Hidden in the spotlight: Media emotion intensity and commodity futures pricing

Yeguang Chi

Business school, University of Auckland
y.chi@auckland.ac.nz

Lina El-Jahel

Business school, University of Auckland
l.eljahel@auckland.ac.nz

Thanh Vu

Business school, University of Auckland
thanh.vu@auckland.ac.nz

Abstract

We investigate the role of media emotion in commodity futures pricing and propose a new factor, media emotion intensity, based on the proportion of emotional content relative to factual content. Our factor exhibits an annual premium of around 14% after we control for other commonly considered benchmark factors. The impact of media emotion is especially strong for commodities with low media coverage, high momentum, high basis-momentum, high hedging pressure, and backwardation. Media emotion intensity significantly predicts the cross-section of commodity futures return both at the portfolio level and the individual commodity level. Our simulated LASSO approach suggests that media emotion intensity is the most robust factor compared to other commonly considered benchmark factors. Furthermore, we investigate various risk channels that are potentially related to media emotion intensity and demonstrate that they cannot subsume the predictability of this media factor.

Keywords: asset pricing, factor model, commodity futures pricing, media news, emotion intensity

1. Introduction

In recent decades, commodity futures have gained increasing importance in investor portfolios (Adams and Glück, 2015; Belousova and Dorfleitner, 2012; Daskalaki and Skiadopoulos, 2011), with substantial inflows and a growing trend towards financialization (Tang and Xiong, 2012). As a result, understanding the sources of risk in commodity markets has become increasingly critical. In response, research has emerged and developed on the factors that can explain the cross-section returns of commodity futures, leading to the identification of several key benchmarks, including the basis (Szymanowska et al., 2014), momentum (Bakshi et al., 2017), and basis-momentum (Boons and Prado, 2019), and other factors (De Roon et al., 2000; Fernandez-Perez et al., 2018; G. Gorton and Rouwenhorst, 2006; Hong and Yogo, 2012; Kang et al., 2020; Sakkas and Tessaromatis, 2020; Szymanowska et al., 2014). The research on factor investing has developed into a thriving field of study, dedicated to discovering new factors that can produce significant abnormal returns and predict the cross-section of returns while considering the predictive power of benchmark factors. Over the past four decades, equity market research has identified and proposed more than 300 factors (Feng et al., 2020; Harvey et al., 2016). Conversely, the number of factors suggested in the commodity market remains limited in comparison.

The literature on media news impact on asset pricing proposes two factors: news sentiment (García, 2013; Smales, 2014; Tetlock, 2007) and media coverage (Fang and Peress, 2009). News sentiment is built on the proportion of positive words and phrases to their negative counterparts. This media news sentiment explains asset price movement via two aspects. On the one hand, the positive and negative levels of media news may be associated with the positive and negative information in the news content that fundamentally explains asset prices (Tetlock et al., 2008). On the other hand, news sentiment can trigger investors' behavioral biases, which drive price movement (García, 2013; Tetlock, 2007). Meanwhile, media coverage measures the volume of news articles that mention a given asset. Fang and Peress (2009) investigate this factor and suggest that stocks without news in the media generate significantly higher returns than those highly featured in the media. In this study, we are interested in an additional narrative-related feature of the media: the use of emotion in news articles. When journalists write news articles based on a set of facts, they would contribute more through their choice of words and phrases, which carry emotional content.

Academic research has placed greater emphasis on the use of emotion in journalism, particularly as digital and social media continue to thrive (Wahl-Jorgensen, 2020). Beckett (2015) argues that journalism uses emotion as a tool to attract attention and engage more readers. Consequently, emotion becomes increasingly central in the content of the news. Goldenberg and Gross (2020) demonstrating that media transmits emotions to readers and increased exposure to emotional content makes perceivers more emotional. The important role of emotion prompts a series of recent studies on the influence of emotion use in media (Bas and Grabe, 2015; Beckett and Deuze, 2016; Murry and Dacin, 1996; Orgeret, 2020; Peters, 2011; Uribe and Gunter, 2007). Since media emotion is highly transmissible, increased exposure to emotional content in media news can amplify the impact of media on readers.

Furthermore, emotion plays a crucial role in information processing and decision-making (Ekman, 2007; Keltner and Lerner, 2010; Keltner et al., 2019; Lerner et al., 2015; G. F. Loewenstein et al., 2001), suggesting that greater exposure to emotional content may trigger a greater impact on reader judgements and decisions. A number of neuroscientific studies have shown that humans tend to give higher priority to emotional stimuli (see Vuilleumier (2005) for a comprehensive review). Rozin et al. (1986) claim that it is difficult to ignore emotion in decision making when emotion is related to it. However, the literature remains inconclusive on whether emotion impairs or enhances judgment (see Pham (2007) and Lerner et al. (2015) for a review). On the one hand, some studies indicate that emotion can trigger behavioral biases in judgment (Han et al., 2007; G. F. Loewenstein et al., 2001; G. Loewenstein, 1996; G. Loewenstein and Lerner, 2003). G. F. Loewenstein et al. (2001) show that stronger emotions associated with decision making can overrule rational behaviors. On the other hand, some studies suggest that emotion can enhance judgement (Loomes and Sugden, 1982; Schwarz, 2000; Solomon, 1993).

In this study, we investigate the impact of emotion delivered by media news on commodity futures returns using a novel factor, media emotion intensity. To construct media emotion intensity for each commodity, we employ the *EmotionVsFact* index from Thomson Reuters *MarketPsych* dataset for each commodity. The index is calculated by subtracting the amount of factual content from the amount of emotional content and scaled by the total amount of content. It represents the excess proportion of emotional content over factual content and is adjusted by the total amount of media coverage. Our media emotion intensity factor is designed to capture how emotional a news article is. It is important to note that emotional words and phrases can be either positive or negative. The emphasis is on how emotional an article is, not on how positive or negative it is.

First, we examine the return of long-short portfolios sorted by media emotion intensity. At the end of each month, we sort the commodities in our sample into three portfolios according to each commodity's media emotion intensity in the month. We group the four commodities with the highest values into the High portfolio; we group the four commodities with the lowest values into the Low portfolio; and we group the remaining commodities into the Mid portfolio. We then construct the High-minus-Low portfolio by longing the High portfolio and shorting the Low portfolio in an equal-weighted manner. We show that this long-short portfolio generates an average annualized return of 13.51% (t-stat = 2.95). In comparison, we use the same sample and construct the basis, momentum and basis-momentum long-short portfolios and find that the basis-momentum factor and momentum factor yield significant annualized returns of 12.50% (t-stat = 2.66) and 9.57% (t-stat = 1.71), respectively, while the basis factor yields nonsignificant returns. After that, we run the time-series spanning test (Fama and French, 1993) using basis, momentum, and basis-momentum as the benchmark factors. Our media emotion intensity factor still generates significant annualized abnormal returns of around 11.5% after we control for the above factors. We find that the overlapping percentages of the High and Low portfolios ranked by media emotion intensity with the other portfolios are small (less than 35%). These results highlight the importance of the media emotion intensity factor, with respect to existing factors such as basis, momentum, and basis-momentum..

Second, we further examine the returns of the media emotion intensity long-short portfolio for different groups of commodities sorted by other factors. We do this by double-sorting commodities based on the media emotion intensity and other factors, including media coverage basis, momentum, and basis-momentum. Our results show that for commodities with low media coverage, higher media emotion intensity generates significantly higher average returns than low media emotion intensity (13.9% higher with a t-stat of 2.54). However, for commodities with high media coverage, there is no significant difference in returns between those with high and low media emotion intensity. This finding suggests that media emotion intensity has a more profound impact on commodities less featured in media news. When single-sorting commodities by basis, momentum, and basis-momentum, our results are in line with Boons and Prado (2019). Additionally, we show that the media emotion intensity factor premium is only significant for commodities with backwardation, high momentum, and high basis-momentum.

Third, we test whether media emotion intensity can predict the cross-section of commodity futures returns and whether its premium is subsumed by other factors. We begin by examining the predictive power of media emotion intensity using the Fama-Macbeth (1973) cross-section test against the benchmark factors, including basis, momentum, and basis-momentum. The results of this test, at both the portfolio and individual commodity levels, reveal that the premium of media emotion intensity remains significant and stable at approximately 1.2% per month or 14% per year, a level that is comparable to what is observed in the univariate sorting. This consistency of the estimated premium of the media emotion intensity factor suggests that its predictive power is not subsumed by the predictive ability of the benchmark factors. In the cross-section tests, only the media emotion intensity and basis-momentum factors can significantly predict the cross-section of both portfolio and individual commodity futures returns.

Further, we run LASSO regressions to identify the factors with the greatest predictive potential. We consider the following factors: basis, momentum, basis-momentum and media factors. First, following Feng et al. (2020), we generate 200 randomized subsamples (trials) for each LASSO regression. Second, we further apply three distinct LASSO regression techniques (regular LASSO, adaptive LASSO, and plug-in LASSO) to rank the strength of factors' predictive power. In the standard LASSO test, we find that the media emotion intensity factor is chosen in 100% of the trials, whereas the basis-momentum factor is ranked second and is selected in 72% of the trials. In the adaptive LASSO test, the media emotion intensity factor is chosen in 99.5% of the trials, and the basis-momentum factor is again the runner-up selected in 55% of the trials. In the plug-in LASSO test, the media emotion intensity and basis-momentum factors are chosen in 84.5% and 84.0% of the trials, respectively. The LASSO test results further corroborate our findings that the media emotion intensity factor is an important predictor for the cross-section of commodity futures returns.

Finally, we investigate whether the media emotion intensity premium is subsumed by other risk sources. These risk sources include hedging pressure (Boons and Prado, 2019; De Roon et al., 2000; Kang et al., 2020; Keynes, 1930), inventory (Boons and Prado, 2019; G. B. Gorton et al., 2013; Kaldor, 1939), market liquidity and funding liquidity (Boons and Prado, 2019; Brunnermeier and Pedersen, 2009), market volatility (Boons and Prado, 2019; Brunnermeier and Pedersen, 2009), stock market and downside stock market (Boons and Prado, 2019; Brunnermeier and Pedersen, 2009), and macroeconomic risk (Le Pen and Sévi, 2018; Szymanowska et al., 2014). The results of the cross-section tests confirm that the media emotion intensity premium remains significant at around 14% annualized after controlling for the above risk sources. The robustness of the media emotion intensity premiums highlights that the predictive power of the media emotion intensity is not subsumed by the commonly considered risk sources in the commodity futures market.

Our key contribution is to the literature on commodity asset pricing. We demonstrate that media emotion intensity is an important new factor that significantly explains the cross-section of returns in the commodity futures market. In our paper, we show that the media emotion intensity factor's average abnormal return is comparable to that of the basis-momentum factor in all the main tests, and far more significant, both statistically and economically, than all the other aforementioned factors. Not only does the media emotion intensity factor survive the times-series spanning test and the cross-section Fama-Macbeth tests, but it also ranks top in the three LASSO regression tests. In addition, we contribute to the literature on the role of media news in commodity markets. We study a novel factor that captures the emotional intensity of news articles. We show that compared to the measures of news sentiment (García, 2013; Smales, 2014; Tetlock, 2007) and media coverage (Fang and Peress, 2009), the emotional intensity factor plays a significant role in explaining the cross-section of returns in the commodity futures market. Additionally, we show that media emotion intensity only generates strong impacts on commodities with low media coverage, backwardation, high momentum, high basis-momentum, and high hedging pressure.

The rest of our paper is organized as follows. Section 2 describes the data and methodology. Section 3 examines the premium of media emotion intensity factor, with respect to other commonly considered risk factors. Section 4 conducts cross-section tests and LASSO regressions to further test the predictive power of media emotion intensity. Section 5 concludes and discusses research extention. The Appendix includes a description of the media data source and further discussions about the analyses in Section 3 and Section 4.

2. Data and Methodology

This section describes our two primary data sources: commodity futures data and media emotion intensity data. Along with the data introduction, we include definitions and measurements for the variables used in our study.

2.1. Commodity futures data

In this study, we collect data for commodity futures from Barchart®, formerly known as Commodity Research Bureau (CRB), the same data source used in the research of Szymanowska et al. (2014) and Boons and Prado (2019). We survey 26 commodity futures traded on CBOT, NYMEX, COMEX, CME and ICE exchanges from January 1998 to February 2020. These commodity futures are selected as those that also have data available for media emotion intensity from Thomson Reuters. We begin our sample in January 1998 because it was the first month Thomson Reuters started compiling the media data used in this research.

Return of commodity futures

Following Boons and Prado (2019), the monthly excess returns on a fully collateralized futures position are calculated as

$$R_{i,t+1}^{T_n} = \frac{F_{i,t+1}^{T_n}}{F_{i,t}^{T_n}} - 1 \tag{1}$$

where $R_{i,t+1}^{T_n}$ is the return of n^{th} nearby futures contract of the commodity i for the month $t+1$. $F_{i,t}^{T_n}$ is the price of the n^{th} nearby futures contract of commodity i at the end of month t. In this study, we focus on the first and second nearby futures, since they are more liquid than further-to-maturity contracts. We follow the approach of Szymanowska et al. (2014) and Boons and Prado (2019). We define the first nearby contract as the contract tradable at the end of month t that expires after month t + *2* and is closest to maturity, with maturity T_1. This selection avoids rolling considerations and ensures that our selected futures prices remain tradable continuously. For example, at the end of June, we will pick the first nearest contract as the nearest contract with the last trading day after the end of August. By doing so, the contract will be held until the end of July, when the contract has at least one month before the expiration date. This gap is secure enough to avoid the unusual trading and price movement when the contract closes to its maturity (Boons and Prado, 2019). We define the second nearby contract as the next nearest to maturity contract tradable at the end of month t that immediately follows the first nearby contract, with maturity T_2.

Boons and Prado (2019) and Szymanowska et al. (2014) show that the premium in commodity futures markets can be categorized into spot premiums and term premiums. Given the illiquidity of underlying commodities, the spot premiums should be better captured by longing the first nearby futures contract. Conversely, term premiums are captured by longing the first nearby contract and shorting the second nearby contract. Specifically, we define the nearby return, R^{nb}_{t+1}, as the return of the first nearby futures contract; and the spreading return, $R^{sp}_{t+1} = R^{T_1}_{t+1} - R^{T_2}_{t+1}$ as the return difference between the first nearby contract and the second nearby contract. We offer descriptive statistics of the nearby and spreading returns of commodity futures in Table 1.

[Table 1 here]

Table 1 shows that most agricultural commodity futures, including grains, softs, and food oils, earn negative returns on average. In contrast, precious metals have positive average returns over the sample period. These statistics are consistent with the commodity futures performance reported in and Boons and Prado (2019). The standard deviation of nearby returns is, however, significantly higher than the corresponding returns, indicating that the movement of commodity futures prices is largely volatile. Hence, selecting commodity futures plays an influential role in short-term investment, while in the long term holding precious metal or energy futures might be more beneficial to investors.

Main benchmark factors in the previous literature

Since Szymanowska et al. (2014) published their study, the set of priced risk factors in the commodity futures market has expanded. Szymanowska et al. (2014) demonstrate that the basis factor can predict both the nearby and spread returns. Bakshi et al. (2017) shows that the momentum factor is priced in the average nearby return. Boons and Prado (2019) introduce the basis-momentum as an additional important factor and demonstrate that the basis, momentum and basis-momentum factors are the three key tradable factors in the commodity futures market. We follow Boons and Prado (2019) to construct these factors as follows. First, we define basis ($Basis_{i,t}$) as

$$Basis_{i,t} = \frac{F^{T_2}_{i,t}}{F^{T_1}_{i,t}} - 1 \tag{2}$$

Second, we define momentum ($Mom_{i,t}$) based on the returns of the past 12 months as:

$$Mom_{i,t} = \prod_{k=t-11}^{t} \left(1 + R_{i,k}^{T_1}\right) - 1 \tag{3}$$

Third, we define basis-momentum ($BM_{i,t}$) as the difference in the momentum of first and second nearby futures contracts as:

$$BM_{i,t} = \prod_{k=t-11}^{t} \left(1 + R_{i,k}^{T_1}\right) - \prod_{k=t-11}^{t} \left(1 + R_{i,k}^{T_2}\right) \tag{4}$$

2.2. Media news data

In this study, we focus on a novel media factor, media emotion intensity, in addition to two other well-documented media factors: news sentiment and media coverage. To measure these factors, we employ the Thomson Reuters Marketpsych Indices (TRMI) database, which analyzes the news content from various financial media news outlets fed by the Refinitiv Machine Readable News (MRN).

Media coverage

We commence with a well-known media effect, media coverage. Media coverage reflects how much content the media news provides readers about a specific asset in a specific period. Fang and Peress (2009) quantified this factor by counting the number of news items referencing each stock. In this study, we favor the MRN because this platform covers hundreds of financial news sites. Thomson Reuters and Refinitiv offer a data system called TRMI that analyses the news content from the MRN. In TRMI, we notice a variable called *"buzz"*, which might more accurately convey the concept of media coverage. First, *"buzz"* sifts through the text of each news article, identifying each term and phrase directly referring to a particular asset. Second, *"buzz"* considers the amount of information contained in each news article. By doing so, *"buzz"* can weigh longer and shorter news items differently. In this study, we use *"buzz"* as a proxy for media coverage.

More specifically, the system will gather all news items related to a particular asset released over a period. After that, each word directly related to the asset and its associated attributes are stored as a so-called variable in the system. The system then scores all variables obtained from the scanned text. The details on how Thomson Reuters defines and scores variables are discussed in Section A1 of Appendix. Finally, the system sums up the scores of all variables associated with the asset and stores this value in an indicator named "*buzz*". To demonstrate how "*buzz*" is calculated, we begin with "*buzz*" for one minute. For a certain asset a and a specific one-minute period, the system records all variables (words and phrases with their attributes). This set of variables is denoted by the symbol $V(a)$. The system then scores each variable v in the set $V(a)$. Let $S(v)$ denote the score value of v. The "*buzz*" generated for the asset a over a period of one minute is calculated as

$$buzz_t^{1-min}(a) = \sum_{v \in V(a)} S(v) \tag{5}$$

Given that the system reports news in the UTC time zone, all time stamps in the TRMI data are converted to the CST time zone. For the month T, the media coverage of the asset a is calculated as the sum of all one-minute "*buzz*" for the asset a encompassing the period from the daily settlement time of the last trading day of the month $T-1$ to the daily settlement time of the last trading day of the month T. The more news content related to the asset a, the more variables are recorded in the system. As a result, "*buzz*" will take on a higher value.

Media emotion intensity

We propose a novel media factor termed media emotion intensity, which quantifies the proportion of emotional content to factual content. This factor is distinct from the well-documented media effect of news sentiment (García, 2013; Smales, 2014; Tetlock, 2007; Tetlock et al., 2008). The precision with which news sentiment is measured depends largely on how words and phrases in media reports are categorized as positive or negative. This classification is challenging, since a single word may communicate different sentimental meanings in different contexts. However, measuring media emotion intensity is more straightforward and transparent because it is less confusing when differentiating emotional content from factual content. Media emotion intensity reflects how emotional media news is towards a specific asset. When journalists generate news articles, they already have a collection of factual materials. They add more to the news through their use of words and phrases that carry emotional content. As a result, it is essential to consider the amount of factual content when calculating the intensity of emotions.

The TRMI data system determines whether each variable v in the set of all variables $V(a)$ for the asset a refers to emotional or factual content. Specifically, it defines the function $I_{emo}(v)$ as

$$I_{emo}(t, v) = \begin{cases} +1 \text{ if v refers to emotional content} \\ -1 \text{ if v refers to factual content.} \end{cases} \tag{6}$$

According to this indicator function, the media emotion intensity of asset a is constructed as

$$Media\ emotion\ intensity(a) = \frac{\sum\limits_{v \in V(a)} (I_{emo}(v) \times S(v))}{buzz(a)}. \tag{7}$$

Because this indicator is scaled by *buzz*, it takes the value from -1 to 1 and reflects the proportion of purely emotional aspects in the news content. The higher value of the indicator corresponds to the higher intensity of emotional content compared with factual content. In the TRMI data, the one-minute media emotion intensity is provided as the one-minute *EmotionVsFact* indicator. To aggregate media emotion intensity for monthly frequency, we take the average of one-minute media emotion intensity weighted by one-minute *buzz*. This measure facilitates the tests for how media emotion intensity reflects the mispricing of commodity futures, particularly when media coverage is controlled.

In this study, we also measure news sentiment using the same approach as media emotion intensity. To assist the sorting of commodities based on news sentiment, we favor a net sentiment measure ranging from -1 to 1. When news sentiment is closer to 1, the media is more positive towards the asset. In contrast, when news sentiment is closer to -1, the media is more negative towards the asset. Appendix A2 discusses the specifics of the news sentiment measurement. The descriptive statistics of media emotion intensity, media coverage and news sentiment are reported in Table 2. The average emotion intensity of all commodities is positive, implying that emotional content makes up a larger part of the news content compared to factual content.

[Table 2 here]

All reported commodities have positive media emotion intensity on average. This demonstrates that media news frequently includes more emotive words and phrases to enhance the story and captivate the audience. Among different commodity types, precious metals have the highest average emotion intensity, with the four examined taking values over 0.4. Some agricultural commodities have the lowest average values of media emotion intensity. Corn, Hogs, and Soybeans have the lowest media emotion intensity. It is not surprising that the media writes more emotionally about precious metals and energy than about agricultural commodities. The low standard deviation of media emotion intensity for each commodity suggests that media emotion intensity may be somewhat related to commodity identification. However, because the maximum value of media emotion intensity for each commodity is larger than the average value of media emotion intensity for most other commodities, there is no commodity whose media emotion intensity value is dominated by other commodities. For media coverage, media outlets prefer energy-related news more, with the highest media coverage going with crude oil. Gold has also attracted the attention of news writers with the position of the second highest media coverage. The difference between the highest and lowest averages of media emotion intensity and media coverage suggests that media emotion intensity represents a distinct feature of the media from news volume. All commodities' average news sentiments are negative and close to zero.

3. Media emotion intensity long-short portfolio

This section investigates the media emotion intensity factor in the commodity futures market. We first test portfolio performance by sorting commodities by media emotion intensity and holding long-short portfolios for one month. We also compare the performance of media emotion intensity with other media and benchmark factors. In the second part, we conduct the double-sorting using media emotion intensity to test whether this factor adds more cross-section explanation when the commodities are already sorted by other factors.

3.1. Single-sorting portfolios

To investigate the long-short portfolios formed by media factors, we sort 26 commodities into three portfolios based on media emotion intensity, media coverage, and news sentiment. At the end of month t, we rank commodity futures according to their signal values for the month t. The High4 portfolio holds four commodities with the highest signal values, while the Low4 portfolio maintains four commodities with the lowest signal values. The Mid portfolio consists of all remaining commodities with their futures contracts available at the end of the month t. Additionally, we construct the High4 minus Low4 portfolio (High4-Low4) by taking long positions on the High4 portfolio and short positions on the Low4 portfolio. We evaluate these portfolios' performance using both nearby and spreading returns. Regarding nearby returns, we take only long positions on the High4-Low4 portfolio using the first nearby commodity futures contracts. As for spreading returns, we take long positions on the High4-Low4 portfolio using the first nearby contracts and short positions on the High4-Low4 portfolio using the second nearby contracts. Each portfolio is formed at the end of the month t and held until the end of the month $t+1$. Boons and Prado (2019) claim that by examining nearby and spreading returns, the sorting would aid in assessing the predictability of factors in the cross-section setting and across maturities.

3.1.1. Portfolio performance

The results of the sorting are presented in Table 3. Panel A shows the performance of portfolios sorted by media emotion intensity, while Panel B reports the sorting result on media coverage. Panel C provides the sorting result for the news sentiment. We also sort the commodities on basis, momentum, and basis-momentum (see Panel D) to compare their performance for our sample period with the reported in earlier studies. Due to the limited availability of the media data, we must restrict our sample to the period from February 1998 to February 2020. This sample period is significantly shorter than the sample period used in the research of Bakshi et al. (2017), Boons and Prado (2019), and Szymanowska et al. (2014). For this reason, we expect that the significance (t-stat) for testing the average returns of the portfolios in our sample will be lower than the significance reported in the previous studies for the identical portfolios.

[Table 3 here]

Among three media factors and three key benchmarks, media emotion intensity presents the most remarkable effects, with the average nearby return of the High4 - Low4 portfolio standing at 13.51% and the highest significance with a t-stat of 2.95. The basis momentum still proves to be the highest effect among the three benchmarks with an average nearby return of 12.50% for the High4 - Low4 portfolio and a t-stat of 2.66. However, the effect of basis-momentum in our sample is weaker than in the sample from August 1960 to February 2014 in Boons and Prado (2019). In this earlier work, all basis, momentum, and basis-momentum significantly affect nearby returns, with the nearby returns of the High4 - Low4 portfolios being -10.61%, 15.02% and 18.38%, respectively.

Given that basis-momentum has been considered the best predictor for commodity futures returns so far, our results show that media emotion intensity is also a potential predictor with the higher average nearby return of the High4-Low4 portfolio and higher t-stat compared with the performance of basis-momentum. Media coverage might also be a good candidate with the high average nearby return of the Low4 - High4 portfolio (9.16% with t-stat of 2.69). This effect of media coverage has a power similar to that of momentum. Furthermore, the effect of media coverage in our research is consistent with Fang and Peress (2009) for the equity market. The higher return on assets with low media coverage may suggest that the media supplies information which facilitates price discovery. Regarding the spread returns, the basis momentum still shows the greatest effect, with the average spreading return of High4 - Low4 standing at 4.64% (t-stat = 4.78). Consistent with Boons and Prado (2019), basis and momentum do not prove their predictability for spreading returns. Interestingly, both media emotion intensity and media coverage can significantly predict spreading returns, with the average spreading returns of High4 - Low4 being 3.35% (t-stat = 4.97) and -2.11% (t-stat = -3.88), respectively. In summary, media emotion intensity, media coverage, and basis-momentum are all powerful predictors for both nearby and spreading returns. Especially, media emotion intensity can show better predictability than basis-momentum with a more significant effect.

We also examine whether news sentiment can predict the returns in the cross-section setting. García (2013) found evidence on the opposite effects of positive and negative news sentiment. Additionally, both García (2013) and Tetlock (2007) confirm the overreaction and reversal following increased news sentiments. However, the effects in these studies are short-term. In Panel C of Table 3, we report the performance of the High4 - Low4 portfolio sorted by news sentiment. The average nearby and spreading return of the High4 - Low4 portfolio sorted by news sentiment are 5.60% (t-stat = 1.13) and 0.72% (t-stat = 0.76). The finding suggests that news sentiment does not significantly predict the cross-section of commodity futures returns in monthly frequency. From here, we will eliminate the news sentiment from further analysis.

According to Boons and Prado (2019), the transaction costs associated with the basis-momentum long-short strategy could be as high as 158 basis points. They assume that investors would update three out of four commodity futures contracts monthly in each of the High4 and Low4 portfolios. Their calculation is based on the 4.4 basis points of the average effective half-spread for trading commodity futures estimated by Marshall et al. (2012). When we sort by media emotion intensity, we observe that precious metals, crude oil, natural gas, and orange juice are regularly featured in the High4 portfolios, whereas agricultural commodity futures are frequently included in the Low4 portfolios. Silver futures appear the most in the High4 portfolio (78%), and hogs futures appear the most in the Low4 portfolio (79%), meaning that we can assume transaction costs similar to Boons and Prado (2019). Additionally, the commodity futures in the High4 and Low4 portfolios are not highly illiquid. As a result, transaction fees in the above estimation are not overlooked. The estimated transaction costs of 158 basis points imply that the 13% returns of the High4-Low4 portfolio ranked by media emotion intensity can comfortably exceed the cost. Similarly, the 9.16% profit of the High4-Low4 portfolio sorted by media coverage also survives the transaction cost. For these reasons, media emotion intensity and media coverage can be favorable for the long-short strategy.

3.1.2. Time-series spanning test

We employ the time series spanning test to determine whether the return of portfolios sorted by other benchmark factors span the return of the portfolio sorted by media emotion intensity. Boons and Prado (2019) conducted the spanning test for basis-momentum against the factor set proposed by Szymanowska et al. (2014) and Bakshi et al. (2017). The Boons and Prado (2019) spanning test highlights the importance of testing the significance of alpha in comparing asset pricing models, which is discussed in Barillas and Shanken (2017, 2018). We follow the same logic as Boons and Prado (2019).

The results of the spanning tests are summarized in Table 4. In Panel A, we examine whether the returns to the portfolio sorted by the media emotion intensity are spanned those sorted by the factors proposed in Bakshi et al. (2017) and Boons and Prado (2019). Similarly, Panel B displays the test employing factors in Szymanowska et al. (2014) and, again, in Boons and Prado (2019). We test four models on each panel to assess the portfolios sorted by media emotion intensity and media coverage. We name the nearby and spreading portfolios sorted by media emotion intensity as Emotion (nb) and Emotion (sp). As for media coverage, we also name the nearby and spreading portfolios as Coverage (nb) and Coverage (sp). The first two portfolios reflect the emotional impact of the media, whereas the latter two describe the influence of media coverage. We incorporate media coverage in the model to test the additional return of media emotion intensity, and vice versa; we include media emotion intensity in the models of media coverage factors. To determine whether the addition of media factors improves mean-variance efficiency, we apply Gibbons et al. (1989) joint GRS test for both the nearby and spreading returns of each factor.

[Table 4 here]

Panel A in Table 4 indicates that after controlling for basis-momentum, basis, momentum, average factor, and media coverage, the alpha of the media emotion intensity nearby portfolio remains significant (t-stat = 2.30). The alpha of nearby media emotion intensity portfolio drops only about 2% from 13.51% to 11.35% after controlling for other benchmarks. The first model in Panel B validates these findings, with an alpha of 11.72% (t-stat = 2.26). Additionally, the R^2 for media emotion intensity is minimal in both Panels A and B (0.09 and 0.04, respectively). This indicates that other benchmark factors only slightly correlate with the performance of the media emotion intensity nearby portfolio, emphasizing the benefit of the media emotion intensity factor. Both Panel A and Panel B report significant alphas for the media emotion intensity spreading portfolio. However, the alphas generated for the spreading portfolio are small, hanging at about 2%. In both Panels A and B, the GRS tests for media emotion intensity portfolios produce a large F-stat with a p-value lower than 1%. These results indicate that media emotion intensity can greatly enhance mean-variance efficiency.

Additionally, media coverage factors generate large alphas when tested against both factor sets in Panels A and B. In Panel A, the alpha for the media coverage nearby portfolio is -7.40% (t-stat = -2.22), while in Panel B, it is -8.19% (t-stat = -2.43). $R2$ is also small (less than 5%) in these two tests, indicating that the media coverage nearby portfolio is also weakly connected to other benchmarks. The media coverage spreading factor's performance is comparable to that of the media emotion intensity spreading portfolio. Although both Panels A and B have significant alphas for this portfolio, its alphas are only below 2%. However, the GRS tests prove that media coverage factors can significantly improve the mean-variance efficiency.

3.1.3. Portfolio overlap

One concern regarding portfolio formation based on media emotion intensity is that the average of this factor may be related to the nature of the commodity and that the standard deviation of this factor for each commodity is relatively small. This issue can diminish the value of commodity ranking based on media emotion intensity. However, statistics on the frequency of each commodity in the High4 and Low4 groups indicate that no commodity is virtually always present in the group with the highest and lowest media emotion intensity. Table 5 provides descriptive information on the percentage of each commodity appearing in the High4 and Low4 groups, sorted by the factors examined in Table 3.

[Table 5 here]

Table 5 reveals that precious metals have the highest proportion of occurrences in the High4 group. This result is consistent with the fact that precious metals are the commodity associated with the highest average emotional intensity in the media. Silver (77.82%) was the most prevalent precious metal in the High4 group, followed by Palladium (66.92%) and Platinum (66.54%). On the contrary, hogs were found more frequently in the Low4 group (78.95%), followed by corn (61.28%), soybeans (60.53%), and canola (53.33%). In general, most commodities have the opportunity to appear in the High4 and Low4 groups.

In contrast, crude oil consistently ranks among the four commodities with the highest media coverage value, whereas soybean oil generally appears among the four commodities with the lowest media coverage value. The highest group also sees the appearance of other energy commodities and gold. Palladium, which is in the group of commodities with the highest media emotion intensity, is mostly in the group with the lowest media coverage. There is no clear similarity between media emotion intensity and media coverage in commodity ranking. Table 5 also shows a different pattern when sorting commodities by news sentiment. This result suggests that the three media factors may reflect different aspects of commodities. In addition, the media emotion intensity ranking differs significantly from the basis, momentum, and basis-momentum rankings.

To further check the overlap between portfolios sorted by three media factors and three benchmark factors, we calculate the percentage that each group of High4 and Low4 sorted by each factor has the same commodities as other groups. To do so, with each factor, we only consider the four commodities with the highest factor values (High4) and the four commodities with the lowest factor values (Low4) at the end of each month. We then count the number of the same commodities in each pair of groups and calculate the overlapping percentage by dividing over four. The result is reported in Table 6.

[Table 6 here]

Table 6 shows that there is a 34.1% overlap between the High4 group ranked by media emotion intensity and the High4 group ranked by news sentiment. However, the High4 portfolio ranked by media emotion intensity also overlaps with the Low4 portfolio ranked by news sentiment at 25.6%. The overlap between the High4 portfolio ranked by media emotion intensity and the High4 and Low4 portfolios ranked by three benchmark factors is not greater than 20%. The similarity between the Low4 portfolio sorted by media emotion intensity and other portfolios sorted by other factors is less than 30%. These statistics imply that the sorting of media emotion intensity differs significantly from the sorting of other investigated factors. The similarities of media emotion intensity sorting are even smaller than those of basis-momentum. The High4 basis-momentum portfolio overlaps at 39.0% with the Low4 basis portfolio and 33.6% with the High4 momentum portfolio. In this analysis, we focus mainly on High4 portfolios sorted by media emotion intensity, momentum, and basis-momentum, and Low4 portfolios sorted by media coverage and basis, as these portfolios are the major contributors to the profit of long-short portfolios presented in Table 3. The overlap analysis implies that the return of the long-short portfolio sorted by media emotion intensity derives little from aspects reflected by media coverage, basis, momentum, and basis-momentum.

3.2. Double-sorting portfolios by media emotion intensity and other factors

In this section, we investigate whether media emotion intensity generates additional returns for portfolios sorted by other factors. To accomplish this, we double-sort commodities by media emotion intensity and each factor independently. Independent sorting is conducted through the separate ranking of the commodities according to media emotion intensity and each factor. To highlight this factor's importance among media factors, we first evaluate the added value of media emotion intensity to media coverage portfolios. After that, we examine the return of High-Low portfolios ranked by media emotion intensity inside each portfolio sorted by other benchmark factors.

3.2.1. Double-sorting approach

For media emotion intensity, we sort our 26 commodities into High and Low portfolios based on the median of media emotion intensity for each month. For other factors, we divide the commodities into portfolios: (1) contango and backwardation based on the positive or negative value of the basis of the month t, (2) high and low based on the median value of the factor at the month t. Because the High4-Low4 portfolio ranked by news sentiment does not generate significant return, we eliminate this factor from the analysis and only consider media coverage in conjunction with media emotion intensity.

Table 7 shows the overlap between each portfolio sorted by media emotion intensity, media coverage, and other factors. For each pair of portfolios generated each month, we count the number of commodities that appear in both portfolios and divide this number by the total number of commodities in the portfolios to determine the overlap percentage. The overall overlapping percentage is the average of the monthly overlapping percentages. Table 7 shows that the commodities with high media emotion intensity appear considerably in both two portfolios sorted by each factor. Hence, if media emotion intensity can predict the profit of commodity futures, it might be meaningful to double-sort commodities by each factor and media emotion intensity.

[Table 7 here]

We double-sort commodities using media emotion intensity and each factor to examine the returns of media emotion intensity High-Low portfolios in each portfolio sorted by other factors. We begin with independent sorting. Because commodities are sorted into two levels for each factor, we form four portfolios: (1) high media emotion intensity and high factor value, (2) high media emotion intensity and low factor value, (3) low media emotion intensity and high factor value, and (4) low media emotion intensity and low factor value.

3.2.2. Double-sorting on emotion and media coverage

Table 8 Panel A presents the result of independent sorting on media emotion intensity and media coverage. The third and fourth columns display average returns and t-stat for testing single-sorted portfolios by each factor. The High-Low portfolio consists of long positions in commodities with high factor value and short positions in commodities with low factor value. When single-sorting commodities into two levels, the High-Low portfolio sorted by media emotion intensity generates a 6.04% return on average with a t-stat of 2.10. Compared to High4-Low4 portfolios, the profit of the High-Low portfolio sorted by media emotion intensity is significantly lower, indicating that the commodities with the highest and lowest values of media emotion intensity may have the most extreme returns.

[Table 8 here]

Looking at media coverage, the Low portfolio generates a higher average return than the High portfolio. This result is consistent with Fang and Peress, 2009 that assets with higher media coverage earn higher returns. However, in the commodity futures market and during our sample period, both the High and Low portfolios sorted by media attention generate statistically insignificant average returns. Within the group of high media coverage commodities, it is interesting that both high and low media emotion intensity portfolios earn similar average returns at around 0.12% and are statistically insignificant with t-stats of 0.03. The High-Low portfolio sorted by media emotion intensity within high media coverage commodities generates only 0.01% (t-stat = 0.00) return on average. This result suggests that when assets are highly featured in the media, media emotion intensity has negligible effects on investors and asset returns. However, among commodities with low media coverage, the High portfolio ranked by media emotion intensity produces a significant 9.17% average return (t-stat = 2.34), while the Low portfolio generates a negative 3.53% average return (t-stat = -1.11). The High-Low portfolio sorted by media emotion intensity within the group of low media coverage commodities earns a return of 12.70% (t-stat = 3.39) on average, a huge improvement from the 2.42% return of the Low portfolio sorted by media coverage. This suggests that media emotion intensity has a significant impact on the return prediction for commodities with less media coverage.

3.2.3. Double-sorting on emotion and other benchmark factors

Next, we conduct the independent double-sorting using media coverage and each of the other benchmark factors. Table 8 Panel B reported the results of these tests. In single sorts, Basis, momentum, and basis-momentum generate significantly 8.11%, 6.88%, and 9.27%, respectively. When separating the high and low media emotion intensity commodities within each portfolio sorted by these factors, we obtain similar results, such that the High portfolios sorted by media emotion intensity produce higher average returns than the Low portfolio. The High-Low portfolio sorted by media emotion intensity improves the average return of backwardation commodities from 7.87% (t-stat = 1.86) to 13.90% (t-stat = 2.54), the average return of high momentum commodities from 4.94% (t-stat = 1.48) to 7.67% (t-stat = 2.07) and the average return of high basis-momentum commodities from 6.00% (t-stat = 1.77) to 8.85% (t-stat = 2.38). The returns of the above High-Low portfolios sorted by media emotion intensity are mainly derived from the high media emotion intensity portfolios with averages of 11.53% (t-stat = 2.34), 8.61% (t-stat = 2.22) and 9.98% (t-stat = 2.54) for basis, momentum and basis-momentum, respectively. Table 8 also reveals that the impact of the intensity of media emotion is stronger for commodities in the groups of low media coverage, backwardation, high momentum, and high basis-momentum. Meanwhile, we observe insignificant returns for the media emotion intensity High-Low portfolios in the portfolios of high media coverage, contango, low momentum, and low basis-momentum.

We also perform conditional sorting to test the returns of portfolios where the ranking is done hierarchically. In these tests, we conduct the following two approaches. First, commodities are sorted by factors, followed by media emotion intensity. In this way, for each portfolio of high and low factor values, we continue to rank the commodities using media emotion intensity. Second, we sort commodities first by media emotion intensity and then by factor value. After classifying commodities into portfolios with high and low media emotion intensity, we divide the commodities within each portfolio into high and low factor values. Conditional double-sorting results are presented in Tables A1 and A2 in the internet appendix. The results of conditional sorting are consistent with those found in Table 8. We provide a more detailed discussion of the conditional sorting results in Internet appendix A3.

3.2.4. Emotion difference in each portfolio sorted by factors

The double-sorting results reveal that media emotion intensity may have stronger effects for specific portfolios sorted by other factors, including low media coverage, backwardation, high momentum, and high basis-momentum. However, the insignificant returns of the media emotion intensity High-Low portfolios within other portfolios might come from the possibility that there is no significant difference in media emotion intensity between the commodities in the groups of high media coverage, contango, low momentum, and low basis-momentum. Hence, we compute the average media emotion intensity for the High and Low portfolios sorted by the media emotion intensity within the above groups. Table 9 presents the calculated results. In Table 9, the last column indicates the difference in media emotion intensity between the High and How media emotion intensity portfolios within each portfolio sorted by other factors.

[Table 9 here]

The last column of Table 9 shows the difference in media emotion intensity between the groups of high and low media emotion intensity within each portfolio sorted by other factors. In the portfolio of commodities that receive high media coverage, the difference in media emotion intensity difference between high and low media emotion intensity is statistically significant at 0.12. In comparison, the gap in the portfolio of low media coverage is significant at 0.17. Combined with the results in Table 8, while the difference in media emotion intensity for the high media coverage portfolio is not minimal compared to that for low media coverage, the High-Low portfolio sorted by media emotion intensity only generates a significant and high return for low media coverage. This supports the view that even if the difference in media emotion intensity is significant, media emotion intensity mostly exerts a significant impact on the commodities having a low level of media coverage. In addition, we discover that the differences in media emotion intensity between the High and Low media emotion intensity in the portfolios sorted by other benchmark factors are statistically significant. These differences are very close when comparing two portfolios sorted by each benchmark. However, similar to media coverage, Table 8 reveals that only portfolios with higher returns in single-sorting, namely backwardation, high momentum, and high basis-momentum, can generate more significant and higher returns.

4. Media emotion intensity and the cross-section of commodity futures return

To examine the predictability of media emotion intensity in the cross-section of commodity futures, we construct our two portfolios pertaining to media emotion intensity as follows. To begin, we use the nearby return and the spreading return of the High4 - Low4 portfolio sorted by media emotion intensity to present the two factors. We name two factors as "emotion (nb)" and "emotion (sp)", respectively. In Table 3, the spanning tests reveal that the returns of the High4-Low4 portfolios sorted by media emotion intensity and media coverage are not spanned by the key benchmark factors. Hence, in this part, we also conduct the cross-section test for media coverage to compare with the performance of media emotion intensity. We use the nearby and spreading returns of the High4 - Low4 portfolio ranked by media coverage to form the next two factors. These two factors are referred to as "Coverage (nb)" and "Coverage (sp)".

The literature has suggested some key benchmark factors for commodity futures returns, as investigated in Table 3. We also include these factors to conduct the cross-section test for our media factors. We first calculate the three benchmarks introduced by Szymanowska et al. (2014): the nearby return of the High4 - Low4 portfolio sorted by basis (basis (nb)) and the spreading returns of the High4 and the Low4 portfolios also sorted by basis (basis (h4-sp) and basis (l4-sp), respectively). Bakshi et al. (2017) suggest the next two benchmarks, which are the equal-weighted average return of all commodity futures (average (nb)), the momentum factor (momentum (nb)) calculated by the nearby return of the High4 - Low4 portfolio sorted by momentum. The last two factors were introduced by Boons and Prado (2019), the nearby and spreading returns of the High4 - Low4 portfolio sorted by basis-momentum (basis-momentum (nb) and basis-momentum (sp), respectively). So far, the basis-momentum (nb) factor is considered the most potent factor to explain the cross-section of commodity futures returns.

The descriptive statistics for our media and benchmark factors are reported in Table 10. The four factors generating the highest returns are, respectively, emotion (nb), basis-momentum (nb), momentum (nb) and coverage (nb). The emotion (nb) factor has relatively comparable descriptive statistics to the basis-momentum factor (nb). All factors exhibit particularly low autocorrelation, with the absolute values of the first-order autocorrelation maximum at 0.06. The pairwise correlations between the 11 factors are shown in Panel B of Table 10. Accordingly, these variables are only weakly associated with one another. Although the emotion (nb) factor correlates most strongly to the average (nb) factor, their correlation is only 0.21. Also, media coverage factors relate very weakly to all the benchmarks factors.

[Table 10 here]

Boons and Prado (2019) argue that little is known about the cross-section of commodity futures. Thus far, the most popular factors are those given by Szymanowska et al. (2014), Bakshi et al. (2017), and, most notably, Boons and Prado (2019) basis-momentum. We are motivated to test the role of media emotion intensity in the cross-section setting as we observe the high abnormal returns of the long-short portfolios sorted by media emotion intensity and media coverage. Also, we observe that the media emotion intensity nearby factor performs at the same level as basis-momentum, and media coverage also presents a comparable effect to that of the momentum factor. The media spreading factors are also significant in the spanning test. However, because they generate only about 2% of return, they might not be attractive when considering transaction costs. Therefore, in the cross-section test, we examine all four media factors but with a particular emphasis on the two nearby factors.

To begin, we examine the predictabilities of (1) our four media factors: media emotion intensity nearby, media emotion intensity spreading, media coverage nearby and media coverage spreading, (2) two factors from Boons and Prado (2019): basis-momentum nearby, basis-momentum spreading, (3) three factors introduced by Szymanowska et al. (2014): basis nearby, basis High4 spreading and basis Low4 spreading, and (4) two factors from Bakshi et al. (2017): average nearby and momentum nearby. In the cross-section test, we employ the Fama-Macbeth (Fama and MacBeth (1973)) regression to perform a two-stage regression. In the first stage, we regress the returns of test assets on the factors to estimate the exposure of each asset to each factor in the time-series setting. In the second stage, we regress the returns of the test assets in the month $t+1$ ($R_{i,t+1}$) on their exposures to the factors of the month t. Denote $\boldsymbol{\beta_{i,t}}$ is the vector of the exposures of the asset i to the factors in the factor set, and $\boldsymbol{\gamma}$ is the vector of factor premiums. We run the following regression in the second stage:

$$R_{i,t+1} = \gamma_0 + \boldsymbol{\beta'_{i,t}\gamma} + u_{i,t} \tag{8}$$

To assess the consistency of factor premiums, we run eight models with the selected factors nested within the factor set. Model (1) evaluates three Szymanowska et al. (2014) factors and two Boons and Prado (2019) basis-momentum factors. In Model (2), we test against the basis nearby, average nearby and momentum nearby factors as suggested in Bakshi et al. (2017) and two basis-momentum factors of Boons and Prado (2019). Model (3) begins to incorporate the media emotion intensity nearby factor into Model (1). Model (4) is Model (3), adding the media emotion intensity spreading factor. Similarly, we include the media emotion intensity nearby factor in Model (2) to create Model (5) and continue to incorporate the media emotion intensity spreading factor to form Model (6). We test the media coverage factors using the same model specifications by adding media coverage nearby and spreading factors to Models (1) and (2) to create Models (7), (8), (9) and (10), respectively. We conduct the cross-section test for two levels: the portfolio and commodity levels, to determine whether (1) media factors can explain the cross-section of tradable portfolio returns, and (2) media factors can explain the cross-section of individual commodity futures returns. The two-level cross-section tests were also used in Boons and Prado (2019). Boons and Prado (2019) note that investors also often favor longer time-to-maturity contracts because they want to hedge risk and carry out rolling strategies. For these reasons, the authors suggest using both nearby and spreading returns of test assets in the cross-section tests. We also employ this approach in our tests.

4.1. Cross-section test at the portfolio level

We begin with a cross-section analysis at the portfolio level. The test portfolios are constructed in two ways. To begin, we divide 26 commodities into three portfolios sorted by their basis, momentum, basis-momentum, media emotion intensity, and media coverage. This first way generates 15 portfolios. We obtain 30 portfolio return series after segregating nearby and spreading returns. Second, we categorize 26 commodities into six categories: energy, grains, metals, soft commodities, food oils, and livestock. This second way generates 12 portfolio return series by considering both the nearby and spreading returns of these six commodity-type portfolios. We finally have 42 portfolio test return series. To conduct the two-stage cross-section tests at the portfolio level, we begin by regressing each portfolio return series against the factor set in each model to determine the exposures of each portfolio to the factors. This time-series regression is conducted for the whole sample from February 1998 to February 2020. Therefore, the exposures (betas) of portfolios in the first stage are all time-invariant. In the second stage, we regress the average return of each portfolio return series to their exposures, estimating the premium of each factor.

Table 11 reports the result of the portfolio-level cross-section tests. In this panel, we present the estimated premium corresponding to each factor. The number in the parenthesis below each estimated premium is the t-stat from the OLS estimation. We further correct the standard errors for the errors-in-variables problem in the first-stage regression following Shanken (1992). The adjusted t-stat by Shanken (1992) are displayed in the square brackets underneath the OLS t-statistic. The first two models show that the explaining ability of factors from Bakshi et al. (2017) is greater than those suggested by Szymanowska et al. (2014) with around 9% higher in R^2 of Model (2) compared with Model (1). Across these two models, only the basis momentum presents a consistent risk price of around 1% a month (equivalent to 12% a year). The Shanken t-statistics of the basis-momentum factor in these models are around 1.75. We do not expect that t-stat in our test is as high as in Boons and Prado (2019) because our sample covers a much shorter period than in the earlier study. Basis nearby also shows a significant premium in Model (1) but loses its significance in Model (2), suggesting that the basis premium might overlap with the risk from other factors in Model (2).

We examine first the tests for media emotion intensity factors in Models (3), (4), (5) and (6). Across these models, the emotion nearby factor shows a very consistent premium at around 1.2% a month (equivalent to around 14.5% a year) and all of the estimated premiums of this factor are significant with Shanken t-statistics greater than 2. These results imply that the media emotion intensity significantly predicts the cross-section of tradable portfolio returns. The premium of this factor is also slightly higher than that of basis-momentum. From Models (3) and (4), basis-momentum shows a consistent risk price at around 1.1% a month (13% a year). However, the significance of the basis-momentum nearby factor is weaker than that of media emotion intensity nearby, with the Shanken t-stat around 1.7. The significance of basis-momentum is not likely to connect to media emotion intensity as both the estimated premium and Shanken t-stat of basis-momentum nearby are remarkably consistent across all ten models. We do not observe the significant premium of the media emotion intensity spreading factor because the corresponding Shanken t-stat stands only around 1. Although the intercepts are significant in two of four models with media emotion intensity factors, they are just around 0.1%, which transfer only around 1.2% a year to portfolios' returns. As for media coverage, we also did not find significant evidence for the premium of media coverage nearby factor. From Models (7) to (10), this factor shows a consistent estimate of its premium at around -0.2% a month (about -2.4% a year). However, both OLS and Shanken t-stat are very small (only around 0.5), suggesting that the media coverage premium is not significant and very small when adding to the key benchmarks set. This leaves media coverage very weak ability to add more predictive value to the cross-section of tradable portfolio returns. The media coverage spreading return only provides a significant premium in

Model (10), while it does not significantly explain the cross-section of portfolio returns in Model (8). In summary, only media emotion intensity nearby, basis-momentum nearby and momentum nearby can provide consistent estimated premiums across all models and survive the threshold for significance based on Shanken t-stat. Further, media emotion intensity nearby factor is the most significant factor, which offers the best explaining ability for the cross-section of tradable portfolio returns.

4.2. Cross-section test at the commodity level

Next, we test the media factors for explaining the cross-section of individual commodity returns. Boons and Prado (2019) conduct this test to provide more solid evidence on the role of basis-momentum factors. This test addresses the arguments of Lewellen et al. (2010) and Ang et al. (2020) about the necessity of testing factors for individual asset returns rather than portfolio returns. Boons and Prado (2019) also overcome the challenge posed by Daskalaki et al. (2014) that predictability of a factor should be tested in individual commodities, although it is difficult to survive this test. We also employ this test to examine further whether media emotion intensity can predict the cross-section of individual commodity returns. For this test, we employ the two-stage Fama Macbeth regression. In the first stage, we regress the nearby and spreading returns of 26 commodity futures on the factor set. We run the rolling time series regression with a 250-day look-back. The reason for selecting the one-year window for daily returns compared to the six-year window for monthly returns is noted in Boons and Prado (2019) that the one-year look back helps estimate factor betas in a more timely manner. By this means, we obtain the series of time-varying exposures of assets to factors. In the second stage, we regress the individual commodity nearby and spreading returns on their exposures to the factors.

Table 12 reports the results of the cross-section test at the commodity level. In this panel, we first show the estimated premium for the factors used in each model. There are two values of t-stat displayed for each estimation. First, we correct standard errors to address autocorrelation using the Newey West method with one lag. We report the t-stat for Newey West standard error in the parenthesis under the estimated risk price. We also correct standard errors using Shanken (1992) method. The Shanken t-stat are reported in the square brackets. The results show that no benchmarks except for basis-momentum can survive this test in our sample period. Across all models, the basis-momentum premium priced in the cross-section of individual commodity returns is 1.5% a month (around 18% a year) on average. We do not find evidence for the significant premium of the basis-momentum spreading factor in this test.

[Table 12 here]

The explaining ability of the media emotion intensity is revealed in Models (3), (4), (5), (6). Across these models, the estimated premiums of this factor are consistent with the estimated value of around 1.3% a month (approx. 15% a year). These estimations all survive the threshold for being significant using the Shanken t-stat. However, for the individual commodity level, the explaining ability of the media emotion intensity is slightly lower than that of the basis-momentum factor. Further, the significance of the media emotion intensity nearby premium is weaker than that of the basis-momentum. We do not observe any significant premiums in Models (4) and (6) for the media emotion intensity spreading factor. The evidence for media coverage is also not consistent across the models. While we find that media coverage can explain the cross-section of individual commodity returns in the last two models, it is not the case for Models (7) and (8). Collecting all results in this test, we observe that only basis-momentum and media emotion intensity can survive the cross-section test at the commodity level.

4.3. LASSO regression results to select factors

Traditionally, cross-section tests have been conducted on a predefined set of factors, and the role of a new factor has been examined by supplementing the model with key benchmarks from the literature. The consistency and significance of the estimated premium for a new factor are the evidence for their ability to explain the cross-section of returns. In this study, so far, we have proved that the premium of the media emotion intensity is significant and consistent at around 1.2% a month (around 14% annualized). This is not the case for basis and momentum factors when they lose their predictive power in the presence of other factors, implying that a factor's ability to explain returns may vanish when the sample changes or when a different set of factors is used. Therefore, the model specification might be sensitive to the sample.

Feng et al. (2020) propose a two-pass regression with a double-selection LASSO technique to select the most relevant factors for the cross-section test. This method is suitable for the equity market because the set of factors suggested by the literature is abundant and selecting the right factors is unquestionably essential. Their method first conducts the LASSO cross-section regression of the average returns against the covariances between the returns and each factor. This step will identify the most significant factors (or the primary factors) that contribute to explaining the cross-section of returns. Other factors are omitted because they have a low connection with portfolio returns. However, exposures to these factors may be highly related to exposures to the selected factors. Hence, eliminating these low-correlated factors may result in an estimation bias for the premiums of the primary factors.

Owing to the limited number of common benchmark factors, we can conveniently conduct the cross-section test with different specifications for the commodity futures market. However, the Feng et al. (2020) approach suggests an interesting test for the relevance of factors. We also employ the same LASSO regression as seen in the first step of their research to select the factors with the predictive ability for the cross-section of returns in the commodity futures market. To test the consistency of factor performance in different periods and for different subsets of test portfolios, we create a random set of 200 samples from our sample with 75% observations. After that, we run the LASSO regression of the average returns of the portfolios on the covariances between the returns of these portfolios and factors. For the cross-section test, each sample will create a unique set of preferred factors. The frequency with which a factor is selected indicates how much it is favored in the cross-section test.

We conduct the LASSO regression with three options. The first approach is standard LASSO regression, which optimizes the CV function to identify the most relevant factors to the cross-section of returns. However, the standard version of LASSO may retain some factors in the model with extremely low loadings. Therefore, we consider the second type of LASSO model, adaptive LASSO, introduced by Zou (2006). This model will eliminate factors with modest coefficients, leaving only predictors with a significant contribution to the return explanation. The third alternative is the LASSO plug-in model (Belloni et al., 2012). The plug-in LASSO regression tends to select the covariates with the highest correlation and the best fit to the out-of-sample test. We expect that the adaptive LASSO will select fewer factors than the standard LASSO, while the plug-in LASSO will select the fewest.

Table 13 shows the relative frequencies of each factor selected in the test for 200 randomized samples and three LASSO models. Furthermore, we independently generated a different set of 200 random samples for each type of LASSO model. The result of the standard LASSO model indicates that the media emotion intensity (nb) factor is selected in 100% of the randomly generated samples, followed by the basis-momentum (nb) factor at a 72% rate of selection. Interestingly, the media emotion intensity (nb) factor is again selected at a high rate of 99.5% for the adaptive LASSO model. The second factor chosen most frequently is also the basis momentum, at 55% of the time. According to the first two types of LASSO regression, emotion is by far the best predictor of the cross-section portfolio returns. The LASSO plug-in will determine which factors are the most predictors of out-of-sample returns. Again, the emotion (nb) factor is the most frequently selected at 84.5%. Similarly, basis-momentum is chosen in 84% of samples. Further, the plug-in LASSO result indicates that the average, basis (nb), and momentum (nb) along with media emotion intensity (nb) and basis-momentum (nb) comprise the set of the best predictors for the portfolio-level cross-section of returns. This selection is comparable to the Boons and Prado (2019) benchmark factors, with the addition of our media emotion intensity factor. In summary, the LASSO test results validate the efficiency of media emotion intensity factors in predicting the cross-section of commodity futures returns.

4.4. Media emotion intensity premium and other risk sources

Some hypotheses have been put forward in the literature to explain premium on the commodity futures market. Given the statistical significance of the media emotion intensity premium in the cross-section test with the benchmark factors, we are interested in testing whether other risk sources subsume this premium. To facilitate these tests, we also employ the two-state Fama-Macbeth regression for the intensity of media emotions and each risk factor. We test the survival of the media emotion intensity premium against hedging pressure, inventory, market volatility, market liquidity, equity market risk, and macroeconomic factors. In this subsection, we focus only on reporting the consistency of the media emotion intensity premium across the tests and leave a detailed discussion of each risk source and corresponding test in Appendix A4.

As the double-sorting results suggest a possible link between hedging pressure and media emotion intensity, we first check to see if hedging pressure can explain the media emotion intensity premium. The predictability of the hedging pressure for commodity futures prices is backed up by the theory of normal backwardation (Keynes (1930)). Accordingly, hedgers trade in the futures market to hedge for spot price fluctuation. We follow Kang et al. (2020) to use the 12-month moving average hedging pressure in the cross-section test to reflect the true hedging pressure in the market. We begin with the time-series spanning test following the Boons and Prado (2019) approach to regress the return of the High4-Low4 portfolio sorted by media emotion intensity on the High4-Low4 portfolios sorted by hedging pressure and other benchmarks. When adding hedging pressure to the model, the alpha decreases from around 12% annualized to 7.74%, suggesting that the return of the media emotion intensity portfolio is partly related to hedging pressure. However, this abnormal return of 7.74% is still significant and large enough. The result implies that a large proportion of the media emotion intensity portfolio return is not explained by hedging pressure and other benchmark factors. In the cross-section test, the premium of media emotion intensity is statistically significant at 1.2% a month when controlling for the hedging pressure and other benchmarks. This level of premium is similar to the result in the model with only the benchmark factors. This evidence shows that the media emotion intensity premium is not subsumed by hedging pressure.

The storage theory (Kaldor, 1939) suggests a link between commodity inventory level and commodity price. Later, G. B. Gorton et al. (2013) argue that the commodity futures premium is driven by the physical inventory level. According to G. B. Gorton et al. (2013) and Boons and Prado (2019), high inventory level is associated with contango, low momentum, and low basis momentum. Hence, in this study, we use basis, momentum, and basis-momentum to represent the inventory level of commodities. By doing this, we come back with the result of the cross-section test with benchmark factors. As the premium of media emotion intensity is statistically significant around 1.2% a month in the test, we conclude that the premium of media emotion intensity is less likely to be subsumed by the inventory risk.

According to Boons and Prado (2019), market volatility risk is priced in the cross-section of commodity futures returns. The authors establish that market volatility could predict and explain the basis-momentum component. The negative risk premium associated with market volatility shows that investors are willing to pay a premium to compensate for increased volatility. In Appendix A4.3, we present the cross-section test for market and average volatility. Market volatility and average volatility are not significantly priced in our sample's cross-section of commodity futures returns. However, the Shanken t-stat for market volatility is -1.42, close to significance at the 10% level. Notably, when other benchmark variables are added to the model, the risk price of market volatility is reduced to -0.4%, with a t-statistic of -0.09, meaning that almost all market volatility risk is captured by the risk of media emotion intensity and other benchmark factors. Across the models, the media emotion intensity factor is still statistically significant at around 1.2%.

The literature also suggests that market liquidity and funding liquidity can explain the premium of commodity futures (Boons and Prado, 2019; Brunnermeier and Pedersen, 2009. We follow Boons and Prado (2019) use Ted spread as a proxy for funding illiquidity and use VIX as a proxy for market illiquidity. The cross-section test shows that the risks of Ted spread and VIX are significantly priced in the cross-section of commodity futures returns. When adding media emotion intensity into the model, the risk prices of these two liquidity factors lose their significance. However, the premium of media emotion intensity is still statistically significant at around 1.2 a month, suggesting that the market and funding liquidity risk cannot subsume the media emotion intensity premium.

Additionally, Boons and Prado (2019) argue that market risk and downside market risk from the equity market are priced in the cross-section of portfolio returns. If the market returns less than the average minus one standard deviation for the entire sample, the market is considered to be on the downside. A bear equity market may deter speculators from clearing the futures market. We use the excess return on the value-weighted CRSP portfolio to approximate the stock market return. After taking into account the intensity of the emotions in the media in the testing models, the equity market's risk premium is reduced to 2% and is no longer significant (t-stat = 1.20). Similarly, the risk price of the downside equity market also drops to 1.5% (t-stat = 1.4). This conclusion shows that media emotion intensity may partly reflect equity market risk and downside market risk. However, similar to the previous test, the premium associated with the media emotion intensity factor is relatively consistent at 1.2% each month and statistically significant. Internet appendix A5 discusses more about the results on equity and downside equity market risks.

Given that the economic situation affects the supply and demand of commodities, macroeconomic factors are closely related to the commodity market (Le Pen and Sévi, 2018; Szymanowska et al., 2014). Owing to the distinct characteristics of the different commodities, each commodity may react differently to macroeconomic shocks. We collect 10 macroeconomic variables that reflect various facets of the United States economy. Although several of these factors exhibit high correlations, they are not identical. We begin by analyzing these macro variables' principal components to isolate the primary information-containing elements. Factor 1 is mainly connected with CPI, PPI, and M2 supply. Consumer sentiment and unemployment are reflected in Factor 2. Factor 3 is associated with the M1 supply and the US dollar index, which measures the strength of the US dollar. The cross-section test result is provided in Appendix Table A9. Only the component associated with consumer confidence and the labor market can significantly explain our sample's cross-section of portfolio returns. When the media emotion intensity factor is included in the model, macroeconomic variables' risk prices and significance of macroeconomic variables are significantly reduced. The premium associated with media emotion intensity again remains statistically significant at around 1.2%. This finding supports the hypothesis that media emotion intensity remains a reliable and consistent predictor of commodity futures returns.

5. Conclusion

Our paper studies the role of a new factor, media emotion intensity, in commodity futures pricing. Media emotion intensity measures the excess proportion of emotional content relative to factual content scaled by the total amount of media coverage. We construct our factor based on this measure and establish its predictive power to commodity futures returns. First, the long-short portfolio sorted by media emotion intensity generates an average annualized return of 13.51% (t-stat = 2.95). Second, we show that the factor premium of media emotion intensity still remains significant after we control for benchmark factors, such as basis, momentum and basis-momentum. Third, we investigate the long-short portfolio returns double-sorted by media emotion intensity and other factors and show that media emotion intensity only exerts a strong influence on returns of commodities with backwardation, low media coverage, high momentum, and high basis-momentum. Fourth, we conduct the cross-section Fama-Macbeth tests and show that the media emotion intensity premium remains statistically and economically significant even after controlling for benchmark risk factors, including basis, momentum and basis-momentum. Finally, our standard, adaptive and plug-in LASSO test results rank the media emotion intensity as the most relevant factor amongst the considered benchmark factors.

Moreover, we also consider other risk channels such as hedging pressure, funding liquidity, market liquidity, macroeconomic risks, market volatility, average volatility, inventory, equity market risk and downside equity market risk. However, we find that none of these factors can subsume the premium of media emotion intensity. The media emotion intensity premium is also statistically and significant at around 1.2% a month or approximately 14% a year across the cross-section test with the benchmark factors and other considered risk sources. This result suggests that media emotion intensity is a reliable and consistent predictor for the cross-section of commodity futures return.

Given the strong impact of media emotion intensity on commodity futures return, we also leave the question of how and by what ways media emotion intensity can affect the investor behaviors for future research. Future studies should continue to expand our understanding on how each type of market participants, such as hedgers or speculators, reacts to media emotion intensity.

References

Adams, Z., & Glück, T. (2015). Financialization in commodity markets: A passing trend or the new normal? *Journal of Banking & Finance*, *60*, 93–111. https://doi.org/10.1016/j.jbankfin.2015.07.008

Ang, A., Liu, J., & Schwarz, K. (2020). Using Stocks or Portfolios in Tests of Factor Models. *Journal of Financial and Quantitative Analysis*, *55*(3), 709–750. https://doi.org/10.1017/S0022109019000255

Bakshi, G., Gao, X., & Rossi, A. G. (2017). Understanding the Sources of Risk Underlying the Cross Section of Commodity Returns. *Management Science*. https://doi.org/10.1287/mnsc.2017.2840

Barillas, F., & Shanken, J. (2017). Which Alpha? *The Review of Financial Studies*, *30*(4), 1316–1338. https://doi.org/10.1093/rfs/hhw101

Barillas, F., & Shanken, J. (2018). Comparing Asset Pricing Models. *The Journal of Finance*, *73*(2), 715–754. https://doi.org/10.1111/jofi.12607

Bas, O., & Grabe, M. E. (2015). Emotion-Provoking Personalization of News: Informing Citizens and Closing the Knowledge Gap? *Communication Research*, *42*(2), 159–185. https://doi.org/10.1177/0093650213514602

Beckett, C., & Deuze, M. (2016). On the Role of Emotion in the Future of Journalism. *Social Media + Society*, *2*(3), 205630511666239. https://doi.org/10.1177/2056305116662395

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica*, *80*(6), 2369–2429. https://doi.org/10.3982/ECTA9626

Belousova, J., & Dorfleitner, G. (2012). On the diversification benefits of commodities from the perspective of euro investors. *Journal of Banking & Finance*, *36*(9), 2455–2472. https://doi.org/10.1016/j.jbankfin.2012.05.003

Boons, M., & Prado, M. P. (2019). Basis-Momentum. *The Journal of Finance*, *74*(1), 239–279. https://doi.org/10.1111/jofi.12738

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market Liquidity and Funding Liquidity. *Review of Financial Studies*, *22*(6), 2201–2238. https://doi.org/10.1093/rfs/hhn098

Daskalaki, C., Kostakis, A., & Skiadopoulos, G. (2014). Are there common factors in individual commodity futures returns? *Journal of Banking & Finance*, *40*, 346–363. https://doi.org/10.1016/j.jbankfin.2013.11.034

Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking & Finance*, *35*(10), 2606–2626. https://doi.org/10.1016/j.jbankfin.2011.02.022

De Roon, F. A., Nijman, T. E., & Veld, C. (2000). Hedging Pressure Effects in Futures Markets. *The Journal of Finance*, *55*(3), 1437–1456. https://doi.org/10.1111/0022-1082.00253

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life* (2nd Owl Books ed). Owl Books
Includes bibliographical references (p. 263-280) and index.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607–636. https://doi.org/10.1086/260061

Fang, L., & Peress, J. (2009). Media Coverage and the Cross-section of Stock Returns. *The Journal of Finance*, *64*(5), 2023–2052. https://doi.org/10.1111/j.1540-6261.2009.01493.x

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, *75*(3), 1327–1370. https://doi.org/10.1111/jofi.12883

Fernandez-Perez, A., Frijns, B., Fuertes, A.-M., & Miffre, J. (2018). The skewness of commodity futures returns. *Journal of Banking & Finance*, *86*, 143–158. https://doi.org/10.1016/j.jbankfin.2017.06.015

García, D. (2013). Sentiment during Recessions. *The Journal of Finance*, *68*(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, *57*(5), 1121–1152. https://doi.org/10.2307/1913625

Goldenberg, A., & Gross, J. J. (2020). Digital Emotion Contagion. *Trends in Cognitive Sciences*, *24*(4), 316–328. https://doi.org/10.1016/j.tics.2020.01.009

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and Fantasies about Commodity Futures. *Financial Analysts Journal*, *62*(2), 47–68. https://doi.org/10.2469/faj.v62.n2.4083

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The Fundamentals of Commodity Futures Returns. *Review of Finance*, *17*(1), 35–105. https://doi.org/10.1093/rof/rfs019

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and Consumer Decision Making: The Appraisal-Tendency Framework. *Journal of Consumer Psychology*, *17*(3), 158–168. https://doi.org/10.1016/S1057-7408(07)70023-2

Harvey, C. R., Liu, Y., & Zhu, H. (2016). . . . and the Cross-Section of Expected Returns. *Review of Financial Studies*, *29*(1), 5–68. https://doi.org/10.1093/rfs/hhv059

Hong, H., & Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, *105*(3), 473–490. https://doi.org/10.1016/j.jfineco.2012.04.005

Kaldor, N. (1939). Speculation and Economic Stability. *The Review of Economic Studies*, *7*(1), 1–27. https://doi.org/10.2307/2967593

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. https://doi.org/10.1111/jofi.12845

Keltner, D., & Lerner, J. S. (2010). Emotion. In *Handbook of social psychology, Vol. 1, 5th ed* (pp. 317–352). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy001009

Keltner, D., Oatley, K., & Jenkins, J. M. (2019). *Understanding emotions* (Fourth edition). John Wiley & Sons, Inc
Includes index.

Keynes, J. M. (1930). *Treatise on money* (Vol. 2). Macmillan. Retrieved May 26, 2022, from https://scholar.google.com/scholar_lookup?title=treatise+on+money&author=Keynes%2C+John+Maynard&publication_year=1930

Le Pen, Y., & Sévi, B. (2018). Futures Trading and the Excess Co-movement of Commodity Prices. *Review of Finance*, *22*(1), 381–418. https://doi.org/10.1093/rof/rfx039

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lettau, M., Maggiori, M., & Weber, M. (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics*, *114*(2), 197–225. https://doi.org/10.1016/j.jfineco.2014.07.001

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 20. https://doi.org/10.1016/j.jfineco.2009.09.001

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292. https://doi.org/10.1006/obhd.1996.0028

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, *92*(368), 805–824. https://doi.org/10.2307/2232669

Ludvigson, S. C. (2004). Consumer Confidence and Consumer Spending. *Journal of Economic Perspectives*, *18*(2), 29–50. https://doi.org/10.1257/0895330041371222

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2012). Commodity Liquidity Measurement and Transaction Costs. *Review of Financial Studies*, *25*(2), 599–638. https://doi.org/10.1093/rfs/hhr075

Murry, J. P., Jr., & Dacin, P. A. (1996). Cognitive Moderators of Negative-Emotion Effects: Implications for Understanding Media Context. *Journal of Consumer Research*, *22*(4), 439. https://doi.org/10.1086/209460

Nagel, S. (2012). Evaporating Liquidity. *Review of Financial Studies*, *25*(7), 2005–2039. https://doi.org/10.1093/rfs/hhs066

Orgeret, K. S. (2020). Discussing Emotions in Digital Journalism. *Digital Journalism*, *8*(2), 292–297. https://doi.org/10.1080/21670811.2020.1727347

Peters, C. (2011). Emotion aside or emotional side? Crafting an 'experience of involvement' in the news. *Journalism*, *12*(3), 297–316. https://doi.org/10.1177/1464884910388224

Pham, M. T. (2007). Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence. *Review of General Psychology*, *11*(2), 155–178. https://doi.org/10.1037/1089-2680.11.2.155

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50*(4), 703–712. https://doi.org/10.1037/0022-3514.50.4.703

Sakkas, A., & Tessaromatis, N. (2020). Factor based commodity investing. *Journal of Banking & Finance*, *115*, 105807. https://doi.org/10.1016/j.jbankfin.2020.105807

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433–440. https://doi.org/10.1080/026999300402745

Shanken, J. (1992). On the Estimation of Beta-Pricing Models. *Review of Financial Studies*, *5*(1), 1–33. https://doi.org/10.1093/rfs/5.1.1

Smales, L. A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, *49*, 275–286. https://doi.org/10.1016/j.jbankfin.2014.09.006

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life*. Hackett Pub. Co.

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. https://doi.org/10.1111/jofi.12096

Tang, K., & Xiong, W. (2012). Index Investment and the Financialization of Commodities. *Financial Analysts Journal*, *68*(6), 54–74. https://doi.org/10.2469/faj.v68.n6.5

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, *63*(3), 1437–1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Uribe, R., & Gunter, B. (2007). Are 'Sensational' News Stories More Likely to Trigger Viewers' Emotions than Non-Sensational News Stories?: A Content Analysis of British TV News. *European Journal of Communication*, *22*(2), 207–228. https://doi.org/10.1177/0267323107076770

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Wahl-Jorgensen, K. (2020). An Emotional Turn in Journalism Studies? *Digital Journalism*, *8*(2), 175–194. https://doi.org/10.1080/21670811.2019.1697626

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735

Tables

Table 1: **Descriptive statistics of commodity futures returns**

This table provides the descriptive statistics for the returns of 26 commodity futures in the study. The first three columns show the name of commodities corresponding to their futures, the type of the commodities and the commodity tickers in the cmdty dataset (formerly CRB data). The last four columns report the average and standard deviation of nearby return and spreading return of each commodity futures for the period from January 1998 to February 2020. The nearby return is defined as the return of the long position on the first nearby futures contract. The spreading return, on the other hand, is defined as the return of simultaneously holding long position on the first nearby futures contract and short position on the second nearby futures contract.

Asset Name	Type	CRB ticker	Nearby return		Spreading return	
			Average	St. dev.	Average	St. dev.
Aluminum	Metals	AL	-4.36%	14.53%	0.83%	1.20%
Brent Crude	Energy	QA	8.88%	30.64%	-0.53%	2.08%
Canola	Food oil	RS	-2.26%	20.51%	-2.03%	3.40%
Cattle	Livestocks	FC	1.01%	15.69%	-2.40%	2.88%
Cocoa	Softs	CC	2.16%	29.87%	-0.79%	2.62%
Coffee	Softs	KC	-8.75%	31.50%	-1.18%	2.38%
Copper	Metals	HG	7.82%	25.93%	-0.17%	1.07%
Corn	Grains	C	-6.89%	26.65%	-3.03%	3.26%
Cotton	Softs	CT	-4.82%	27.60%	-3.46%	5.30%
Crude Oil	Energy	CL	6.87%	31.27%	-1.26%	2.48%
Ethanol	Energy	AK	17.54%	30.50%	5.11%	7.75%
Gasoline	Energy	RB	14.22%	33.90%	1.06%	5.98%
Gold	Metals	GC	6.60%	16.49%	-0.05%	0.26%
Heating Oil	Energy	HO	6.83%	31.27%	-0.54%	3.22%
Hogs	Livestocks	LH	-9.55%	26.34%	-8.68%	9.80%
Natural Gas	Energy	NG	-10.77%	42.87%	-4.49%	9.91%
Orange Juice	Softs	OJ	-2.08%	29.71%	-0.24%	3.45%
Palladium	Metals	PA	17.81%	35.50%	0.74%	3.77%
Palm Oil	Food oil	CU	-4.98%	21.64%	0.10%	2.58%
Platinum	Metals	PL	7.05%	21.99%	0.77%	1.84%
Rough Rice	Grains	RR	-9.47%	23.81%	-3.22%	5.36%
Silver	Metals	SI	6.27%	29.40%	-0.31%	0.46%
Soybean Oil	Food oils	BO	-2.01%	25.17%	-0.97%	1.50%
Soybeans	Grains	S	4.69%	24.88%	-0.47%	2.48%
Sugar	Softs	SB	0.82%	31.93%	-1.27%	7.41%
Wheat	Grains	W	-8.56%	28.98%	-2.84%	3.36%

Table 2: **Descriptive statistics of media factors**

The following table contains descriptive statistics on media emotion intensity, media coverage and media news sentiment. Media emotion intensity quantifies the media intensity of emotion content towards an asset in comparison to its factual content. The higher the media emotion intensity value, the more emotional content dominates the asset's factual commentary. Media coverage indicates how much content in the news is directly relevant to an item. Media news sentiment measures the tone of media news that takes a positive values when the aggregated tone of media news in the month is positive and take a negative values when the aggregated tone of media news in the month is negative. Mean, standard deviation, min and max values are reported for each commodity. The sample covers the period from January 1998 to February 2020.

Commodity	Ticker	Media emotion intensity				Media coverage				News sentiment			
		Mean	St.dev	Min	Max	Mean	St. dev	Min	Max	Mean	St.dev	Min	Max
Ethanol	AK	0.391	0.058	0.239	0.532	0.955	0.548	0.200	3.200	-0.014	0.059	-0.185	0.152
Aluminum	AL	0.422	0.045	0.323	0.555	3.803	3.192	1.515	25.737	-0.055	0.062	-0.211	0.047
Soybean oil	BO	0.232	0.172	-0.415	0.833	0.064	0.065	0.001	0.309	-0.068	0.124	-0.370	0.309
Corn	C	0.196	0.048	-0.019	0.307	3.704	2.538	0.657	24.243	-0.066	0.035	-0.160	0.012
Cocoa	CC	0.311	0.084	0.075	0.496	1.762	0.862	0.273	4.616	-0.041	0.055	-0.168	0.130
Crude oil	CL	0.378	0.040	0.251	0.466	41.985	34.210	3.074	185.962	-0.118	0.041	-0.265	0.002
Cotton	CT	0.263	0.070	0.030	0.424	1.804	1.097	0.365	7.156	-0.058	0.035	-0.168	0.043
Palm oil	CU	0.354	0.043	0.270	0.448	1.570	0.673	0.666	4.107	-0.096	0.053	-0.226	0.023
Cattle	FC	0.309	0.086	0.087	0.459	3.940	2.540	0.793	13.804	-0.128	0.036	-0.223	0.011
Gold	GC	0.413	0.036	0.284	0.477	13.839	11.494	0.604	68.075	-0.037	0.048	-0.173	0.077
Copper	HG	0.335	0.075	0.175	0.490	5.178	3.639	0.635	17.039	-0.082	0.055	-0.214	0.109
Heating oil	HO	0.312	0.097	0.023	0.566	0.526	0.501	0.058	2.887	-0.144	0.093	-0.423	0.094
Coffee	KC	0.342	0.086	0.145	0.508	2.891	1.440	0.663	6.526	-0.036	0.050	-0.145	0.076
Hogs	LH	0.166	0.066	0.011	0.366	0.749	0.311	0.280	2.022	-0.127	0.044	-0.224	0.001
Natural gas	NG	0.366	0.036	0.271	0.463	8.052	6.575	0.568	27.704	-0.080	0.049	-0.243	0.023
Orange juice	OJ	0.263	0.139	-0.064	0.661	0.208	0.171	0.021	1.607	-0.103	0.069	-0.380	0.128
Palladium	PA	0.457	0.123	0.180	0.755	0.404	0.323	0.007	2.020	-0.071	0.144	-0.474	0.309
Platinum	PL	0.446	0.095	0.131	0.670	1.251	1.444	0.051	12.551	-0.011	0.094	-0.321	0.220
Brent crude	QA	0.425	0.112	0.178	0.688	1.127	1.072	0.075	6.456	-0.155	0.101	-0.443	0.086
Gasoline	RB	0.294	0.047	0.166	0.419	5.864	4.510	0.473	27.813	-0.111	0.048	-0.264	0.056
Rough rice	RR	0.313	0.102	0.001	0.502	3.388	2.937	0.245	25.639	-0.068	0.036	-0.161	0.033
Canola	RS	0.203	0.084	-0.045	0.462	0.459	0.243	0.069	1.349	-0.110	0.063	-0.253	0.038
Soybean	S	0.186	0.051	0.051	0.368	3.554	1.791	0.513	14.387	-0.071	0.036	-0.183	0.020
Sugar	SB	0.298	0.072	0.151	0.456	3.613	1.971	0.489	8.681	-0.072	0.028	-0.158	-0.009
Silver	SI	0.448	0.056	0.239	0.544	2.324	2.214	0.068	18.144	-0.021	0.072	-0.202	0.175
Wheat	W	0.243	0.044	0.136	0.372	4.718	1.943	1.093	14.847	-0.087	0.037	-0.185	0.015

Table 3: **Commodity portfolios sorted on factors**

This table reports the unconditional performance of the portfolios sorted by media emotion intensity, media coverage, news sentiment and other key benchmarks: basis, momentum and basis-momentum. We evaluate portfolio performance in terms of both nearby and spreading returns. For each factor, we present the nearby returns in the upper half of the table and the spreading return in the lower half. When sorting on a factor, we divided commodity futures into three portfolios: High4 (four commodity futures having the highest values of the signal), Low4 (four commodity futures having the lowest values of the signal) and Mid (the remaining commodity futures which were available at the end of the month). High4 - Low4 is the portfolio of longing on the High4 portfolio and shorting on the Low4 portfolio. The nearby return of a portfolio is calculated as the equal-weighted average of the returns of the first nearby contracts in this portfolio. The spreading return is the difference between the equal-weighted average of the returns of the first and second nearby contracts. The returns are computed for the period beginning in February 1998 and ending in February 2020. Panels A, B and C illustrate the performance of portfolios sorted by media emotion intensity, media coverage and news sentiment, respectively. Panel D summarizes the performance of High4-Low4 portfolios sorted by basis, momentum and basis-momentum.

	Panel A: Univariate sort on media emotion intensity			
	High 4	Mid	Low 4	High4 - Low 4
	Nearby returns			
Ave. Ret	11.16%	0.19%	-2.35%	13.51%
t-stat	2.62	0.06	-0.65	2.95
Sharpe	0.56	0.01	-0.14	0.63
	Spreading returns			
Ave. Ret	0.11%	-1.20%	-3.24%	3.35%
t-stat	0.40	-4.63	-5.16	4.97
Sharpe	0.09	-0.99	-1.10	1.06
	Panel B: Univariate sort on media coverage			
	High 4	Mid	Low 4	High4 - Low 4
	Nearby returns			
Ave. Ret	-3.55%	1.79%	5.62%	-9.16%
t-stat	-0.92	0.62	1.44	-2.69
Sharpe	-0.20	0.13	0.31	-0.57
	Spreading returns			
Ave. Ret	-2.16%	-1.44%	-0.04%	-2.11%
t-stat	-4.37	-5.27	-0.14	-3.88
Sharpe	-0.93	-1.12	-0.03	-0.82
	Panel C: Univariate sort on new sentiments			
	High 4	Mid	Low 4	High4 - Low 4
	Nearby returns			
Ave. Ret	6.25%	0.54%	0.65%	5.60%
t-stat	1.37	0.18	0.15	1.13
Sharpe	0.29	0.04	0.03	0.24
	Spreading returns			
Ave. Ret	0.10%	-1.79%	-0.62%	0.72%
t-stat	0.12	-6.80	-1.31	0.76
Sharpe	0.03	-1.45	-0.28	0.16

Panel D: Univariate sorts on other factors			
	Basis	Momentum	Basis-Momentum
	High4 - Low 4	High4 - Low 4	High4 - Low 4
Nearby returns			
Ave. Ret	-5.49%	9.57%	12.50%
t-stat	-1.17	1.71	2.66
Sharpe	-0.25	0.36	0.57
Spreading returns			
Ave. Ret	1.47%	-1.18%	4.64%
t-stat	1.41	-1.24	4.78
Sharpe	0.30	-0.26	1.02

Table 4: **Time-series spanning test for media factors**

This table presents the results of the time-series spanning test for four media factors. We define four media factors constructed by the nearby and spreading returns of the High4 - Low4 portfolio sorted by media emotion intensity and media coverage . The note "(nb)" and "(sp)" after the name of each factor correspond to nearby and spreading returns, respectively. Regarding benchmark factors, we employ six factors suggested by Szymanowska et al. (2014), Bakshi et al. (2017), and Boons and Prado (2019). Szymanowska et al. (2014) introduced three factors: the nearby return of the High - Low portfolio sorted by basis and the spreading returns of the High and the Low portfolios sorted by basis. To match with Boons and Prado (2019), we use the nearby return of the High4 - Low4 portfolio sorted by basis and the spreading returns of the High4 and Low4 portfolio sorted by basis to represent the three factors from Szymanowska et al. (2014). We also use the nearby average return of commodity futures (average (nb)), the nearby return of High4 - Low4 portfolio sorted by momentum to represent two factors suggested by Bakshi et al. (2017). Finally, we use the nearby return and spreading return of portfolios sorted by basis-momentum as the last two factors introduced by Boons and Prado (2019). Panel A shows the results of the spanning tests for the media factors versus the factors from Boons and Prado (2019) and Bakshi et al. (2017). Panel B shows the results of the spanning tests using the factors suggested by Szymanowska et al. (2014) and two factors introduced by Boons and Prado (2019). The standard errors of coefficients are corrected using the Newey West method with one lag. Coefficients with "***", "**", and "*" are statistically significant at 1%, 5% and 10% levels, respectively. The results are estimated for the sample from February 1998 to February 2020.

Panel A: Factors from Boons and Prado (2019) and Bakshi et al. (2017)				
	Media emotion		Media coverage	
Model	(1)	(2)	(3)	(4)
	Nearby return	Spreding return	Nearby return	Spreding return
Alpha	0.1135**	0.0243***	-0.0740**	-0.0193***
t-stat (Alpha)	(2.3020)	(3.2127)	(-2.2221)	(-2.7806)
Basis-Momentum (nb)	0.0260	-0.0064	-0.1068*	-0.0017
Basis-Momentum (sp)	-0.2682	0.2152***	-0.2274	-0.0491
Basis (nb)	-0.1103	-0.0080	-0.0093	-0.0113
Momentum (nb)	0.1044*	0.0086	0.0013	-0.0029
Average (nb)	0.3020***	0.0062	-0.0144	0.0194
Coverage (nb)	0.0716	0.0193		
Coverage (sp)	-0.7817	-0.0211		
Emotion (nb)			-0.0050	-0.0108
Emotion (sp)			0.5111	0.0363
R-squared	0.0899	0.1096	0.0373	0.028
GRS-F	15.0534		12.4255	
p-value	0.0005		0.0020	
Panel B: Factors from Boons and Prado (2019) and Szymanowska et al. (2014)				
	Media emotion		Media coverage	
Model	(5)	(6)	(7)	(8)
	Nearby return	Spreding return	Nearby return	Spreding return
Alpha	0.1172**	0.0187**	-0.0819**	-0.0149***
t-stat (Alpha)	(2.2645)	(2.5460)	(-2.4383)	(-2.9980)

Basis-Momentum (nb)	0.0766	-0.0090	-0.1053*	0.0063
Basis-Momentum (sp)	-0.3289	0.1780**	-0.2057	-0.0718
Basis (nb)	-0.1730**	-0.0049	-0.0192	-0.0033
Basis (high4 - sp)	-0.0848	-0.3426***	0.3522	0.0834
Basis (low4 - sp)	-0.2133	-0.1995**	0.1247	0.1800**
Coverage (nb)	0.0583	0.0146		
Coverage (sp)	-0.5703	0.0852		
Emotion (nb)			-0.0099	-0.0103
Emotion (sp)			0.6506*	0.1163*
R-squared	0.0426	0.2422	0.0408	0.0879
GRS-F	11.2556		8.8423	
p-value	0.0036		0.0120	

Table 5: **The percentage of each commodity futures appearing in High4 and Low4 portfolio sorted by factors**

This table displays the percentage of each commodity that appears in the High4 and Low4 groups, sorted by media factors (media emotion intensity, media coverage, and news sentiment) and benchmark factors (basis, momentum, and basis-momentum). Each month's commodity ranking is performed on the last day of the month. For each sort, only the four commodities with the highest factor values (High4 group) and the four commodities with the lowest factor values (Low4 group) are considered (Low4 group). The percentage of appearances is computed by dividing the total number of appearances by the sample period's number of months.

		Basis		Momentum		Basis-Momemtum		Emotion		Coverage		Sentiment	
Asset Name	CRB ticker	Low 4	High 4	Low 4	High 4	Low 4	High 4	Low 4	High 4	Low 4	High 4	Low 4	High 4
Ethanol	AK	34.21%	0.75%	0.00%	20.68%	5.64%	43.61%	0.00%	9.77%	7.89%	0.00%	1.50%	27.07%
Aluminum	AL	2.63%	0.00%	2.26%	4.14%	0.00%	4.14%	0.00%	4.14%	0.00%	1.50%	2.63%	2.26%
Soybean Oil	BO	2.26%	0.00%	9.77%	6.77%	0.75%	5.26%	39.47%	9.40%	100.00%	0.00%	26.32%	32.71%
Corn	C	4.89%	52.26%	20.68%	9.40%	18.80%	1.13%	61.28%	0.00%	0.00%	7.14%	2.63%	8.65%
Cocoa	CC	10.15%	5.64%	25.19%	23.31%	9.77%	12.78%	5.26%	1.50%	0.00%	0.38%	4.14%	28.20%
Crude Oil	CL	21.43%	4.14%	15.79%	21.80%	12.78%	12.03%	0.00%	19.55%	0.00%	100.00%	24.06%	1.13%
Cotton	CT	19.55%	21.43%	22.93%	10.15%	37.59%	15.41%	10.90%	0.00%	0.00%	0.00%	0.38%	12.03%
Palm Oil	CU	4.14%	1.13%	3.76%	0.00%	0.38%	5.64%	0.00%	0.00%	0.00%	0.00%	1.88%	0.38%
Cattle	FC	24.06%	7.89%	8.65%	12.41%	26.69%	12.41%	6.39%	1.50%	0.00%	15.79%	36.09%	1.88%
Gold	GC	6.39%	0.00%	1.50%	16.92%	0.00%	16.92%	0.00%	40.60%	0.00%	74.44%	1.88%	27.44%
Copper	HG	15.04%	0.00%	4.51%	16.17%	0.38%	11.28%	1.50%	4.14%	0.00%	27.82%	10.53%	9.40%
Heating Oil	HO	17.67%	2.26%	17.29%	21.43%	5.64%	19.17%	10.15%	7.89%	45.49%	0.00%	51.50%	11.65%
Coffee	KC	3.76%	46.62%	31.95%	10.15%	7.14%	2.26%	1.13%	6.77%	0.00%	2.63%	1.50%	33.46%
Hogs	LH	29.32%	51.88%	28.95%	7.89%	71.80%	13.53%	78.95%	0.00%	7.14%	0.00%	33.08%	2.26%
Natural Gas	NG	16.54%	27.82%	39.47%	13.16%	47.37%	11.65%	0.00%	10.90%	0.00%	61.28%	8.65%	6.39%
Orange Juice	OJ	14.29%	23.31%	24.81%	22.18%	6.02%	25.19%	32.71%	10.53%	80.45%	0.00%	23.31%	11.28%
Palladium	PA	12.03%	0.38%	15.79%	38.72%	1.13%	21.05%	0.75%	66.92%	71.43%	0.00%	30.83%	35.71%
Platinum	PL	19.92%	0.00%	1.13%	10.53%	0.00%	14.66%	0.00%	66.54%	19.55%	0.38%	6.39%	55.64%
Brent Crude	QA	22.18%	1.50%	16.17%	22.56%	3.01%	16.92%	0.38%	52.63%	19.17%	0.00%	56.77%	7.14%
Gasoline	RB	41.35%	10.53%	5.26%	29.32%	16.92%	37.59%	1.88%	1.50%	0.00%	33.08%	24.81%	6.77%
Rough Rice	RR	7.14%	42.11%	25.56%	6.39%	42.48%	9.77%	9.40%	7.52%	0.00%	5.26%	2.63%	6.77%
Canola	RS	10.90%	11.28%	16.92%	18.42%	19.17%	14.66%	53.38%	0.38%	35.71%	0.00%	30.45%	5.64%
Soybeans	S	19.55%	0.00%	3.01%	15.79%	7.52%	19.17%	60.53%	0.00%	0.00%	28.57%	2.63%	6.39%
Sugar	SB	32.33%	31.58%	23.31%	18.80%	38.35%	32.71%	2.63%	0.00%	0.00%	3.38%	1.50%	6.02%
Silver	SI	5.64%	0.00%	7.14%	16.92%	0.00%	9.77%	0.00%	77.82%	13.16%	0.38%	6.39%	52.26%
Wheat	W	2.63%	57.52%	28.20%	6.02%	20.68%	11.28%	23.31%	0.00%	0.00%	37.97%	7.52%	1.50%

Table 6: **Pairwise average overlapping percentage between factor three-level single-sorting portfolios (High4, Mid, Low4)**

The table reports the similarity in sorting commodities based on three media factors (media emotion intensity, media coverage, and news sentiment) and three benchmark factors (basis, momentum, basis-momentum). First, at the end of each month, the commodities are ranked based on the value of each factor. Second, we form the High4 group incorporates the four commodities with the highest factor values and the Low4 group contains the four commodities with the lowest factor values. Third, for each pair of groups, we calculate the number of commodities that appear in both groups. Finally, the overlapping percentage is calculated by dividing the number by four, the number of commodities in each group.

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Media emotion intensity	High 4	(1)	100.0											
	Low 4	(2)	0.0	100.0										
Media coverage	High 4	(3)	11.9	4.3	100.0									
	Low 4	(4)	27.7	23.8	0.0	100.0								
News sentiment	High 4	(5)	34.1	9.6	8.0	21.5	100.0							
	Low 4	(6)	25.6	15.6	15.8	32.4	0.0	100.0						
Basis	High 4	(7)	3.8	28.2	13.7	7.7	9.6	12.3	100.0					
	Low 4	(8)	12.5	16.4	17.3	12.6	17.1	13.3	0.0	100.0				
Momentum	High 4	(9)	18.0	12.7	16.9	17.3	25.6	11.7	4.2	33.0	100.0			
	Low 4	(10)	10.4	20.5	17.2	15.9	8.8	21.0	42.0	6.1	0.0	100.0		
Basis-momentum	High 4	(11)	16.0	16.4	14.7	17.5	17.4	13.6	6.2	39.0	33.6	5.4	100.0	
	Low 4	(12)	2.5	24.4	18.5	3.6	7.0	17.2	27.3	18.7	13.9	22.7	0.0	100.0

Table 7: **Pairwise average overlapping percentage between factor two-level single-sorting portfolios**

The table reports the similarity in sorting commodities based on two media factors (media emotion intensity and media coverage) and three other factors (basis, momentum, and basis-momentum). First, at the end of each month, the commodities are ranked based on the value of each factor. Second, we divide commodities into two levels, high and low, based on factor value for all factors except for basis. For basis, we sort commodities into two levels: contango and backwardation. Third, for each pair of groups, we calculate the number of commodities that appear in both groups. Finally, the overlapping percentage is calculated by dividing the number by the number of commodities in each pair of groups.

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Emotion	High	(1)	100.0									
	Low	(2)	0.0	100.0								
Coverage	High	(3)	51.8	48.2	100.0							
	Low	(4)	48.2	50.1	0.0	100.0						
Basis	Contango	(5)	50.5	49.4	52.7	47.1	100.0					
	Backwardation	(6)	24.5	25.5	23.3	26.9	0.0	100.0				
Momentum	High	(7)	54.0	46.0	48.4	51.6	41.7	36.6	100.0			
	Low	(8)	45.8	52.2	51.2	46.5	58.0	12.8	0.0	100.0		
Basis-momentum	High	(9)	52.9	44.7	46.5	51.1	45.3	30.6	58.3	39.2	100.0	
	Low	(10)	41.6	54.3	53.1	42.7	53.0	18.8	38.1	58.4	0.0	100.0

Table 8: **Independent double-sorting on Emotion and Other factors**

This table reports the nearby returns of portfolios independently double-sorted by media emotion intensity and other variables. We first report the return with t-statistics of the portfolios single-sorted by each factor in the third and four columns. For media emotion intensity, media coverage, momentum, and basis-momentum, we sort commodities into two portfolios, High and Low, based on the factor's median. For basis, we sort commodities into two portfolios: contango and backwardation. All the portfolios are created by sorting commodities using factor value at the end of each month and then held for one month. The last six columns report the returns of the portfolios sorted by media emotion intensity and each variable. Panel A presents the result of single-sorting commodities into two portfolios based on the median value of media emotion intensity and the result of double-sorting commodities using media emotion intensity and media coverage. Panel B reports the results of double-sorting commodities using media emotion intensity and other factors.

Panel A: Double-sorting on media emotion intensity and media coverage

		Single Sort on Row Variable		Independent sorting on emotion and media coverage					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Emotion	High	4.48	1.31						
(t)	Low	-1.56	-0.51						
	High - Low	6.04	2.10						
Coverage	High	0.73	0.23	0.12	0.03	0.11	0.03	0.01	0.00
(t)	Low	2.42	0.80	9.17	2.34	-3.53	-1.11	12.70	3.39
	High - Low	-1.69	-0.89	-9.05	-3.14	3.64	1.17		

Panel B: Double-sorting on media emotion intensity and other factors

		Single Sort on Row Variable		Independent sorting on emotion and other factors					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Basis	Contango	-0.24	-0.08	0.45	0.12	-3.41	-1.02	3.86	1.11
(t)	Backwardation	7.87	1.86	11.53	2.34	-2.36	-0.48	13.90	2.54
	Conta. - Back.	-8.11	-2.26	-11.08	-2.54	-1.05	-0.22		
Momentum	High	4.94	1.48	8.61	2.22	0.94	0.26	7.67	2.07
(t)	Low	-1.94	-0.61	-1.52	-0.37	-3.92	-1.13	2.40	0.58
	High - Low	6.88	2.30	10.13	2.42	4.86	1.31		
Basis-Momentum	High	6.00	1.77	9.98	2.54	1.13	0.29	8.85	2.38
(t)	Low	-3.27	-1.06	-2.10	-0.53	-4.56	-1.38	2.45	0.61
	High - Low	9.27	3.34	12.08	2.92	5.69	1.64		

Table 9: **Emotion difference in the portfolios sorted by factors**

This table reports the nearby returns of portfolios conditionally double-sorted by media emotion intensity and other variables. More specifically, the commodities are sorted into high and low media emotion intensity groups first and then, within each level of media emotion intensity, we continue to sort the commodities by other factors' value. We first report the return with t-statistics of the portfolios single-sorted by each factor in the third and four columns. For media emotion intensity, media coverage, momentum, and basis-momentum, we sort commodities into two levels, High and Low, based on the factor's median. For basis, we sort commodities into two portfolios: contango and backwardation. All the portfolios are This table reports the difference in media emotion intensity within each portfolio sorted by factors. First, for each factor, we sort commodities into two levels using the value of these factors. Based on the value of basis, we divide commodities into two portfolios of contango and backwardation commodities. Using the median value of each factor, we divide commodities into High and Low portfolios for all other factors. We also separate commodities into two groups, high and low media emotion intensity, based on the median value of this factor. Hence, for a pair of media emotion intensity and each factor, we double-sort commodities into four portfolios using the above single-sorting. The numbers in Columns 3 and 4 are the average media emotion intensity of the portfolios formed by high and low media emotion intensity commodities, respectively, which belongs to each portfolio defined in the first two columns. The final column shows the difference in media emotion intensity between two portfolios in the same row that appear in Columns 3 and 4. "***", "**", and "*" denote statistical significance at the 1%, 5%, and 10% levels, respectively. by sorting commodities using factor value at the end of each month and then held for one month. The last six columns report the returns of the portfolios sorted by each factor within each level single-sorted by media emotion intensity.

		Media emotion intensity		
		High	Low	High - Low
Media coverage	High	0.37	0.25	0.12***
	Low	0.40	0.23	0.17***
Basis	Contango	0.38	0.24	0.15***
	Backwardation	0.38	0.25	0.13***
Momentum	High	0.39	0.25	0.14***
	Low	0.38	0.24	0.14***
Basis-momentum	High	0.39	0.25	0.15***
	Low	0.36	0.24	0.12***

Table 10: **Descriptive statistics of Media factors and Benchmark factors**

This table provides the descriptive statistics of media factors and benchmark factors. We define four media factors constructed by nearby and spreading returns of the High4 - Low4 portfolio sorted by media emotion intensity and media coverage. The note "(nb)" and "(sp)" after the name of each factor correspond to nearby and spreading returns, respectively. Regarding benchmark factors, we employ six factors suggested by Szymanowska et al. (2014), Bakshi et al. (2017), and Boons and Prado (2019). Szymanowska et al. (2014) introduced three factors: the nearby return of the High - Low portfolio sorted by Basis and the spreading returns of the High and the Low portfolios sorted by Basis. To match with Boons and Prado (2019), we use the nearby return of the High4 - Low4 portfolio sorted by Basis and the spreading returns of the High4 and Low4 portfolio sorted by Basis to represent the three factors from Szymanowska et al. (2014). We also use the nearby average return of commodity futures (Average (nb)), the nearby return of High4 - Low4 portfolio sorted by momentum to represent two factors suggested by Bakshi et al. (2017). Finally, we use the nearby return and spreading return of portfolios sorted by Basis-momentum as the last two factors introduced by Boons and Prado (2019). Panel A reports the descriptive statistics of these 11 factors, and Panel B shows the pairwise correlation between them. The results are calculated for the period from February 1998 to February 2020.

Panel A: Descriptive Statistics of Media and Benchmark factors

Factor	Ave. Ret	St. Dev.	Skew.	Kurt.	AR(1)
Emotion (nb)	13.51%	21.51%	-0.92	37.52	0.04
Emotion (sp)	3.35%	3.16%	2.03	74.92	0.04
Coverage (nb)	-9.16%	16.02%	-1.06	45.78	-0.05
Coverage (sp)	-2.11%	2.56%	-5.78	162.07	-0.06
Basis (nb)	-5.49%	22.08%	0.44	35.91	0.01
Basis (h4 - sp)	-1.41%	2.84%	0.78	48.74	-0.02
Basis (l4 - sp)	-2.87%	3.94%	-1.46	77.99	0.08
Average (nearby)	1.65%	13.67%	-1.66	69.84	0.10
Momentum (nb)	9.57%	26.26%	-0.97	41.78	-0.05
Basis-Momentum (nb)	12.50%	22.08%	0.59	39.16	0.00
Basis-Momentum (sp)	4.64%	4.55%	0.18	55.97	0.02

Table 10: **Descriptive statistics of Media factors and Benchmark factors (cont.)**

Panel B: Pairwise correlations between factors

		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Emotion (nb)	F1	1.00										
Emotion (sp)	F2	0.27	1.00									
Coverage (nb)	F3	0.01	0.06	1.00								
Coverage (sp)	F4	-0.05	0.00	0.36	1.00							
Basis (nb)	F5	-0.17	-0.12	0.02	-0.07	1.00						
Basis (h4 - sp)	F6	-0.08	-0.38	0.05	0.06	0.44	1.00					
Basis (l4 - sp)	F7	-0.01	-0.20	0.02	0.25	-0.32	-0.02	1.00				
Average (nearby)	F8	0.21	0.03	-0.04	0.09	-0.12	0.11	0.07	1.00			
Momentum (nb)	F9	0.18	0.13	-0.03	-0.02	-0.37	-0.27	0.12	0.05	1.00		
Basis-Momentum (nb)	F10	0.09	0.12	-0.16	-0.02	-0.24	-0.17	-0.06	0.19	0.24	1.00	
Basis-Momentum (sp)	F11	0.00	0.30	-0.10	-0.07	-0.14	-0.30	0.04	0.01	0.16	0.48	1.00

Table 11: **Cross-section tests for media emotion intensity and media coverage factors at portfolio level**

This table reports the cross-section tests for media factors and other benchmark factors for the sample from February 1998 to February 2020 at the portfolio level. Model (1) presents the test for the factors introduced by Szymanowska et al. (2014) and Boons and Prado (2019). Model (2) considers factors from Bakshi et al. (2017) and Boons and Prado (2019). We add media emotion intensity nearby factor to Model (1) to form Model (3) and continue to add media emotion intensity spreading factor to build Model (4). Similarly, Model (5) and Model (6) are Model (2) adding media emotion intensity nearby factor and then media emotion intensity spreading factor. We also add media coverage factors to Model (1) to form Models (7) and (8) and to Model (2) to form Models (9) and (10). To conduct the cross-section test at the portfolio level, we employ Fama Macbeth (1973) regression. We used 42 portfolios: including (1) 30 portfolios by sorting each factor of basis, momentum, basis-momentum, media emotion intensity and media coverage into three groups (High4, Mid, Low4) and separate to nearby and spreading strategies; (2) 12 portfolios by the two above strategies for commodity types (Energy, Grains, Livestocks, Metals, Softs and Food oils). We regress the cross-section regression of average returns of portfolios against their exposures to factors. We report the estimated risk price of each factor with the t-statistic from OLS estimation in the parentheses underneath and the t-statistic after correcting standard error using Shanken (1992)) method in the square brackets.

Portfolio-level Fama-Macbeth cross-section Test										
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Basis (nb)	-0.011	-0.007	-0.007	-0.007	-0.005	-0.005	-0.011	-0.012	-0.008	-0.009
	(-2.469)	(-1.721)	(-1.648)	(-1.692)	(-1.233)	(-1.238)	(-2.565)	(-2.635)	(-1.887)	(-2.207)
	[-1.757]	[-1.258]	[-1.146]	[-1.165]	[-0.877]	[-0.880]	[-1.808]	[-1.836]	[-1.354]	[-1.515]
Basis (h4 - sp)	0.000		-0.000	-0.000			0.000	0.000		
	(0.094)		(-0.193)	(-0.201)			(0.105)	(0.354)		
	[0.080]		[-0.165]	[-0.171]			[0.089]	[0.292]		
Basis (l4 - sp)	-0.004		-0.002	-0.002			-0.004	-0.004		
	(-2.455)		(-1.576)	(-1.463)			(-2.475)	(-2.427)		
	[-2.039]		[-1.316]	[-1.233]			[-2.050]	[-1.987]		
Average		0.003			0.003	0.003			0.003	0.003
		(1.204)			(1.080)	(1.098)			(1.218)	(1.198)
		[0.845]			[0.747]	[0.759]			[0.855]	[0.806]
Momentum (nb)		0.011			0.008	0.008			0.011	0.010
		(2.196)			(1.646)	(1.648)			(2.209)	(2.017)
		[1.557]			[1.138]	[1.139]			[1.565]	[1.362]
Basis-Momentum (nb)	0.012	0.01	0.011	0.011	0.010	0.010	0.012	0.012	0.01	0.009

	(2.390)	(2.404)	(2.220)	(2.281)	(2.422)	(2.422)	(2.426)	(2.649)	(2.400)	(2.332)
	[1.745]	[1.748]	[1.630]	[1.660]	[1.735]	[1.734]	[1.755]	[1.865]	[1.705]	[1.585]
Basis-Momentum (sp)	0.001	-0.000	0.004	0.004	0.003	0.003	0.001	0.001	-0.000	0.001
	(0.707)	(-0.080)	(2.381)	(2.217)	(2.815)	(2.680)	(0.699)	(0.580)	(-0.128)	(0.571)
	[0.595]	[-0.072]	[1.984]	[1.866]	[2.253]	[2.189]	[0.587]	[0.473]	[-0.114]	[0.472]
Emotion (nb)			0.013	0.013	0.012	0.012				
			(2.961)	(2.959)	(2.918)	(2.911)				
			[2.073]	[2.072]	[2.106]	[2.098]				
Emotion (sp)				0.002		0.001				
				(1.337)		(0.883)				
				[1.107]		[0.761]				
Coverage (nb)							-0.002	-0.002	-0.003	-0.001
							(-0.521)	(-0.455)	(-0.745)	(-0.396)
							[-0.379]	[-0.329]	[-0.573]	[-0.290]
Coverage (sp)								-0.002		-0.003
								(-0.986)		(-2.445)
								[-0.848]		[-2.088]
Constant	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001
	(-1.172)	(-5.239)	(-1.272)	(-1.147)	(-4.464)	(-4.989)	(-1.383)	(-1.416)	(-5.521)	(-4.588)
	[-1.051]	[-5.110]	[-1.147]	[-1.034]	[-4.241]	[-4.732]	[-1.243]	[-1.252]	[-5.382]	[-4.136]
R^2	0.430	0.517	0.527	0.601	0.613	0.669	0.483	0.552	0.572	0.602

Table 12: **Cross-section tests for media emotion intensity and media coverage factors at commodity level**

This table reports the cross-section tests for media factors and other benchmark factors for the sample from February 1998 to February 2020 at the commodity level. Model (1) presents the test for the factors introduced by Szymanowska et al. (2014) and Boons and Prado (2019). Model (2) considers factors from Bakshi et al. (2017) and Boons and Prado (2019). We add media emotion intensity nearby factor to Model (1) to form Model (3) and continue to add media emotion intensity spreading factor to build Model (4). Similarly, Model (5) and Model (6) are Model (2) adding media emotion intensity nearby factor and then media emotion intensity spreading factor. We also add media coverage factors to Model (1) to form Models (7) and (8) and to Model (2) to form Models (9) and (10). To conduct the cross-section test at the portfolio level, we employ Fama Macbeth (1973) regression. We used 42 portfolios: including (1) 30 portfolios by sorting each factor of basis, momentum, basis-momentum, media emotion intensity and media coverage into three groups (High4, Mid, Low4) and separate to nearby and spreading strategies; (2) 12 portfolios by the two above strategies for commodity types (Energy, Grains, Livestocks, Metals, Softs and Food oils). We regress the cross-section regression of average returns of portfolios against their exposures to factors. To run the cross-section tests at the commodity level, we employ Fama-Macbeth (1973) regression for the nearby and spreading returns of 26 commodity futures. At the first stage, we run the rolling time-series regression with one-year rolling of daily returns. In the second stage, we run the monthly cross-section regression of nearby and spreading returns of commodity futures on their time-varying exposures to factors. We report the estimated risk prices of factors with their t-statistic corrected using the Newey West method for one lag in the parenthesis and their t-statistic corrected using Shanken, 1992 method in the square brackets.

Commodity-level Fama-Macbeth cross-section Test										
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Basis (nb)	-0.006	-0.007	-0.007	-0.008	-0.005	-0.007	-0.006	-0.007	-0.010	-0.010
	(-0.928)	(-1.140)	(-1.148)	(-1.340)	(-0.821)	(-1.059)	(-0.968)	(-1.076)	(-1.441)	(-1.477)
	[-0.747]	[-0.976]	[-0.908]	[-1.018]	[-0.686]	[-0.887]	[-0.772]	[-0.813]	[-1.216]	[-1.170]
Basis (h4 - sp)	0.002		0.001	0.001			0.002	0.001		
	(1.406)		(0.863)	(0.954)			(1.227)	(1.075)		
	[1.236]		[0.767]	[0.799]			[1.060]	[0.866]		
Basis (l4 - sp)	0.001		0.002	0.002			0.001	0.002		
	(0.470)		(1.015)	(1.202)			(0.651)	(0.956)		
	[0.405]		[0.857]	[0.939]			[0.557]	[0.746]		
Average		0.004			0.003	0.003			0.003	0.003
		(1.419)			(1.367)	(1.371)			(1.324)	(1.244)
		[1.010]			[0.977]	[0.980]			[0.938]	[0.839]

Momentum (nb)		0.002			0.006	0.007			-0.000	0.000
		(0.353)			(0.921)	(1.124)			(-0.001)	(0.055)
		[0.273]			[0.722]	[0.878]			[-0.001]	[0.041]
Basis-Momentum (nb)	0.009	0.015	0.012	0.014	0.016	0.015	0.012	0.011	0.014	0.014
	(1.813)	(2.770)	(2.239)	(2.665)	(2.842)	(2.770)	(2.119)	(2.084)	(2.565)	(2.529)
	[1.359]	[2.179]	[1.676]	[1.884]	[2.280]	[2.221]	[1.623]	[1.521]	[1.998]	[1.871]
Basis-Momentum (sp)	0.001	0.003	0.001	-0.000	0.002	0.002	0.002	0.003	0.003	0.004
	(0.797)	(1.501)	(0.279)	(-0.195)	(1.483)	(1.237)	(1.086)	(1.566)	(1.599)	(2.082)
	[0.660]	[1.348]	[0.234]	[-0.153]	[1.286]	[1.090]	[0.920]	[1.240]	[1.427]	[1.713]
Emotion (nb)			0.014	0.013	0.011	0.011				
			(2.791)	(2.621)	(2.294)	(2.189)				
			[2.083]	[1.886]	[1.760]	[1.677]				
Emotion (sp)				0.002		0.001				
				(1.646)		(1.154)				
				[1.252]		[0.953]				
Coverage (nb)							-0.004	-0.004	-0.009	-0.01
							(-1.062)	(-1.302)	(-2.395)	(-2.874)
							[-0.757]	[-0.891]	[-1.797]	[-2.050]
Coverage (sp)								-0.003		-0.003
								(-1.936)		(-2.531)
								[-1.670]		[-2.146]
Constant	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(-1.096)	(-3.473)	(-1.212)	(-1.869)	(-3.757)	(-3.587)	(-1.366)	(-1.644)	(-3.344)	(-2.799)
	[-1.044]	[-3.386]	[-1.144]	[-1.653]	[-3.600]	[-3.439]	[-1.298]	[-1.452]	[-3.286]	[-2.532]
R^2	0.368	0.412	0.438	0.474	0.488	0.521	0.417	0.459	0.464	0.503

Table 13: **LASSO regression for selecting predictive factors**

This table summarizes the relative frequency of being selected for each factor in the first step of the two-pass regression with double-selection LASSO introduced by Feng et al. (2020). In this test, we employ LASSO cross-section regression to regress portfolios' average returns on the covariances between the portfolio returns with each factor. This technique will select the best predictors for the cross-section of portfolio returns. We employ three types of LASSO models: standard LASSO, adaptive LASSO introduced by Zou (2006) and plug-in LASSO from Belloni et al. (2012). For each type of LASSO model, we randomly generate 200 sub-samples from our main sample. Each sub-sample contains 75% observations of the whole sample. We run the LASSO model on the 200 samples and count the number of times each factor is selected among these 200 samples. The numbers reported in this table are the relative frequencies each factor appears in the set of selected factors.

	Standard LASSO	Adaptive LASSO	Plugin LASSO
Emotion (nb)	100.0%	99.5%	84.5%
Emotion (sp)	68.5%	28.0%	31.5%
Coverage (nb)	67.0%	46.0%	0.5%
Coverage (sp)	24.0%	17.0%	0.0%
Basis-Momentum (nb)	72.0%	55.0%	84.0%
Basis-Momentum (sp)	57.5%	46.0%	9.5%
Average (nb)	53.0%	27.5%	15.5%
Basis (nb)	53.5%	34.0%	36.5%
Basis (h4-sp)	25.0%	13.0%	4.5%
Basis (l4-sp)	23.0%	15.5%	0.0%
Momentum (nb)	52.5%	29.5%	50.0%

Internet appendix

A1 - Thomson Reuters MarketPsych Indices data and Lexical analysis

We use the Thomson Reuters MarketPsych Indices (TRMI) to calculate media emotion intensity. TRMI is based on a system of indicators that reflect a variety of psychological aspects extracted from media news and analyzed by the Thomson Reuters news reading system. The strength of TRMI is that this index system uses a unique Natural Language Processing algorithm based on a complicated system of Lexicon dedicated to business and investment as well as advanced grammatical analysis for news and social media textual analysis.

Traditionally, textual analysis often employs the "bag of word" approach. This method defines a dictionary of sentiment words and scores each word and phrase in a text using this predefined dictionary. There are some potential challenges with sentimental textual analysis. First, polarized measures such as sentiment with two aspects (positive and negative) might be confusing to a certain extent. Some information might be positive for some investors but negative for others based on the context. Second, the dictionary is required to be updated frequently to fit the current context. Third, the content should be carefully analyzed to focus only on the related information to a particular asset. Thomson Reuters overcomes these challenges by creating dynamic dictionaries for each aspect of information and regularly updating them to fit the financial and business environment. To associate the text's content with a particular asset, the system maintains a list of asset names and their aliases. This list is periodically updated and human-reviewed to maintain its accuracy. To exclude unrelated tokens generated by keyword scanning, the system uses a supervised machine learning algorithm to determine which tokens correspond with the asset names and which do not. For instance, when associating news material with gold, the machine can scan the token "gold medals" and relate the phrase to the Olympic Games rather than the gold commodity. As in the preceding example, the algorithm will employ anti-correlation filtering and case sensitivity to eliminate the perplexing tokens. Additionally, the system uses a correlation filter to determine whether the contents are associated with the asset. For instance, if the algorithm encounters the content "I love eating corn," this content will not be counted for "corn". The algorithm counts references only if they contain important identifiers such as "return" or "futures".

The Thomson Reuters news reading system collects words and phrases and categorizes them into two groups: psychological and factual. The psychological category encompasses words and phrases such as joy, hope, fear, stress, happy, hate and love that reflect psychological aspects. The factual category contains factual information about a subject, such as earnings, price, or return. The system then examines the attributes of each word. TRMI calls words or phrases with their attributes are "Variable". Accounting, earning, ambiguity, and anxiety are all examples of words. In these examples, accounting and earning might connect to factual information, while ambiguity and fear are emotional aspects. When scanning the text, the system will add the grammatical tense to the words and phrases and add the directional sentiment to them (if any) to form Variables. For instance, a sentence contains good information on accounting, and this information is from the past so that the system will save it as a Variable, namely "AccountingGood_p".

With regards to scoring a Variable, looking at an example of a sentence: "Experts expect gold will have much higher price next month". First, the system will scan the word "gold" to associate this information with the commodity "gold". Second, they identify the word "price" as a word in the dictionary. Third, the word "expect" relates the information to the future tense. Forth, identifying the word "higher" as an "up" word. Hence, the system will record a Variable "PriceUp_f" to refer to a piece of information on higher price in the future for the asset "gold". A standard score for a Variable is 1. However, because the word "higher" goes with the word "much", which intensifies the level of "higher", the score for the Variable "PriceUp_f" in this context is doubled to 2. If the system scans the text and sees some weaker modifiers, such as minimally, it will halve the score of the Variable to 0.5. Similarly, we can see an example of emotional content as the sentence "I love investing on gold futures". This sentence shows the view of the authors, which is not factual information of the asset “gold”. In this example, the word “love” refers to the psychological aspect of “love”. The word “love” also becomes a Variable “love” and got a score of 1. This Variable is recorded for the asset “gold” and the category “love”.

A2 - News sentiment measure

The term ”news sentiment” is frequently used to refer to two sentimental aspects of media news, positive and negative. García (2013), Tetlock (2007), and Tetlock et al. (2008) measure positive and negative news sentiment by the proportion of positive and negative words in the news articles based on predefined dictionaries suggested in the literature. We use the sentiment indicator in TRMI data to measure news sentiment for each commodity. TRMI uses the self-defined dictionary for positive and negative words to calculate news sentiment. This dictionary is designed specifically for investment and business contexts and is frequently updated and validated by humans. For a specific period, define the function $I_{sen}(v)$ as

$$I_{sen}(t,v) = \begin{cases} +1 \text{ if v conveys a positive view on the asset} \\ -1 \text{ if v conveys a negative view on the asset.} \end{cases} \tag{9}$$

The news sentiment of the asset a for the specific period is calculated as

$$News\ sentiment(a) = \frac{\sum\limits_{v \in V(a)} (I_{sen}(v) \times S(v))}{buzz(a)}. \tag{10}$$

This sentiment indicator is the net of positive to negative news sentiment and ranges from -1 to 1. A positive value reflects the positive average tone of media toward the asset. In comparison, a negative value indicates that the media, on average, views the asset negatively. Many previous studies, such as García (2013) and Smales (2014), measure news sentiment for each news item and then aggregate the news-level sentiment to calculate the daily news sentiment. In this study, we aggregate the monthly news sentiment from word-level sentiments. This measure is to consider the length of the news content in each news item. Further, this study features long-short strategies to evaluate the factors. Separating positive from negative sentiments might not accurately reflect the media's overall view of an asset. Therefore, we prefer to use the net news sentiment to rank the commodity futures and test the performance of the portfolio constructed based on this factor.

A3 - Conditional double-sorting on media emotion intensity and other factors

We further test conditional double-sorting on media emotion intensity and other factors. Table A1 presents the performance of the portfolios double-sorted by each factor first and then by media emotion intensity. Again, there is no significant difference in performance between high and low emotion intensity commodities within the high media coverage portfolio. The High-Low portfolio sorted by media emotion intensity for high media coverage commodities generates an average return of -2.66% (t-stat = -0.70). We only observed a significant difference within the group of the low media coverage commodities as the High-Low media emotion intensity portfolio generates a return of 12.19% (t-stat = 3.36). However, the portfolio's return mainly comes from the high media emotion intensity with low media coverage.

[Table A1 here]

Within each portfolio sorted by the other factors, the high media emotion intensity commodities bring higher average returns than low media emotion intensity. It should be noted that backwardation, high momentum, high basis-momentum, and high hedging pressure are the portfolios that generate better and more significant (higher t-stat) returns than the counterparts in each factor's single-sorting. Although the positive returns of High-Low portfolios sorted by media emotion intensity are only significant for high momentum, high basis-momentum, and high monthly hedging pressure, we observe the markedly higher and more significant returns of the High portfolios sorted by media emotion intensity within backwardation, high momentum, high basis-momentum, high hedging pressure commodities, which we call the “preferred” portfolio by each single-sorting. This indicates that emotion contributes to selecting commodities with superior performance (higher returns and higher portfolio t-stat) within each ”preferred” portfolio single-sorted by each other factor. This result also supports the long-only strategy in the commodity futures market, where investors should prioritize commodities with high media emotion intensity in “preferred” portfolios sorted by other factors.

When double-sorting on media emotion intensity and then each factor, we obtain a similar pattern. The performance of the portfolio resulting from this double-sorting is presented in Table A2. There is no significant difference between the average returns of commodities with high media coverage in the group with high media emotion intensity and those in the group with low media emotion intensity. Again, this implies that media emotion exerts little impact when media supply investors with abundant information. However, the low media coverage commodities create a much greater average return in the high media emotion intensity group than in the low media emotion intensity group. The difference in return between these two portfolios is approximately 14.45% (t-stat = 3.88). Additionally, the return difference between commodities with high and low media coverage is approximately double for the group with high media emotion intensity than low media emotion intensity (10.92% versus 5.62%).

[Table A2 here]

The Contango-Backwardation and High-Low portfolios sorted by other factors also generate significant returns for the high media emotion intensity commodities group. These portfolios get the majority of their profits from the “preferred” commodities in each single-sorting. Except for 12-month hedging pressure, the High-Low portfolios sorted by other factors provide higher returns for the group with high media emotion intensity than for the group with low media emotion intensity. Furthermore, portfolios that select “preferred” commodities by each factor among commodities with high media emotion intensity have the highest returns compared to the High-Low and “preferred” portfolios that are single-sorted by these factors. This again suggests that the media emotion intensity helps to filter the better commodity futures for one-month holding periods. The

High-Low portfolio sorted by 12-month HP generates a smaller but similar return in the high media emotion intensity group compared to the low media emotion intensity group, implying a potential link between media emotion intensity and the current trend in hedging pressure.

A4 - Discussion on whether media emotion intensity premium is subsumed by other risk sources

A4.1 - Hedging pressure

The role of hedging pressure stretches back to the view of Keynes (1930) on the theory of normal backwardation. According to this theory, hedgers participate in the futures market to hedge against spot price fluctuations. As hedgers take more short positions in the market, speculators share the risk with hedgers by setting the futures price lower than the expected spot price. In this way, the futures price is affected by hedging pressure. If the media emotion intensity implies hedging pressure, we can observe the media emotion intensity factor's performance being adjusted in response to hedging pressure.

To measure hedging pressure, we employ the data on trader positions from the Commodity Futures Trading Commission (CFTC). The CFTC publishes weekly reports, Commitment of Trader, that summarize the commercial and non-commercial traders' long and short positions in the commodity futures market. These reports are frequently provided on Friday of each week and contain data for the preceding Tuesday-to-Tuesday period. There is no perfect method for aggregating weekly data to monthly data. In this study, we assign a week (Tuesday to Tuesday) to a specific month if the last day of this week falls on that month. Kang et al. (2020) smooth the hedging pressure measure by taking the 52-week moving average of the weekly hedging pressure and show that this measure better captures the true hedging pressure in the commodity futures market. We use the same rolling period to calculate our hedging pressure variable (*12-month HP*) but modify it to suit the monthly frequency (12-month rolling). Furthermore, we also consider short-term HP (*monthly HP*).

First, we calculate the time series and cross-section correlations between hedging pressure and other factors to determine whether it interacts with our media emotion intensity factor and other benchmarks. We employ the Boons and Prado (2019) approach. To calculate time-series correlation, we first calculate the correlation between HP and the factor F for each commodity and then take the median of the correlations (across all commodities) as the time-series correlation between HP and the factor F. By comparison, the cross-section correlation between HP and the factor F is determined as the median of a series of monthly cross-section correlations between HP and F. These numbers are shown in Table A3 Panel A. Only momentum is found to be highly linked with hedging pressure over time. While it is clear that basis-momentum and basis have a minor correlation with hedging pressure, media emotion intensity is largely unrelated. Regarding cross-section associations, media emotion intensity has the strongest link with hedging pressure, at 0.37, followed by momentum (0.30). This could suggest that media emotion intensity might be partly related to hedging pressure in the commodity futures market.

[Table A3 here]

Further, we examine whether the long-short portfolio's abnormal return is primarily explained by hedging pressure. Panel B of Table A3 summarized the findings of the spanning test for the media emotion intensity long-short portfolio return when compared with hedging pressure, other benchmarks, and by all of them. We begin with sorting commodities into three portfolios (High4, Mid and Low4) based on hedging pressure for this spanning test. The nearby return of the High4-Low4 portfolio sorted by hedging pressure will be used to test the association of this factor to the media emotion intensity factor. The standard errors of all coefficients are adjusted using the Newey West method with one lag. Further, we compare media emotion intensity and basis-momentum performance using the same spanning test for the basis-momentum long-short portfolio. Model 1 shows that the return of hedging pressure portfolio is significantly correlated with the return of media emotion intensity portfolio; even after controlling for other factors, the correlation between hedging pressure and media emotion intensity remains highly significant. Model 2 indicates that media emotion intensity can generate a high abnormal return of about 12% annualized, even after controlling for the main benchmarks. When hedging pressure is considered, the alpha decreases to 7.74 %, and the corresponding significance decreases from t-stat = 2.56 to t-stat = 1.83. This result implies that media emotion intensity abnormal return is linked to hedging pressure. Media emotion intensity still generates a high abnormal return of nearly 8% after controlling for hedging pressure, implying that a substantial portion of the media emotion intensity return cannot be explained by hedging pressure or other benchmarks. When controlling for media emotion intensity and other benchmarks, basis-momentum can generate a 10% abnormal return. However, adding hedging pressure to the test reduces the alpha of basis-momentum to 7.85%, and the t-stat also decreases significantly from 2.25 to 1.80. Furthermore, we discover a substantial link between the long-short return of the hedging pressure portfolio and the long-short return of the basis-momentum portfolio. This result also suggests that the basis-momentum abnormal return is strongly related to hedging pressure. After accounting for hedging pressure, both media emotion intensity and basis-momentum continue to perform well and generate large abnormal returns at a similar power and significance level.

Although the spanning test indicates that a portion of the media emotion intensity's abnormal return might be attributed to hedging pressure, the cross-section analysis should be used to examine whether the media emotion intensity premium still significantly predicts the cross-section of commodity futures returns if hedging pressure is controlled. The result of this cross-section test is shown in Pane C of Table A3. We again assess the relevance of factors using Fama-Macbeth regression and Shanken (1992) correction. Model (1) includes the average factor representing market risk and hedging pressure. The result shows that the premium of hedging pressure is high and significant at 1.7% a month (nearly 20% annualized). When adding media emotion intensity to form Model (3), the risk price of hedging pressure falls to 1.2% a month (t-stat decreases from 2.58 to 1.77, suggesting that media emotion intensity does relate to hedging pressure. First, the positive and significant premium of hedging pressure is consistent with the theory of backwardation. Also, this estimation is consistent with Kang et al. (2020) that long-term risk is mainly driven by hedging pressure from commercial traders. Second, the decline in the hedging pressure premium while the premium of media emotion intensity remains unchanged indicates that media emotion intensity might partly link to hedging pressure but not be subsumed by this risk factor.

Interestingly, when hedging pressure and basis-momentum are combined in the same model, the risk price of basis-momentum falls to 0.8% per month and loses its significance. This implies that hedging pressure risk is also related and explains basis-momentum risk. When all factors are combined in the same test, we notice that the premium of hedging pressure decreases dramatically to 0.7% per month and becomes insignificant. We further observe that, in this model, the premium of basis-momentum changes a little, but the momentum factor substantially loses its significance and premium compared to Table 11. In summary, the cross-section test results indicate that the risk of hedging pressure might be largely linked to media emotion intensity and momentum. Unlike momentum, the estimated media emotion intensity premiums are consistent and high across all models. This conclusion is important because it indicates that a substantial portion of media emotion intensity premium cannot be explained by other factors.

A4.2 - Inventory

The evidence on the relation between media emotion intensity risk and hedging pressure suggests a link between media emotion intensity and the inventory state of commodities. G. B. Gorton et al. (2013) suggest that the physical inventory level is what drives the risk premiums of commodity futures. They note that basis and past futures return could capture the inventory level of commodities and explain the risk premiums. The origins of these interactions can be traced back to storage theory (Kaldor, 1939). At the high inventory level, the commodity supply is high and the futures price pattern is more likely to be contango. This happens because the futures price will increase to compensate for the storage cost. In contrast, when a commodity's inventory is low, the spot price will increase to reflect the scarcity of supplies. Therefore, the futures price pattern is more likely to be backwardation. According to G. B. Gorton et al. (2013), a low inventory level is frequently related to a negative basis (backwardation) and higher past futures returns. Thus, inventory risk can be represented by the basis and momentum of commodity futures returns.

The single sorts in Table 3 and Table 8 already demonstrate that backwardation and high momentum portfolios have a larger return than their counterparts. This result is consistent with Boons and Prado (2019) and G. B. Gorton et al. (2013), and supports the theory of storage. However, the spanning tests in Table 4 and Table A3 reveal that when controlling for basis, momentum, and hedging pressure, the alphas are still high above 7.6%, suggesting that these proxies for the inventory level cannot explain a large proportion of media emotion intensity abnormal returns. The cross-section test in Table 11 and Table A3 also confirm that the media emotion intensity premium is not subsumed by hedging pressure, basis, and momentum factors. Unfortunately, we do not have the direct measures for inventory and storage level as in Boons and Prado (2019). However, these results might suggest that a large part of media emotion intensity premium is not explained by the inventory level of commodities.

A4.3 - Market volatility and average volatility

According to Boons and Prado (2019), market volatility risk is priced in the cross-section of commodity futures returns. The concept of market liquidity stretches all the way back to Brunnermeier and Pedersen (2009). They contended that market liquidity is related to market volatility and that traders' ability to provide liquidity relies on their funding availability. Additionally, they demonstrated that speculators' funding availability might explain risk premiums. Boons and Prado (2019) argued that the negative risk premium associated with market volatility indicates that investors are willing to pay a premium to offset the increased volatility. In accordance with Boons and Prado (2019), we additionally examine whether media emotion intensity links to market volatility risk.

This study uses Boons and Prado (2019) methodology to calculate market and average volatility. We begin by building a market portfolio that is equally weighted in commodity futures to calculate market volatility. Then, the monthly market volatility is determined as the sum square of the market portfolio's daily returns. The average volatility is calculated as the mean of the monthly volatility of commodity futures' daily returns, where the monthly volatility of each commodity futures equals the sum square of its daily returns. First, we examine whether the market and average volatility can be used to predict the returns of portfolios sorted by media emotion intensity. The outcome is shown in Table A4. To conduct the test, we estimate WLS regression models of the nearby returns of the High4, Low4, Mid, and High4-Low4 portfolios at the month $t+1$ against the market volatility and the average volatility at the time t. These regressions are similar to those described in Boons and Prado (2019). In WLS regression, the weight is the inverted conditional volatility, measured by the standard deviation of the last 12 months' returns. The result indicates that both market volatility and average volatility have limited predictive potential, with $R2$ values of approximately 1% for the High4, Low4, and Mid portfolios and even as low as 0.01% for the High4-Low4 portfolio. Additionally, the t-stat for the market and average volatility coefficients in the High4-Low4 return model is quite minimal, at roughly 0.2. This indicates that market volatility or average volatility only explains a small portion of the following month performance of the long-short portfolio ranked by media emotion intensity.

[Table A4 here]

Furthermore, we investigate whether the media emotion intensity premium is subsumed by market volatility and average volatility. We use Fama-Macbeth regression with Shanken (1992) adjustment to do the cross-section test. The results of this test are summarized in Table A5. We begin with only the market factor and each volatility factor and then add the media emotion intensity factor into the model, followed by other benchmark factors. Model (1) indicates that the risk price of market volatility is -7.2% per month when only the average factor is used. This estimate is comparable to the -8.0% monthly recorded by Boons and Prado (2019). Additionally, our estimate of -2.8% in Model (2) is consistent with the 2019 paper's estimate of -2.4%. The only difference is that market volatility and average volatility are not significantly priced in the cross-section in our test. However, this may be because our sample size is smaller than that of Boons and Prado (2019). The Shanken t-statistic for market volatility is -1.42, which is close to being significant. The Shanken t-stat for average volatility, on the other hand, is -0.34, which is far from being significant. Interestingly, the media emotion intensity factor's premium remains highly constant and significant across models. The premium associated with media emotion intensity is approximately 1.2% a month (with t-stat ¿ 2), consistent with the results of all previous cross-section tests in this document. Therefore, we can conclude that the media emotion intensity premium is not subsumed by market volatility and average volatility risk.

[Table A5 here]

A4.4 - Market liquidity and funding liquidity

Brunnermeier and Pedersen (2009) argue that market liquidity and funding liquidity are connected to market volatility. They also point out that these two types of liquidity might link together. When the funding capital is tighter, investors appear to be less interested in trading. As a result, the market becomes increasingly illiquid, resulting in an increase in market volatility. The risk price of the market volatility measured for commodity futures markets might also suggest media emotion intensity might partly capture the market liquidity risk in the commodity futures market. We use the Ted spread as a proxy for funding illiquidity in the same way as Boons and Prado (2019). Previous studies often favor VXO over VIX owing to the rich historical data. Also, VIX and VXO are highly correlated. In our study, we employ VIX as a proxy for equity market volatility as our sample covers the period from 1998 when VIX was already available. As increased volatility may result from the increased market and funding illiquidity, VIX also reflects the risk associated with these two types of illiquidity.

We apply the two-stage Fama Macbeth regression with Shanken (1992) correction to evaluate these liquidity proxies. We use TED spread directly in the models as TED spread does not contain a time trend component and is stationary. However, because VIX is highly volatile and non-stationary, we conduct the test using the first-order difference of VIX. This transformation is required to assist the time series regression in the first stage. First, we run the cross-section test with only the average factor and each illiquidity proxy in the first two models. Models (3) and (4) are built by respectively supplementing Models (1) and (2) with media emotion intensity factors. Additionally, we incorporate basis-momentum, basis, and momentum factors to create Models (5) and (6). The results are reported in Table A6.

[Table A6 here]

Models (1) and (2) demonstrate that when only the average factor is controlled, both illiquidity proxies can explain the cross-section of portfolio returns. When media emotion intensity factors are included in Model (3), the estimated risk coefficient decreases from 0.227 to 0.181, and the t-stat drops to 1.365. This finding implies that media emotion intensity may partially represent the funding liquidity risk. When the basis-momentum, basis and momentum factors are added to Model (5), the coefficient of Ted spread continues to decrease and loses its significance. This change is consistent with Boons and Prado (2019) finding that basis-momentum also reflects funding liquidity risk. Similarly, the coefficient of the VIX factor is significant in Model (2) but less so in Model (4), highlighting that media emotion intensity might be related to funding and market liquidity. We calculate the correlation between the returns of the High4-Low4 portfolio ranked by media emotion intensity and the short-term reversal factor (Nagel, 2012) published on Kenneth French's website, following Boons and Prado (2019). The correlation of less than 0.1 indicates a weak relation between media emotion intensity premium in the commodity markets and liquidity provision in the equity market. The premium of media emotion intensity is relatively stable and consistent, at around significantly 1.2% a month. In other words, the premium of media emotion intensity is not subsumed by liquidity risks.

A4.5 - Equity market risk and Downside equity market risk

Boons and Prado (2019) suggests that both market risk and downside market risk of the equity market are priced in the cross-section of portfolio returns in the commodity futures markets. They argued that the bear equity market might discourage speculators from clearing the commodity futures market. Table A7 illustrates the cross-section test results when the equity market risk and equity downside market risk are considered. Similar to other cross-section tests, we apply the two-stage Fama-Macbeth regression with Shanken (1992) correction. We use the value-weighted CRSP portfolio's excess return to represent the equity market return. If the market return is less than the average minus one standard deviation for the entire sample, the market is deemed to be on the downside. This measure is suggested by Lettau et al. (2014).

[Table A7 here]

Model (1) shows that equity market risk accounts for 3.1% of the cross-section of portfolio returns in the commodity futures market each month. When downside market risk is added in Model (2), the equity market and downside equity market risk premiums are 3.7% (t-stat = 1.60) and 2.7% (t-stat = 1.62), respectively. After incorporating the media emotion intensity factor into the testing models, the equity market's risk premium is decreased to 2% and no longer significant (t-stat = 1.20). Similarly, the risk price of the downside equity market also drops to 1.5% (t-stat = 1.4). This finding implies that the media emotion intensity may relate to both the equity market risk and the equity market's downside risk. The premium associated with media emotion intensity is fairly stable at 1.2% per month and significant, as seen in the previous test. In our sample, the risk price of the downside market risk is smaller than in the sample of Boons and Prado (2019), suggesting that the commodity futures market is being less affected by the sentiment from the equity market. This test also indicates that the premium of media emotion intensity is not subsumed by equity market risk and downside equity market risk.

A4.6 - Macroeconomic factors

Given the fact that the situation of the economy has an impact on the supply and demand of commodities, the macroeconomic factor is inextricably linked to the commodity market. Due to the unique nature of different commodities, each commodity might respond differently to macroeconomic shocks. Numerous studies have established the macroeconomic influences on the returns on commodity prices (see Le Pen and Sévi, 2018). This section tests the predictive role of macroeconomic risks in the cross-section of commodity futures returns and whether the macroeconomic risks subsume the premium of media emotion intensity.

We collect ten macroeconomic variables: CPI, CPI for commodity, real M1 supply, real M2 supply, PPI for final demand of commodities, PPI for commodities in the US, PPI for manufacturing, Unemployment, US Dollar index, and Michigan consumer sentiment index. These are the key macroeconomic variables that reflect various facets of the US economy. We are aware that while several of these variables are highly correlated, they are not identical. As a result, we begin by doing a principal component analysis of these macro variables to identify the key factors containing the primary information. Because none of these variables is stationary, they are transformed to the first-order difference form to facilitate the time-series regression in the first stage of the cross-section test. The results of the principal component analysis are summarized in Table A8. Panel A demonstrates that the third factor's eigenvalue is 1.09, whereas the fourth factor's eigenvalue is 0.84. As a result, following PCA analysis, we retain only three factors. Panel B reports the loading of each factor on the macroeconomic variables. Factor 1 is mainly associated with CPI, PPI and M2 supply. These factors reflect the price level of the economy. Factor 2 mostly captures the change in consumer sentiment and unemployment. Both of these variables are related to the labor market and consumer confidence (see Ludvigson, 2004). Factor 3 is tied to the M1 supply and the US dollar index, which measures the US currency's strength.

[Table A8 here]

The cross-section test for macroeconomic factors using two-stage Fama-Macbeth regression with Shanken, 1992 correction is summarized in Table A9. Only the factor related to consumer confidence and the labor market can significantly explain the cross-section of portfolio returns in our sample. Factor 1 risk is not significantly priced (t-stat = -1.11), showing that the inflation risk contributes only a negligible amount to the cross-section explanation. This finding contrasts with that of Szymanowska et al. (2014), who discovered that inflation risk has a larger predictive power. It should be emphasized, however, that the Shanken correction significantly reduces the t-statistic compared to the OLS estimate or Newey West correction. Furthermore, given our short sample period, it is more difficult for factors to survive the significant thresholds. However, we can observe the relation between media emotion intensity and macroeconomic risks through the change in coefficients and their corresponding t-statistic. Feng et al. (2020) argued that we should still bring back the variables that correlate with the main factors to reduce the estimation bias for premiums.

[Table A9 here]

Two major findings are revealed in Table A9. First, when adding media emotion intensity to the model of macroeconomic factors with the average factor, the estimated coefficients of the macroeconomic factors change greatly. Also, their t-stat decrease considerably. Particularly for Factor 2, the estimated coefficient loses its significance markedly from t-stat = -1.80 to t-stat = -0.196. This means that the premium associated with media emotion intensity is related to the risk associated with the labor market and consumer spending. Even this relationship is obscure in the previous literature. Second, after controlling for all the macroeconomic factors and other benchmarks, the estimated premium of the media emotion intensity factor remains significant and rather steady, staying at around 1.2% monthly. This result confirms that the media emotion intensity premium is not subsumed by macroeconomic risks.

Appendix tables

Table A1: Double-sorting on factor first and emotion second

This table reports the nearby returns of portfolios conditionally double-sorted by media emotion intensity and other variables. More specifically, the commodities are sorted into two levels by each factor first and then, within each level of each factor, we continue to sort the commodities by media emotion intensity. We first report the return with t-statistics of the portfolios single-sorted by each factor in the third and four columns. For media emotion intensity, media coverage, momentum, basis-momentum, and hedging pressure, we sort commodities into two levels, High and Low, based on the factor's median. For basis, we sort commodities into two portfolios: contango and backwardation. All the portfolios are created by sorting commodities using factor value at the end of each month and then held for one month. The last six columns report the returns of the portfolios sorted by media emotion intensity within each level single-sorted by other factors. Panel A presents the result of single-sorting commodities into two portfolios based on the median value of media emotion intensity and the result of double-sorting commodities using media emotion intensity and media coverage. Panel B reports the results of double-sorting commodities using media emotion intensity and other factors.

Panel A: Double-sorting on media emotion intensity and media coverage

		Single Sort on Row Variable		Sorting on factor first, and then emotion					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Emotion	High	4.48	1.31						
(t)	Low	-1.56	-0.51						
	High - Low	6.04	2.10						
Coverage	High	0.73	0.23	-0.57	-0.16	2.08	0.56	-2.66	-0.70
(t)	Low	2.42	0.80	8.20	2.16	-3.99	-1.25	12.19	3.36
	High - Low	-1.69	-0.89	-8.78	-2.98	6.07	1.87		

Panel B: Double-sorting on media emotion intensity and other factors

		Single Sort on Row Variable		Conditional sorting on emotion and other factors					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Basis	Contango	-0.24	-0.08	2.13	0.60	-3.63	-1.12	5.76	1.83
(t)	Backwardation	7.87	1.86	10.72	2.27	3.62	0.67	7.10	1.29
	Conta. - Back.	-8.11	-2.26	-8.59	-1.99	-7.25	-1.42		
Momentum	High	4.94	1.48	8.48	2.15	1.28	0.35	7.20	2.05
(t)	Low	-1.94	-0.61	0.28	0.07	-4.48	-1.28	4.76	1.21
	High - Low	6.88	2.30	8.20	2.05	5.76	1.54		
Basis-Momentum	High	6.00	1.77	8.81	2.34	2.85	0.74	5.96	1.71
(t)	Low	-3.27	-1.06	-2.46	-0.67	-4.42	-1.31	1.96	0.55
	High - Low	9.27	3.34	11.27	3.09	7.27	2.18		
Monthly HP	High	2.37	0.71	6.93	1.85	-2.20	-0.58	9.13	2.54
(t)	Low	-1.34	-0.46	-0.06	-0.02	-2.66	-0.76	2.61	0.68
	High - Low	3.71	1.44	6.98	1.94	0.46	0.13		
12-month HP	High	5.83	1.71	8.90	2.35	2.80	0.70	6.09	1.60
(t)	Low	-2.46	-0.78	-0.10	-0.03	-4.87	-1.35	4.76	1.19
	High - Low	8.29	2.83	9.00	2.25	7.67	1.90		

Table A2: Double-sorting on emotion first and factor second

This table reports the nearby returns of portfolios conditionally double-sorted by media emotion intensity and other variables. More specifically, the commodities are sorted into high and low media emotion intensity groups first and then, within each level of media emotion intensity, we continue to sort the commodities by other factors' value. We first report the return with t-statistics of the portfolios single-sorted by each factor in the third and four columns. For media emotion intensity, media coverage, momentum, basis-momentum, and hedging pressure, we sort commodities into two levels, High and Low, based on the factor's median. For basis, we sort commodities into two portfolios: contango and backwardation. All the portfolios are created by sorting commodities using factor value at the end of each month and then held for one month. The last six columns report the returns of the portfolios sorted by each factor within each level single-sorted by media emotion intensity. Panel A presents the result of single-sorting commodities into two portfolios based on the median value of media emotion intensity and the result of double-sorting commodities using media emotion intensity and media coverage. Panel B reports the results of double-sorting commodities using media emotion intensity and other factors.

Panel A: Double-sorting on media emotion intensity and media coverage

		Single Sort on Row Variable		Sorting on emotion first, and then factor					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Emotion	High	4.48	1.31						
(t)	Low	-1.56	-0.51						
	High - Low	6.04	2.10						
Coverage	High	0.73	0.23	-0.93	-0.26	1.16	0.32	-2.09	-0.57
(t)	Low	2.42	0.80	9.99	2.57	-4.46	-1.38	14.45	3.88
	High - Low	-1.69	-0.89	-10.92	-3.86	5.62	1.77		

Panel B: Double-sorting on media emotion intensity and other factors

		Single Sort on Row Variable		Independent sorting on emotion and other factors					
				High (Emotion)		Low (Emotion)		High - Low (Emotion)	
		Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat	Ave. Ret	t-stat
Basis	Contango	-0.24	-0.08	0.22	0.06	-3.35	-1.01	3.57	1.04
(t)	Backwardation	7.87	1.86	11.53	2.34	-2.36	-0.48	13.90	2.54
	Conta. - Back.	-8.11	-2.26	-11.31	-2.65	-0.99	-0.21		
Momentum	High	4.94	1.48	8.05	2.02	0.15	0.04	7.89	2.07
(t)	Low	-1.94	-0.61	1.07	0.27	-3.58	-1.02	4.65	1.19
	High - Low	6.88	2.30	6.98	1.76	3.73	1.05		
Basis-Momentum	High	6.00	1.77	9.01	2.34	1.24	0.35	7.77	2.11
(t)	Low	-3.27	-1.06	-1.05	-0.26	-4.77	-1.43	3.72	0.95
	High - Low	9.27	3.34	10.06	2.57	6.01	1.85		
Monthly HP	High	2.37	0.71	8.23	2.15	-0.30	-0.09	8.53	2.42
(t)	Low	-1.34	-0.46	-0.41	-0.11	-3.44	-1.00	3.03	0.79
	High - Low	3.71	1.44	8.64	2.80	3.14	0.91		
12-month HP	High	5.83	1.71	9.29	2.40	3.49	0.96	5.80	1.65
(t)	Low	-2.46	-0.78	1.27	0.33	-6.01	-1.72	7.28	1.77
	High - Low	8.29	2.83	8.02	2.30	9.50	2.71		

Table A3: Media emotion intensity versus hedging pressure

This table shows how hedging pressure interacts with media emotion intensity in explaining returns in the commodity futures market. We measure hedging pressure as the number of short positions minus the number of long positions of commercial traders scaled by total interest. Following Kang et al. (2020), we smooth hedging pressure using a one-year moving average of monthly hedging pressure. Panel A presents the correlation of hedging pressure with media emotion intensity and other benchmarks. The correlation is calculated for two settings: time-series and cross-section. We follow Boons and Prado (2019) to calculate time-series correlation using the median of time-series correlations of individual commodity hedging pressure with the other factor. The cross-section correlation is the median of the monthly cross-section correlations between hedging pressure and media emotion intensity of commodities. Panel B reports the result of the spanning test for media emotion intensity and the basis-momentum factor adding hedging pressure to the interacting factor list. We first test for only hedging pressure in the first model, only other factors in the second model and combine hedging pressure and other factors in the third model. The standard errors of estimated coefficients are all corrected using the Newey West method with one lag when doing the spanning test. The t-statistics of alpha is reported in the parenthesis. Panel C reports the cross-section test adding hedging pressure to the factor list. We employ two-stage Fama-Macbeth (1973) regression for 42 factor-sorted and commodity-type portfolio returns. We correct the standard errors using Shanken (1992) and report the corresponding t-statistic in the parenthesis. Coefficients with "***", "**", and "*" are statistically significant at 1%, 5% and 10% levels, respectively. The results are estimated for the sample from February 1998 to February 2020.

Panel A: Correlation between hedging pressure and other factors

	Correlation	Basis	Momentum	Basis-Momentum	Emotion	Buzz
Hedging pressure	Time-series	-0.17	0.47	-0.13	0.03	0.05
	Cross-section	-0.20	0.30	0.15	0.37	-0.11

Panel B: Spanning test for media emotion intensity and basis-momentum using hedging pressure

	Media emotion			Basis-Momentum		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	0.0763*	0.1178**	0.0774*	0.0798*	0.1004**	0.0785*
t-stat (Alpha)	(1.7863)	(2.5620)	(1.8254)	(1.7719)	(2.2523)	(1.8018)
Hedging pressure (nb)	0.5348***		0.5104***	0.3251***		0.2719***
Emotion (nb)					0.0080	-0.0658
Basis-Momentum (sp)		0.0078	-0.0565			
Basis (nb)		-0.1234*	-0.0573		-0.1747**	-0.1467*
Momentum (nb)		-0.0973	-0.0008		-0.1962	-0.0411
Average (nb)		0.3978***	0.2013		0.4667***	0.3178**

R-squared	0.2351	0.0692	0.2548	0.0858	0.0990	0.1431
GRS-F	3.3679	6.8975	3.5016	3.0404	4.8760	3.0991
p-value	0.0665	0.0086	0.0613	0.0812	0.0272	0.0783

Panel C: Cross-section test for hedging pressure risk

Model	(1)	(2)	(3)	(4)	(5)	(6)
Average (nb)	0.004	0.003	0.004	0.003	0.004	0.004
	(1.062)	(0.865)	(1.083)	(0.853)	(1.051)	(1.052)
Hedging pressure (nb)	0.017**		0.012*		0.016**	0.007
	(2.580)		(1.769)		(2.381)	(0.938)
Emotion (nb)		0.013**	0.013**			0.012**
		(2.232)	(2.109)			(2.079)
Basis-momentum (nb)				0.013**	0.008	0.010*
				(2.132)	(1.394)	(1.710)
Basis (nb)						-0.005
						(-0.852)
Momentum (nb)						0.007
						(1.031)
cons	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(-4.756)	(-5.170)	(-4.761)	(-5.401)	(-4.686)	(-4.589)
R-squared	0.393	0.413	0.492	0.347	0.446	0.659

Table A4: Market volatility and average volatility in predicting media emotion intensity portfolio returns

This table reports the results of the time-series regression for the predictive role of market volatility and average volatility on returns of portfolios sorted by media emotion intensity. We follow Boons and Prado (2019) to measure market volatility as the sum of squared daily returns of market portfolio (equal-weighted portfolio of commodity futures) and average volatility as the mean of monthly volatility of daily returns of individual commodity futures. Panel A shows the predictive power of market volatility for four portfolios constructed based on media emotion intensity: High4, Low4, Mid and High4-Low4. Panel B reports the results of similar tests for average volatility. For each test, we employ the method seen in Boons and Prado (2019) that regresses the portfolio returns on a one-month lag of the volatility using WLS regression weighted on the inverted standard deviation of monthly returns calculated for the recent 12 months. The results are reported for the sample from February 1998 to February 2020.

	Nearby return (t + 1)			
VARIABLES	Emotion (High4)	Emotion (Mid)	Emotion (Low4)	Emotion (nb)
Market volatility (t)	-2.8236	-3.4915	-2.5856	-0.2380
	(-0.8857)	(-1.1022)	(-1.1995)	(-0.1130)
R^2	0.0064	0.0191	0.0074	0.0000
Average volatility (t)	-0.9480	-1.5765	-1.1989	0.2510
	(-0.6040)	(-1.0590)	(-1.1072)	(0.2186)
R^2	0.0024	0.0127	0.0053	0.0001

Table A5: Cross-section tests for market volatility and average volatility.

This table reports the results of the cross-section tests for market volatility and average volatility. We employ the two-stage Fama-Macbeth (1973) regression to estimate the premium of each factor. In the first stage, we run the time series regression for the whole sample to estimate the time-series exposures of each portfolio to the factors. In the second stage, we run the cross-section regression of portfolio returns on these exposures. The standard errors of all estimated coefficients are adjusted following Shanken (1992) correction. We report the estimated premiums and their Shanken t-statistic in the parenthesis underneath each estimated coefficient for each model. For each selected factor, we run (1) the test for average (market) factor and the factor, (2) model adding media emotion intensity factor to the first model, and (3) model adding basis-momentum, basis and momentum factor to the second model. We follow Boons and Prado (2019) to measure market volatility as the sum of squared daily returns of the market portfolio (equal-weighted on commodity futures) and average volatility as the average monthly volatility of daily returns of commodity futures. Coefficients with "***", "**", and "*" are statistically significant at 1%, 5% and 10% levels, respectively. The results are estimated for the sample from February 1998 to February 2020.

Model	(1)	(2)	(3)	(4)	(5)	(6)
Average (nb)	0.003	0.003	0.003	0.003	0.003	0.003
	(0.959)	(0.956)	(0.875)	(0.866)	(0.792)	(0.794)
Market volatility	-0.072		-0.019		-0.004	
	(-1.418)		(-0.451)		(-0.094)	
Average volatility		-0.028		-0.012		-0.011
		(-0.342)		(-0.142)		(-0.124)
Emotion (nb)			0.013**	0.013**	0.012**	0.012**
			(2.224)	(2.240)	(2.090)	(2.082)
Basis-momentum (nb)					0.010*	0.011*
					(1.823)	(1.830)
Basis (nb)					-0.005	-0.005
					(-0.923)	(-0.910)
Momentum (nb)					0.008	0.008
					(1.171)	(1.176)
cons	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
	(-5.662)	(-6.034)	(-5.329)	(-5.189)	(-4.737)	(-4.702)
R-squared	0.332	0.324	0.458	0.461	0.653	0.652

Table A6: Cross-section tests for funding liquidity and market liquidity risk.

This table reports the results of the cross-section tests for market liquidity and funding liquidity risk. We employ the two-stage Fama Macbeth regression to estimate the risk price of each factor. In the first stage, we run the time series regression for the whole sample to estimate the time-series exposures of each portfolio to the factors. In the second stage, we run the cross-section regression of portfolio returns on these exposures. The standard errors of all estimated coefficients are adjusted following Shanken (1992) correction. We report the estimated risk prices and their Shanken t-statistic in the parenthesis underneath each estimated coefficient for each model. For each selected factor, we run (1) the test for average (market) factor and the factor, (2) model adding media emotion intensity factor to the first model, and (3) model adding basis-momentum, basis and momentum factor to the second model. The table shows the result of testing funding iliquidity and markert iliquidity. We employ two proxies for funding liquidity: Ted spread (the gap between 3-month LIBOR rate and 3-month Treasury rate) and VIX (CBOE volatility index). Coefficients with "***", "**", and "*" are statistically significant at 1%, 5% and 10% levels, respectively. The results are estimated for the sample from February 1998 to February 2020.

Model	(1)	(2)	(3)	(4)	(5)	(6)
Average (nb)	0.003	0.003	0.003	0.003	0.003	0.003
	(0.738)	(0.897)	(0.711)	(0.897)	(0.716)	(0.698)
Ted spread	0.227*		0.181		0.119	
	(1.656)		(1.365)		(0.845)	
ΔVIX		-2.736*		-0.927		1.770
		(-1.834)		(-0.772)		(1.228)
Emotion (nb)			0.013**	0.012**	0.012**	0.013**
			(2.048)	(2.096)	(2.065)	(2.088)
Basis-momentum (nb)					0.010*	0.011*
					(1.670)	(1.776)
Basis (nb)					-0.006	-0.004
					(-0.941)	(-0.702)
Momentum (nb)					0.008	0.008
					(1.111)	(1.057)
cons	-0.001***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
	(-4.789)	(-4.613)	(-4.353)	(-4.894)	(-4.539)	(-4.100)
R^2	0.344	0.360	0.480	0.465	0.659	0.640

Table A7: Cross-section tests for equity market risk.

This table reports the results of the cross-section tests for equity market risk and downside equity market risk. We employ the two-stage Fama-Macbeth (1973) regression to estimate the risk price of each factor. In the first stage, we run the time series regression for the whole sample to estimate the time-series exposures of each portfolio to the factors. In the second stage, we run the cross-section regression of portfolio returns on these exposures. The standard errors of all estimated coefficients are adjusted following Shanken (1992) correction. We report the estimated risk prices and their Shanken t-statistic in the parenthesis underneath each estimated coefficient for each model. For each selected factor, we run (1) the test for average (market) factor and the factor, (2) model adding media emotion intensity factor to the first model, and (3) model adding basis-momentum, basis and momentum factor to the second model. We examine the equity market risk using value-weighted CRSP excess return and also consider the downside equity market risk as suggested by Boons and Prado (2019). Coefficients with "***", "**", and "*" are statistically significant at 1%, 5% and 10% levels, respectively. The results are estimated for the sample from February 1998 to February 2020.

Model	(1)	(2)	(3)	(4)	(5)	(6)
Average (nb)	0.003	0.003	0.003	0.003	0.003	0.003
	(0.891)	(0.791)	(0.855)	(0.824)	(0.777)	(0.734)
CRSP excess return	0.031	0.037	0.020	0.023	-0.005	-0.012
	(1.519)	(1.604)	(1.065)	(1.197)	(-0.285)	(-0.757)
CRSP excess return (downside)		0.027		0.015		-0.013
		(1.620)		(1.142)		(-1.286)
Emotion (nb)			0.012*	0.012*	0.012**	0.012**
			(1.926)	(1.859)	(2.084)	(1.971)
Basis-momentum (nb)					0.011*	0.010*
					(1.846)	(1.695)
Basis (nb)					-0.005	-0.006
					(-0.876)	(-0.951)
Momentum (nb)					0.008	0.008
					(1.172)	(1.096)
cons	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
	(-4.841)	(-4.041)	(-4.712)	(-4.334)	(-4.676)	(-4.148)
R^2	0.331	0.395	0.468	0.520	0.630	0.666

Table A8: Principal component analysis of macroeconomic variables

This table summarizes the results of the Principle component analysis of macroeconomic variables. We employ 11 macroeconomic variables, including CPI, CPI for commodities, 3-month LIBOR, real M1, real M2, PPI for final demand of commodities, PPI for the US, PPI for total manufacturing, Unemployment, US Dollar index and Michigan consumer sentiment. All macroeconomic variables are used in difference to ensure their stationary. We conduct the principal component analysis (CPA) to extract orthogonal factors from the set of variables with varimax rotation. Panel A reports the result of PCA results with eigenvalues of adding factors. Panel B shows the loadings of factors on variables.

Panel A: Factor analysis and Eigenvalues

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	5.1804	3.9672	0.5180	0.5180
Factor2	1.2132	0.1263	0.1213	0.6394
Factor3	1.0869	0.2425	0.1087	0.7480
Factor4	0.8444	0.0266	0.0844	0.8325
Factor5	0.8178	0.4094	0.0818	0.9143
Factor6	0.4084	0.1809	0.0408	0.9551
Factor7	0.2276	0.1077	0.0228	0.9779
Factor8	0.1198	0.0418	0.0120	0.9898

Panel B: Factor loadings and Uniqueness

Variable	Factor1	Factor2	Factor3	Uniqueness
ΔConsumer sentiment		0.7059		0.2015
ΔCPI (commodity)	0.9269			0.1360
ΔCPI	0.9323			0.1266
ΔM1 (real)			0.6593	0.2817
ΔM2 (real)	-0.7818			0.2453
ΔPPI (commodity - final demand)	0.9278			0.1175
ΔPPI (commodity - US)	0.9115			0.1452
ΔPPI (total manufacturing)	0.9294			0.1234
ΔUnemployment		-0.7273		0.3183
ΔUS Dollar index			-0.6823	0.4575

Table A9: Cross-section tests for macroeconomic factors.

This table presents the cross-section tests for macroeconomic factors in the model with other benchmark factors. We employ 11 macroeconomic variables, including CPI, CPI for commodities, 3-month LIBOR, real M1, real M2, PPI for final demand of commodities, PPI for the US, PPI for total manufacturing, unemployment, US Dollar index and Michigan consumer sentiment. All macroeconomic variables are used in difference to ensure their stationary. First, we conduct Principal Component Analysis (PCA) to extract the main orthogonal factors from the macroeconomic variables. After that, we use the factors predicted from PCA analysis to test their risks to the cross-section of commodity futures returns. In the first four models, we include only the average (market) factor and each macro factor. In the following four models, from Model (5) to Model (8), we add media emotion intensity factors to the set of factors employed in Model (1) to Model (4), respectively. In the last four models, Model (9) to Model (12), we further add other benchmark factors to the models of the macro factor and media factors models. We use the two-stage Fama Macbeth regression with Shanken (1992) correction to run the cross-section test. The t-statistic of each estimated risk price is calculated based on the Shanken (1992) standard error and reported in the parenthesis.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average (nb)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	(0.907)	(0.763)	(0.924)	(0.865)	(0.865)	(0.870)	(0.797)	(0.769)	(0.791)
Macro factor 1	0.253			-0.199			-0.041		
	(1.113)			(-0.786)			(-0.159)		
Macro factor 2		-0.902*			0.083			-0.295	
		(-1.797)			(0.196)			(-0.634)	
Macro factor 3			-0.159			0.082			-0.018
			(-0.569)			(0.309)			(-0.068)
Emotion (nb)				0.014**	0.013**	0.013**	0.012**	0.012**	0.012**
				(2.364)	(2.223)	(2.294)	(2.151)	(2.047)	(2.085)
Basis-Momentum (nb)							0.010*	0.010*	0.011*
							(1.773)	(1.718)	(1.813)
Basis (nb)							-0.006	-0.005	-0.005
							(-0.979)	(-0.853)	(-0.892)
Momentum (nb)							0.008	0.008	0.008
							(1.150)	(1.139)	(1.185)
cons	-0.002***	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(-5.532)	(-4.529)	(-6.002)	(-4.962)	(-4.991)	(-5.327)	(-4.510)	(-4.434)	(-4.925)
R2	0.398	0.349	0.351	0.499	0.463	0.478	0.661	0.647	0.647