
	4.51		3.54		3.64		2.86		3.72	
ΔSentiment		0.026*		0.029*		0.023*		0.024*		0.024*
		3.49		3.58		3.15		2.91		2.96
ΔVolume			-0.000	-0.000			-0.000	-0.000	-0.000	-0.000
			-1.17	-1.16			-1.33	-1.32	-1.34	-1.33
Basis			-0.001	0.008			0.039	0.050	0.040	0.052
			-0.02	0.11			0.48	0.61	0.40	0.52
Past return			0.076*	0.079*			0.093*	0.095*	0.089*	0.091*
			11.43	11.80			12.96	13.10	12.47	12.62
Basis-momentum									0.123	0.121
									1.46	1.40
Relative-basis									-0.104	-0.107
									-1.29	-1.30
Year/Month FE	N	N	N	N	Y	Y	Y	Y	Y	Y
Commodity FE	N	N	N	N	Y	Y	Y	Y	Y	Y
NW	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	3844	3736	3380	3329	3844	3736	3380	3329	3362	3315
Adj. R2	0.006	0.003	0.072	0.074	0.042	0.041	0.111	0.113	0.114	0.114

Table 3 Performance summary

This table reports the performance of long-short sentiment and commodity factors in Panel A and long-only investments in Panel B. AVG denotes an equally weighted and monthly rebalanced portfolio of all commodities in the sample. To construct long-short portfolios, we sort all commodities into high and low groups based on the respective characteristic at the end of each month, then take long (short) positions in commodities within the high (low) group (except for skewness). The portfolios are weighted equally and rebalanced monthly. To match the time across strategies, the sample period covers January 2010-December 2020.

	Sentiment	Carry	Momentum	Hedging Pressure	Skewness	Basis-Momentum	Relative-Basis
Panel A: Long-short benchmarks							
Annualized Mean	7.2%	3.3%	0.2%	5.0%	4.3%	4.4%	3.4%
t-statistics	2.9	1.6	0.4	2.0	1.7	2.0	1.7
Annualized Volatility	9.6%	11.2%	12.9%	12.9%	12.7%	10.9%	10.0%
Annualized Downside	5.6%	6.6%	8.4%	8.3%	6.7%	6.2%	6.4%
Sharpe Ratio	0.75	0.30	0.02	0.39	0.34	0.40	0.33
Sentiment Sharpe/Benchmark	0.0x	1.5x	46.3x	0.9x	1.2x	0.9x	1.3x
Sortino Ratio	1.42	0.61	0.12	0.73	0.78	0.82	0.61
Omega Ratio	1.77	1.31	1.03	1.40	1.33	1.36	1.30
Skewness	0.062	0.021	-0.150	-0.131	0.618	0.185	-0.083
Excess Kurtosis	0.407	0.476	0.289	0.053	1.212	0.259	0.718
99%VaR(Cornish-Fisher)	7.2%	8.2%	9.4%	9.6%	7.8%	7.5%	7.7%
%Positive Months	61%	53%	48%	56%	49%	54%	56%
Maximum Drawdown	-12.2%	-44.8%	-57.6%	-29.1%	-33.9%	-29.9%	-32.5%
CER	5.3%	0.8%	-3.2%	1.6%	1.2%	2.0%	1.3%
	AVG	**Energy**	**Grains**	**Livestock**	**Metals**	**Precious**	**Softs**
Panel B: Long-only benchmarks							
Annualized Mean	-2.0%	-2.6%	-0.5%	-4.3%	4.9%	6.2%	-1.5%
t-statistics	-0.2	0.3	0.4	-0.9	1.4	1.9	0.0
Annualized Volatility	13.6%	29.7%	22.3%	15.1%	20.8%	20.6%	17.8%
Annualized Downside	9.3%	20.6%	13.7%	10.2%	13.0%	13.2%	10.7%
Sharpe Ratio	-0.15	-0.09	-0.02	-0.28	0.23	0.30	-0.09
Sentiment Sharpe/Benchmark	5.1x	8.4x	33.4x	2.7x	2.2x	1.5x	8.7x
Sortino Ratio	-0.1x	0.1x	0.1x	-0.3x	0.6x	0.7x	0.0x
Omega Ratio	0.90	1.02	1.05	0.84	1.31	1.34	0.98
Skewness	-0.29	-0.33	0.12	-0.28	-0.10	-0.20	0.03
Excess Kurtosis	1.084	1.594	0.617	-0.064	1.419	0.553	0.715
99%VaR(Cornish-Fisher)	0.107	0.250	0.155	0.105	0.169	0.161	0.127
%Positive Months	48.5%	53.4%	47.4%	47.8%	54.6%	55.8%	49.4%
Maximum Drawdown	-56%	-93%	-65%	-70%	-59%	-57%	-57%
CER	-5.9%	-25.0%	-10.8%	-9.1%	-4.3%	-2.7%	-8.1%

Table 4 Time-series spanning tests

This table reports the regression results of sentiment profits using the Bakshi et al. (2019) model and augmented versions of the model. The dependent variable is the sentiment factor returns. The baseline model includes AVG, Basis and Momentum factors and the augmented models include hedging pressure, skewness, relative-basis and basis-momentum. *t*-statistics underneath each regression coefficient are based on Newey-West standard errors. The last row reports the adjusted *R*-squared. The sample period covers January 2010-December 2020. * denotes significance at 10% or better.

	(1)	(2)	(3)	(4)	(5)	(6)
Annualized Alpha	0.0684*	0.0696*	0.0684*	0.0684*	0.0660*	0.0684*
	2.45	2.52	2.43	2.45	2.42	2.45
AVG	0.0610	0.0654	0.0614	0.0634	0.0719	0.0781
	0.97	0.94	0.96	1.03	1.16	1.16
Basis	-0.1868*	-0.1824*	-0.1840*	-0.1783*	-0.2281*	-0.2165*
	-1.94	-1.92	-1.97	-1.74	-2.55	-2.38
Momentum	-0.0236	-0.0187	-0.0240	-0.0253	-0.0424	-0.0413
	-0.35	-0.25	-0.35	-0.38	-0.62	-0.54
Hedging pressure		-0.0175				-0.0123
		-0.18				-0.12
Skewness			-0.0106			0.0069
			-0.19			0.11
Relative-basis				-0.0273		-0.0355
				-0.32		-0.41
Basis-momentum					0.1343	0.1368
					1.37	1.29
Adj. R2	0.029	0.022	0.022	0.022	0.038	0.016

Table 5 Cross-sectional pricing tests

This table reports prices of risk from Fama-MacBeth regressions. The test assets are a panel of 46 portfolios that includes 40 quintile portfolios sorted by commodity fundamentals and 6 sector portfolios. The risk factors used for obtaining step one betas include the Bakshi et al. (2019) factors as well as other commodity fundamentals including hedging pressure, skewness, relative-basis and basis-momentum. *t*-statistics underneath each step two lambdas are based on Shanken (1992) corrected errors. The last two rows report the adjusted *R*-squared and MAPE (mean absolute pricing error) of each model. The sample period covers January 2010-December 2020. * denotes significance at 10% or better.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Sentiment	0.010*		0.010*		0.005		0.007*		0.011*		0.008*		0.007*
	1.89		1.98		1.48		1.66		2.28		1.83		1.95
AVG		-0.004	-0.002	-0.005	-0.004	-0.005	-0.003	-0.004	-0.002	-0.005	-0.004	-0.006	-0.004
		-0.63	-0.30	-0.78	-0.63	-0.76	-0.49	-0.66	-0.26	-0.47	-0.28	-0.66	-0.43
Basis		-0.008	-0.008	-0.001	-0.002	-0.001	-0.002	-0.006	-0.005	-0.004	-0.004	0.002	0.001
		-1.29	-1.01	-0.24	-0.48	-0.23	-0.27	-0.91	-0.53	-0.34	-0.27	0.29	0.14
Momentum		-0.010	-0.010	0.001	-0.001	0.005	0.004	-0.009	-0.009	-0.005	-0.005	0.007	0.005
		-1.10	-0.85	0.08	-0.16	0.67	0.49	-0.94	-0.59	-0.28	-0.22	0.59	0.41
Hedging pressure				0.008	0.006							0.007	0.002
				1.40	0.99							0.74	0.20
Skewness						0.012*	0.011					0.008	0.01
						1.99	1.59					0.86	0.89
Relative-basis								0.004	0.007			0.003	0.005
								0.57	0.58			0.30	0.42
Basis-momentum										-0.017	-0.017	-0.007	-0.007
										-1.29	-1.07	-0.76	-0.66
Adj-R2	0.051	0.143	0.191	0.197	0.236	0.185	0.232	0.212	0.260	0.198	0.248	0.341	0.381
MAPE(%)	0.039	0.053	0.053	0.036	0.039	0.041	0.042	0.060	0.063	0.045	0.044	0.036	0.044

Table 6 Sources of sentiment returns

This table reports the average return of short/long leg to and the long-short portfolio to the sentiment strategy in various market and economic conditions. VIX denotes the CBOE Volatility Index. B&W is the Baker and Wurgler (2006) sentiment index. S&P500 and S&P/GSCI denote the 24-month return of the respective index. Basis and hedging pressure are the monthly cross-sectional mean of basis and hedging pressure. Yield curve is the one-year change in the slope of US treasury curve. Term spread is the difference between 10-year and 3-month Treasury Yield. TED spread is the difference between 3-month LIBOR based on US dollars and 3-month Treasury Bill. GDP and Inflation denote year-on-year change in global industrial production and G7 CPI, respectively. CFNAI is the Chicago Fed National Activity Index. BDI, EPU and GPR denote the Baltic dry, economic policy uncertainty, and geopolitical risk index, respectively. High (low) indicates periods when the value of the macro variable is above (below) the sample mean. Bold denotes significance at 10% or better. *p* denotes the *p*-value of the null hypothesis that returns in the high versus the low state are identical. The sample period covers January 2010-December 2020.

	VIX			B&W			S&P500		
	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Panel A: market sentiment and states									
High	-0.51%	0.03%	0.54%	-0.79%	-0.24%	0.55%	-0.20%	0.49%	0.69%
	-0.7	0.1	1.5	-1.6	-0.5	**2.4**	-0.4	1.0	**2.5**
Low	-0.21%	0.46%	0.68%	0.24%	0.93%	0.70%	-0.60%	-0.10%	0.50%
	-0.5	1.2	**2.3**	0.4	1.9	1.8	-1.0	-0.2	1.5
p(Ho: H=L)	0.68	0.56	0.77	0.15	0.12	0.74	0.57	0.43	0.69

	Basis			Hedging pressure			S&P/GSCI		
	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Panel B: Commodity market states									
High	-1.25%	-0.44%	0.81%	-0.20%	0.44%	0.64%	-0.03%	0.64%	0.67%
	-2.2	-0.7	**2.5**	-0.3	0.8	2.0	-0.1	1.6	**2.3**
Low	0.52%	0.94%	0.41%	-0.52%	0.05%	0.58%	-0.84%	-0.32%	0.52%
	1.2	**2.6**	1.5	-1.0	0.1	**2.1**	-1.2	-0.5	1.7
p(Ho:H=L)	**0.01**	**0.06**	0.39	0.65	0.60	0.88	0.26	0.20	0.74

	Yield curve			Term spread			TED spread		
	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Panel C: Business cycle									
High	-1.00%	0.10%	1.10%	-0.78%	0.21%	0.99%	-0.68%	0.35%	1.03%
	-1.9	0.2	**4.0**	-1.2	0.4	**2.9**	-1.3	0.8	**3.0**
Low	0.28%	0.40%	0.12%	0.06%	0.29%	0.23%	-0.09%	0.16%	0.25%
	0.5	0.8	0.4	0.1	0.6	1.0	-0.2	0.3	0.8
p(Ho: H=L)	**0.07**	0.69	**0.03**	0.24	0.92	**0.10**	0.41	0.80	**0.09**

	GDP			Inflation			CFNAI		
	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Panel D: Macroeconomy									
High	-0.23%	0.47%	0.70%	-0.57%	0.33%	0.91%	0.13%	0.34%	0.20%
	-0.4	1.1	**2.4**	-1.2	0.8	**3.0**	0.3	0.6	0.7
Low	-0.50%	0.02%	0.52%	-0.15%	0.16%	0.31%	-0.83%	0.16%	0.99%
	-0.8	0.0	1.6	-0.3	0.3	1.1	-1.6	0.3	**3.0**
p(Ho: H=L)	0.70	0.54	0.70	0.55	0.82	0.20	0.17	0.81	**0.09**

	BDI			EPU			GPR		
	Short	Long	L-S	Short	Long	L-S	Short	Long	L-S
Panel E: Shipping cost, policy, and political risks									
High	0.01%	0.58%	0.57%	0.37%	0.61%	0.24%	-0.43%	-0.02%	0.40%
	0.0	1.2	1.8	0.6	1.1	0.7	-0.9	-0.1	1.6
Low	-0.73%	-0.08%	0.65%	-1.09%	-0.12%	0.98%	-0.30%	0.52%	0.82%
	-1.5	-0.2	2.0	**-2.7**	-0.3	**4.2**	-0.5	0.9	**2.1**
p(Ho: H=L)	0.30	0.37	0.86	**0.04**	0.32	0.11	0.86	0.47	0.37

Table 7 Sentiment with influential tweets

This table reports the performance of the sentiment strategy with modified sentiment measures. Instead of computing sentiment based on all tweets (i.e., regardless of whether a tweet is liked or retweeted), we re-measure sentiment based on tweets that contain only non-zero versus zero retweets, likes and jointly. In other words, we re-calculate sentiment based on a sub-group of tweets. At the end of each month, we sort all commodities into high and low groups based on sentiment shifts re-calculated using the respective sub-group of tweets, then take long (short) positions in commodities within the high (low) group. The portfolios are weighted equally and rebalanced monthly. *p* denotes the *p*-value of the null hypothesis that sentiment strategies using tweets with non-zero versus zero retweet/likes generate identical average returns. The sample period covers January 2010-December 2020.

	Retweet		Likes		Retweet + Likes	
	non-zero	zero	non-zero	zero	non-zero	zero
Annualized Mean	4.4%	3.7%	1.8%	4.9%	4.0%	3.6%
t-statistics	1.8	1.4	0.8	1.6	1.4	1.2
Annualized Volatility	9.6%	9.6%	12.1%	11.4%	10.8%	11.0%
Annualized Downside Volatility	5.5%	5.8%	8.0%	8.1%	7.6%	7.5%
Sharpe Ratio	0.46	0.38	0.15	0.43	0.37	0.33
Sortino Ratio	0.91	0.72	0.32	0.70	0.62	0.57
Omega Ratio	1.45	1.31	1.16	1.40	1.35	1.28
Skewness	0.089	-0.015	0.266	-0.406	-0.552	-0.457
Excess Kurtosis	0.242	0.691	2.398	1.803	2.866	1.563
99%VaR(Cornish-Fisher)	0.069	0.072	0.095	0.103	0.106	0.097
%Positive Months	56.1%	53.8%	50.8%	56.1%	57.6%	55.3%
Maximum Drawdown	-14%	-17%	-26%	-32%	-22%	-32%
CER	2.6%	1.8%	-1.1%	2.2%	1.6%	1.1%
p(Ho: non-zero=zero)	0.84		0.54		0.94	

Table 8 Tweet intensity

This table reports the performance of the sentiment strategy in sub-groups. Instead of deploying the strategy on the full cross-section, we implement the same strategy in two sub-group of commodities (i.e., 50% most active versus least active) divided by the total number of unique tweets, retweets and likes. We measure sentiment based on all tweets irrespective of retweets or likes status. At the end of each month, we sort all commodities within each sub-group into high and low portfolios based on sentiment shifts, then take long (short) positions in commodities within the high (low) portfolio. The portfolios are weighted equally and rebalanced monthly. *p* denotes the *p*-value of the null hypothesis that the sentiment strategy deployed in the sub-group of commodities with the most twitter activity generates lower average returns than those in the least active sub-group. The sample period covers January 2010-December 2020.

	Tweets		Retweets		Likes	
	Least Active	Most Active	Least Active	Most Active	Least Active	Most Active
Annualized Mean	3.7%	6.7%	2.7%	10.1%	4.0%	8.7%
t-statistics	1.2	2.0	0.9	2.6	1.3	1.8
Annualized Volatility	12.4%	13.8%	14.2%	15.5%	13.0%	16.9%
Downside Volatility	6.9%	6.2%	8.3%	7.8%	7.8%	7.6%
Sharpe Ratio	0.30	0.49	0.19	0.65	0.31	0.51
Sortino Ratio	0.66	1.27	0.46	1.52	0.63	1.39
Omega Ratio	1.24	1.58	1.36	1.67	1.29	1.51
Skewness	0.037	0.704	-0.114	0.657	-0.105	0.753
Excess Kurtosis	-0.396	1.602	-0.466	1.239	0.022	1.273
99%VaR(Cornish-Fisher)	0.083	0.086	0.097	0.097	0.094	0.099
%Positive Months	50.8%	56.1%	54.5%	52.3%	53.0%	57.6%
Maximum Drawdown	-28%	-15%	-29%	-21%	-31%	-24%
CER	0.6%	3.1%	-1.3%	5.5%	0.6%	3.4%
p(Ho:most<least)	0.13		0.01		0.13	

Table 9 Alternative Lexica

This table reports the performance of sentiment strategies using alternative sentiment lexica. At the end of each month, we sort all commodities into high and low groups based on sentiment shifts measured using Bing, NRC, TextBlob and AFINN, respectively. We then take long (short) positions in commodities within the high (low) group (except for skewness). The portfolios are weighted equally and rebalanced monthly. The sample period covers January 2010-December 2020.

	(1) L&M	(2) Bing	(3) NRC	(4) TextBlob	(5) AFINN
Annualized Mean	7.2%	2.2%	-0.2%	1.5%	2.8%
t-statistics	2.9	0.9	0.1	0.7	1.1
Annualized Volatility	9.6%	9.9%	9.3%	9.8%	10.3%
Annualized Downside Volatility	5.6%	6.8%	5.6%	5.5%	6.7%
Sharpe Ratio	0.75	0.22	-0.02	0.16	0.28
Sortino Ratio	1.42	0.40	0.04	0.37	0.51
Omega Ratio	1.77	1.15	1.11	1.17	1.22
Skewness	0.062	-0.057	0.011	0.273	-0.110
Excess Kurtosis	0.407	2.002	-0.040	0.734	0.197
99%VaR(Cornish-Fisher)	0.072	0.083	0.062	0.066	0.076
%Positive Months	61.3%	57.0%	50.7%	49.2%	49.2%
Maximum Drawdown	-12%	-18%	-28%	-25%	-33%
CER	5.3%	0.2%	-1.9%	-0.3%	0.7%

Table 10 Alternative signals

This table reports the performance of sentiment strategies using alternative sentiment signals. Panels A and B report results based mean and sd sentiment measures, respectively. Level denotes the raw $Sentiment_{i,t}$ whereas Change denotes $\Delta Sentiment_{i,t}$. DMA represents $\Delta Sentiment_{i,t}$ at time t minus the $\Delta Sentiment_{i,t}$ averaged over the last $N \in \{6, 9, 12\}$ months. At the end of each month, we sort all commodities into high and low groups based on the alternative sentiment signal, then take long (short) positions in commodities within the high (low) group. The portfolios are weighted equally and rebalanced monthly. The sample period covers January 2010-December 2020.

	(1)	(2)	(3)	(4)	(5)
Panel A: Mean sentiment	Level	Change	DMA=6	DMA=9	DMA=12
	3.7%	7.2%	5.2%	7.7%	6.6%
Annualized Mean	1.7	2.9	2.1	3.3	2.7
t-statistics	8.8%	9.6%	9.8%	9.5%	9.7%
Annualized Volatility	4.2%	5.6%	5.0%	4.6%	5.2%
Annualized Downside Volatility	0.42	0.75	0.53	0.80	0.68
Sharpe Ratio	0.99	1.42	1.17	1.84	1.39
Sortino Ratio	1.50	1.77	1.68	1.86	1.67
Omega Ratio	0.290	0.062	0.282	0.400	0.176
Skewness	-0.253	0.407	-0.096	-0.002	0.369
Excess Kurtosis	0.055	0.072	0.063	0.061	0.069
99%VaR(Cornish-Fisher)	51.0%	61.3%	58.2%	58.9%	59.6%
%Positive Months	-17%	-12%	-15%	-14%	-14%
Maximum Drawdown	2.2%	5.3%	3.3%	5.8%	4.7%
CER	3.7%	7.2%	5.2%	7.7%	6.6%

	(6)	(7)	(8)	(9)	(10)
Panel B: SD sentiment	Level	Change	DMA=6	DMA=9	DMA=12
Annualized Mean	2.7%	4.2%	4.7%	4.4%	3.4%
t-statistics	1.1	1.8	2.3	1.8	1.5
Annualized Volatility	9.5%	8.9%	8.1%	9.3%	8.6%
Annualized Downside Volatility	7.0%	6.1%	4.1%	6.5%	5.9%
Sharpe Ratio	0.28	0.47	0.58	0.48	0.40
Sortino Ratio	0.45	0.77	1.25	0.76	0.65
Omega Ratio	1.24	1.39	1.76	1.39	1.35
Skewness	-0.442	-0.418	0.058	-0.614	-0.538
Excess Kurtosis	0.193	1.529	-0.494	1.013	1.069
99%VaR(Cornish-Fisher)	0.075	0.079	0.055	0.081	0.074
%Positive Months	57.6%	59.1%	57.6%	61.4%	59.8%
Maximum Drawdown	-29%	-20%	-10%	-19%	-19%
CER	0.8%	2.6%	3.3%	2.6%	1.9%

Table 11 Placebo test

This table reports the performance of strategies using $\Delta GoogleTrends_{i,t}$ as sorting signals. DMA represents $\Delta GoogleTrends_{i,t}$ at time t minus the average $\Delta GoogleTrends_{i,t}$ over the last $N \in \{3, 6, 9, 12\}$ months. At the end of each month, we sort all commodities into high and low groups based on the respective GoogleTrends signal, then take long (short) positions in commodities within the high (low) group. The portfolios are weighted equally and rebalanced monthly. The sample period covers January 2010-December 2020.

	(1) GoogleTrends	(2) DMA=3	(3) DMA=6	(4) DMA=9	(5) DMA=12
Annualized Mean	0.8%	2.5%	0.3%	0.1%	1.4%
t-statistics	0.5	1.1	0.3	0.2	0.7
Annualized Volatility	10.7%	10.4%	10.4%	11.3%	10.7%
Annualized Downside Volatility	8.1%	6.9%	6.6%	8.2%	8.0%
Sharpe Ratio	0.07	0.24	0.03	0.01	0.13
Sortino Ratio	0.17	0.45	0.13	0.09	0.25
Omega Ratio	1.04	1.20	1.03	1.01	1.10
Skewness	-0.855	-0.422	-0.177	-0.857	-0.975
Excess Kurtosis	5.870	1.670	1.273	4.611	5.772
99%VaR(Cornish-Fisher)	0.127	0.091	0.083	0.123	0.126
%Positive Months	52.6%	55.6%	50.3%	51.9%	52.4%
Maximum Drawdown	-34%	-35%	-38%	-40%	-34%
CER	-1.8%	0.3%	-1.9%	-2.7%	-1.1%

Table 12 Liquidity, portfolio construction and sub-period

This table reports the performance of the sentiment strategy in various specifications. (1) At the end of each month, we exclude commodities with the 20% least volume in the cross-section before deploying the strategy. (2) Instead of equally weighting the portfolio, a rank-weight is implemented where the weight of commodity $w_t^i = c\left(Rank(\Delta sentiment_t^i) - \frac{N_t+1}{2}\right)$, where c is a scaling factor that ensures the weights sum up to one. (3) and (4) Instead of medians, we sort commodities intro terciles and quintiles, respectively. Then take positions in commodities within the most extreme quantiles. The sample period for all specifications covers January 2010-December 2020, except for (5), where the performance in January 2015-December 2020 is reported.

	(1) 80% most liquid	(2) Rank-weights	(3) Breakpoint Q3	(4) Breakpoint Q5	(5) Last 5 years
Annualized Mean	6.8%	6.2%	7.9%	6.2%	6.7%
t-statistics	2.6	2.2	2.7	1.7	2.3
Annualized Volatility	10.2%	11.4%	11.9%	15.0%	9.3%
Annualized Downside Volatility	5.8%	6.6%	6.6%	9.4%	6.2%
Sharpe Ratio	0.67	0.54	0.67	0.41	0.71
Sortino Ratio	1.31	1.07	1.37	0.80	1.18
Omega Ratio	1.72	1.48	1.64	1.37	1.71
Skewness	0.148	0.028	0.127	0.164	0.051
Excess Kurtosis	0.318	0.166	0.107	0.795	1.270
99%VaR(Cornish-Fisher)	0.073	0.083	0.084	0.109	0.076
%Positive Months	57.7%	60.6%	60.6%	52.8%	62.5%
Maximum Drawdown	-15%	-16%	-17%	-23%	-8%
CER	4.8%	3.6%	5.1%	1.7%	4.9%

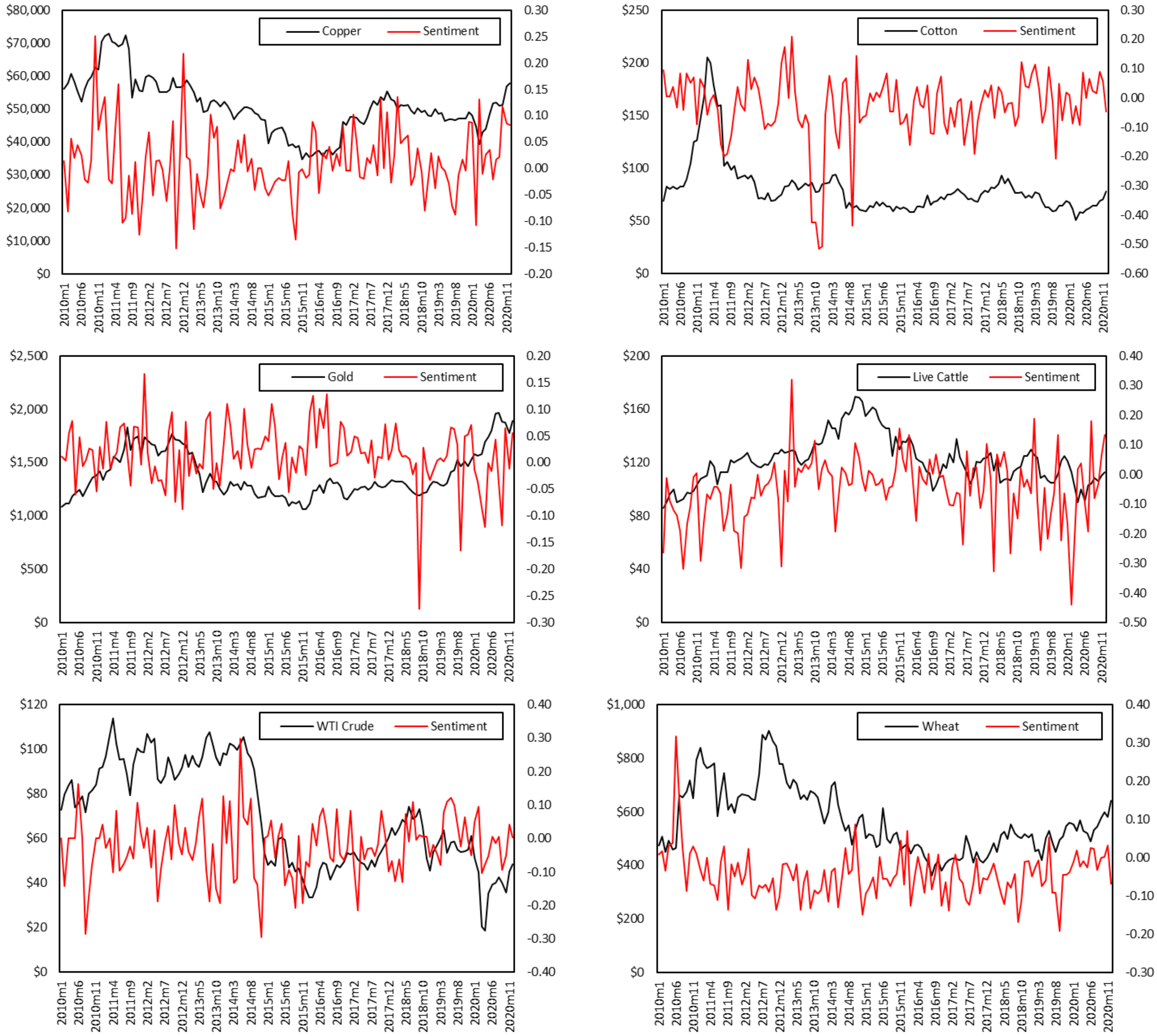


Figure 1 Commodity prices and sentiment

This figure illustrates the settlement price (primary y-axis) and sentiment (secondary y-axis) of selected commodities across six sectors. The red (black) line demotes the sentiment (prices).

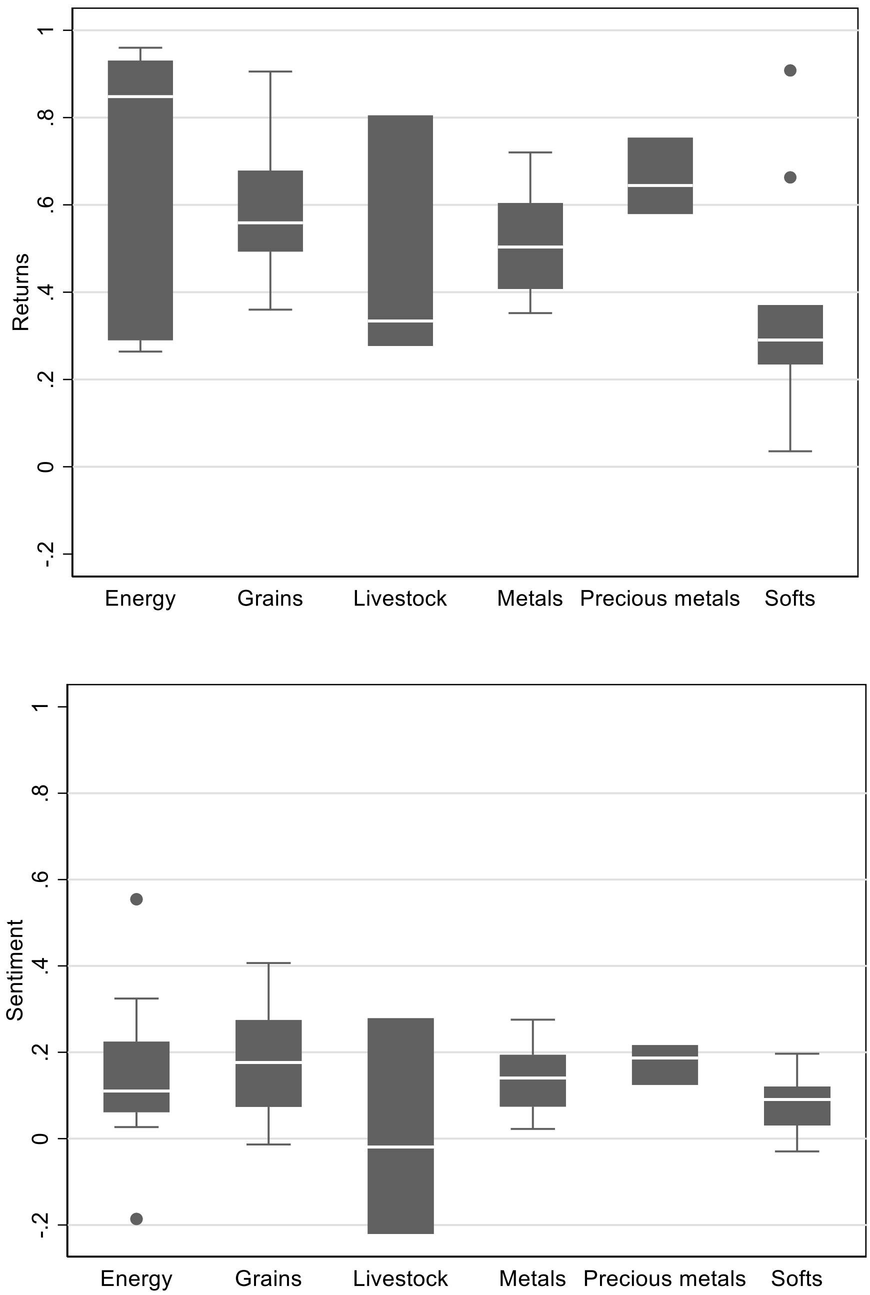


Figure 2 Correlations

This figure illustrates Spearman's rank correlations of returns (upper) and sentiment (lower) within each commodity sector. The sample period covers January 2010-December 2020.

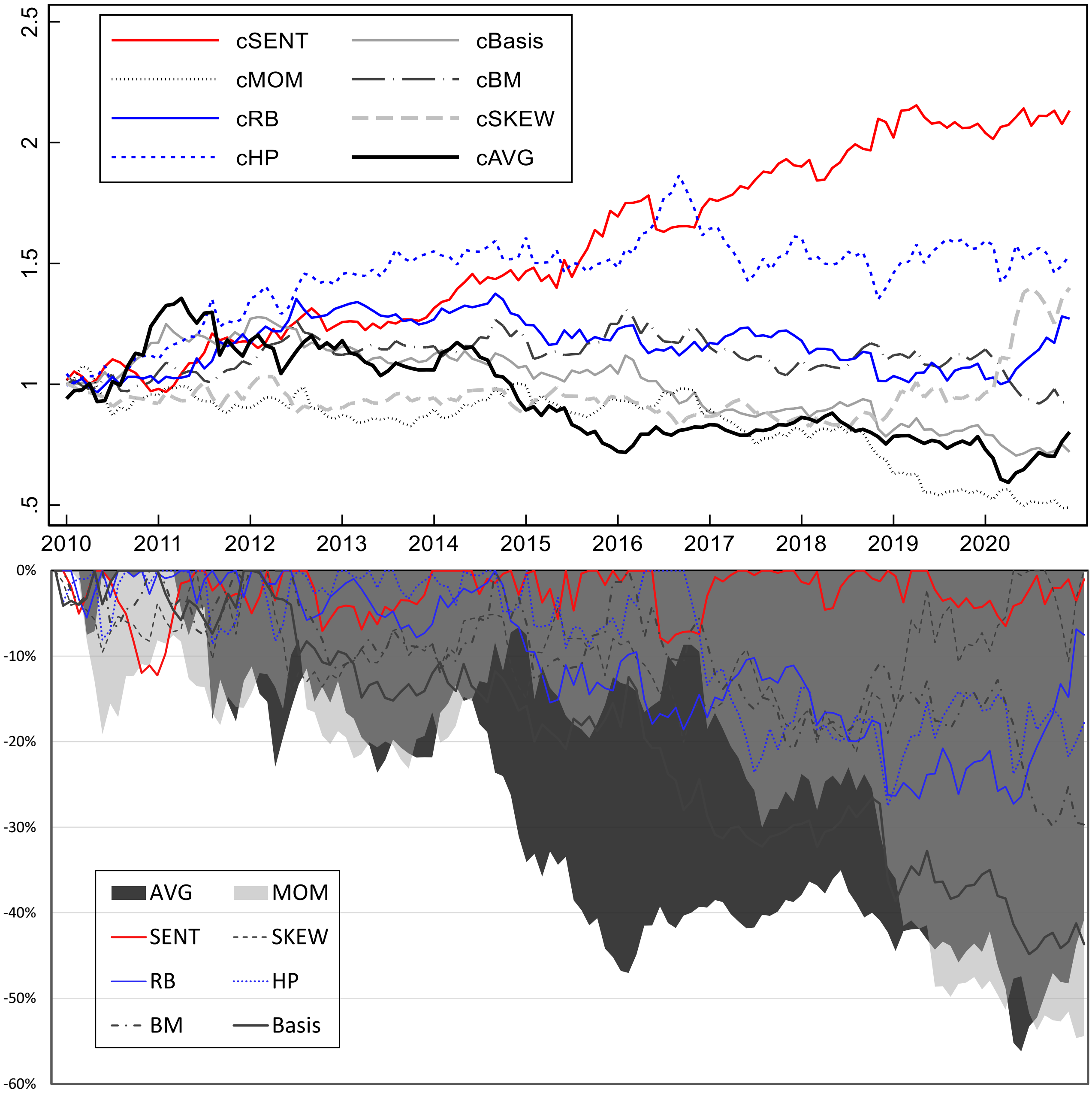


Figure 3 Cumulative performance

This figure illustrates the cumulative excess return (upper) and drawdown (lower) of commodity factor strategies. MOM, HP, RB, BM and SENT represent long-short strategy based on momentum, hedging pressure, relative-basis, basis-momentum and sentiment signals. AVG is an equal-weighted portfolio of all commodities rebalanced monthly.

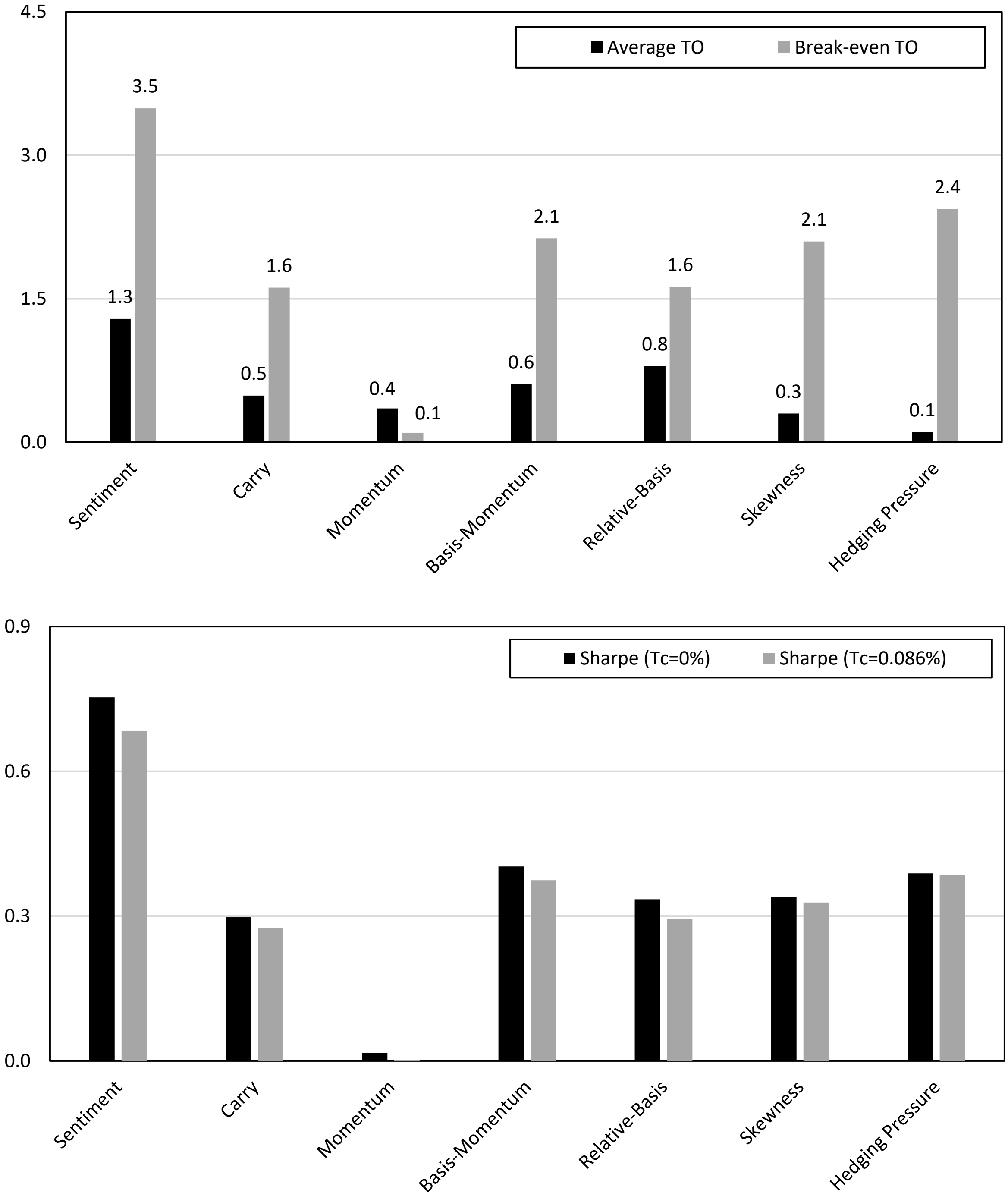
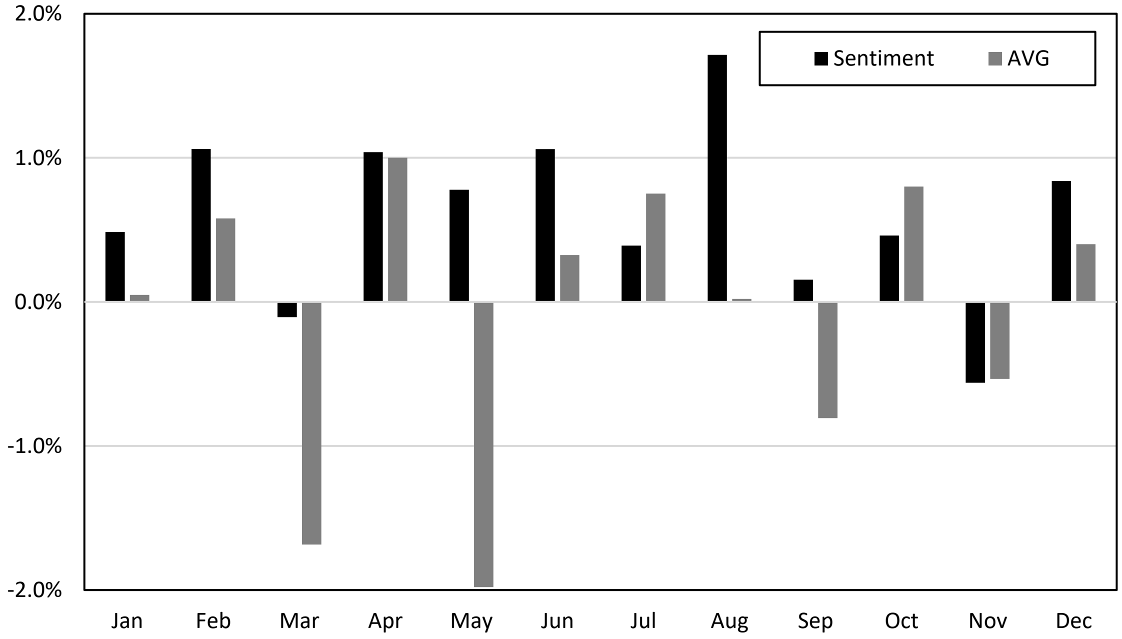


Figure 4 Transaction costs

This figure illustrates the turnover TO (upper) and Sharpe ratios (lower) of commodity factor strategies. where $TO = \frac{1}{T-1}\sum_{t=1}^{T-1}\sum_{i=1}^{N}\left(\left|w_{t+1}^{i} - w_{t+}^{i}\right|\right)$. The after-cost Sharpe ratio is based on net returns, where $\tilde{r}_{P,t} = r_{P,t} - 0.5 \times \sum_{c=1}^{N} TO_t^i \times TC_t^i$. $w_{t+1}^{c,i}$ is the weight of commodity i at month t+1 for the strategy under consideration, w_{t+}^{i} denotes the actual weight of commodity i at the end of month t+1 prior to the rebalancing of the strategy and after accounting for the performance of the commodity from t to t+1.



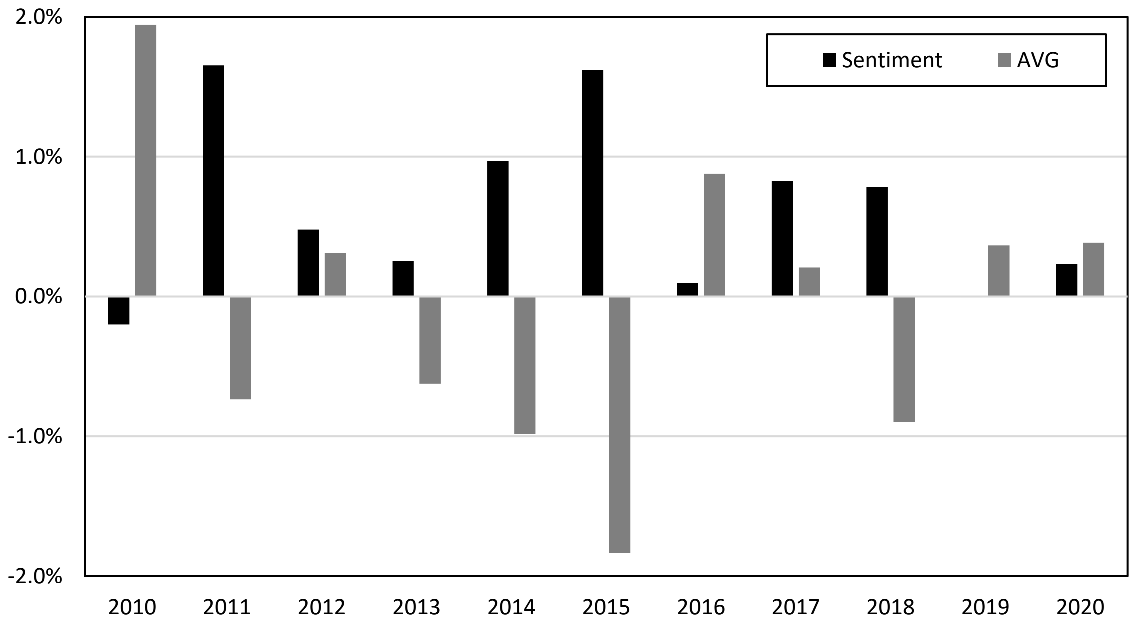


Figure 5 Seasonality

This figure illustrates the average returns of the sentiment strategy in calendar months (upper) and years (lower). AVG is an equal-weighted portfolio of all commodities rebalanced monthly.