Financialization and Commodity Market Serial Dependence*

Zhi Da[†], Ke Tang[‡], and Yubo Tao[§]

First Draft: November, 2018
This Draft: October, 2019

Abstract

Recent financialization in commodity markets makes it easier for institutional investors to trade a portfolio of commodities via various commodity index products. Using news-based sentiment measures, we find that such trading can propagate non-fundamental shocks from some commodities to others in the same index, giving rise to price overshoots and subsequent reversals at a daily frequency. Price overshooting results in negative return autocorrelations for indexed commodities (and commodity indices) but not for non-indexed commodities, and such autocorrelations are closely linked to index exposure measures. Taking advantage of the fact that index weights of the same commodity can vary across different indices in a relatively ad-hoc and pre-determined fashion, we provide causal evidence that index trading drives return autocorrelation. Our results speak to efficient price discovery in financialized commodity futures markets.

JEL Classification: G12, G40, Q02.
Keywords: Financialization; Return autocorrelation; Index trading; News sentiment; Price discovery.

*We have benefited from comments and suggestions from Shaun Davies (discussant), Prachi Deuskar (discussant), Li Guo, Jun Li (discussant), Christina Nikitopoulos (discussant), Neil Pearson, Marcel Prokopczuk, Jing Wu, Liyan Yang (discussant), Qifei Zhu (discussant), and seminar participants at University of Notre Dame, the 3rd Australasian Commodity Markets Conference, the 2019 Commodity and Energy Markets Association Annual Conference, the 7th ABFER Annual Conference, the 2019 China International Conference in Finance, the 2019 Asian Meeting of Econometric Society, the 3rd JPMCC International Symposium, the 2019 Summer Institute of Finance, and the 2019 University of Oklahoma Energy and Commodities Finance Research Conference.

[†]University of Notre Dame, Mendoza College of Business, Notre Dame, IN, 46556, USA. Email: zda@nd.edu.
[‡]Tsinghua University, Institute of Economics, Beijing, 100084, China. Email: ketang@tsinghua.edu.cn.
[§]Singapore Management University, School of Economics, 90 Stamford Road, 178903, Singapore. Email: ybtao@smu.edu.sg

1 Introduction

The last two decades witnessed the financialization of the commodity markets. According to the staff report from the Commodity Futures Trading Commission (CFTC), investment flows to various commodity indices increased from $15 to $200 billion during the period from 2003 to 2008. Barclays estimated the commodity index investment rose to $319 billion in the first quarter of 2019.[1] The rapid money inflow in commodity markets especially in the years 2007 and 2008 leads to hot debate among researchers and policymakers about the influence of financialization on commodity price discovery and return dynamics.

Coinciding with the large investment inflow to commodity indices, different commodities started to display synchronized boom and bust cycles. In addition, Tang and Xiong (2012) find such comovement to be more severe for commodities in popular indices (indexed commodities) than for those excluded from indices (non-indexed commodities), as shown in Figure 1.[2]

[Figure 1 is about here.]

As mentioned in Henderson, Pearson and Wang (2014), "The hypothesis that the flows of financial commodity investors did impact the commodity markets is referred to as the 'financialization of commodity markets.'" Comovement among indexed commodities in itself, however, does not necessarily imply that financialization is the cause, since indexed commodities could have been endogenously selected into an index, precisely because they are exposed to the same fundamental shocks. Therefore, direct tests along the lines of price discovery of financialization require clean identification strategies.

Our main variable of interest is the daily return autocorrelation instead of return correlation

[1] https://www.kitco.com/news/2019-04-08/Barclays-Commodity-Assets-Under-Management-Rise-In-1Q.html.

[2] We first calculate an equal-weighted index for each sector of indexed and non-indexed commodities, then calculated the average correlation among five sector indices for an annual rolling window. Since there are no non-indexed commodities in energy and live cattle sectors, we take heating oil and RBOB and lean hogs as non-indexed commodities due to their small weights in the index.

of different commodities. When we do that, we observe a clear divergence between the indexed commodity portfolio and the non-indexed commodity portfolio, as evident in Figure 2. Similar to Baltussen, van Bekkum and Da (2019), we draw the first-order autocorrelation coefficient of returns on commodity indices with a backward 10-year rolling window. We observe a slight increasing trend in the past 38 years in the daily autocorrelation in returns of the non-indexed commodity portfolio (NIDX).[3] In sharp contrast, the daily autocorrelations in popular commodity indices (S&P Goldman Sachs Commodity Index (GSCI) and the Bloomberg Commodity Index (BCOM)) have steadily declined since 2004 when financialization began.[4] They entered the negative territory around 2005 and became significantly negative since 2006. While the early-stage declining index return autocorrelation before 2004 can be consistent with improved information efficiency when common fundamental shocks are simultaneously and efficiently incorporated into the prices of multiple indexed commodities, *a negative return autocorrelation unambiguously signals inefficiency in price discovery.* Figure 2 thus corroborate the theoretical hypothesis by Goldstein and Yang (2017), who write "a process of increased financialization first increases and then decreases price informativeness." It suggests that prices across multiple indexed commodities can overshoot and subsequently revert at the same time, resulting in negative autocorrelation, even at the index level.

[Figure 2 is about here.]

Negative return autocorrelation at daily frequency is hard to explain using fundamental factors. For example, common discount rate or risk premium variations which can also cause negative return autocorrelations tend to operate at business cycle frequency. Instead, we attribute it to financialization and the resulting commodity index trading that propagate "non-fundamental shocks" from

[3] The return of the equal-weighted non-indexed commodity portfolio is the simple average of non-indexed commodities' returns. To mitigate the effect of outliers, we winsorize each nonindexed commodity's return at the top and bottom 5 percentile before taking the average.

[4] GSCI was originally developed in 1991, by Goldman Sachs. In 2007, ownership was transferred to Standard & Poor's. BCOM was originally launched in 1998 as the Dow Jones-AIG Commodity Index (DJ-AIGCI) and renamed to Dow Jones-UBS Commodity Index (DJ-UBSCI) in 2009, when UBS acquired the index from AIG. On July 1, 2014, the index was rebranded under its current name.

some commodities in the index to the rest. Such "non-fundamental shocks" can reflect investor sentiments and/or transitory price pressures demanded by market makers for providing liquidity (see Campbell, Grossman and Wang, 1993; Nagel, 2012).[5]

Figure 3 provides some supporting evidence. We plot the return AR(1)s of three indices against our total index exposure measure where all series are estimated in 10-year backward rolling windows with a minimum window length of 2 years. We see a clear negative relationship between the index autocorrelation and our index exposure measure (-0.32 for GSCI and -0.42 for BCOM). In other words, during periods when institutional investors trade commodity index more actively, the commodity index return (on both GSCI and BCOM) becomes more negatively autocorrelated. No such negative relation is observed for the portfolio of non-indexed commodities (NIDX). In fact, the correlation between NIDX and total index exposure measure is positive (0.25).

[Figure 3 is about here.]

The rest of the paper provides additional evidence that links financialization to negative return autocorrelation among indexed commodities.

We ran three sets of tests. In the first, we provide direct evidence that index trading propagates investor sentiment across commodities in the same index. We measure daily sentiment on a commodity as the deseasonalized *negative* news tone innovation. We then study the spillover of such sentiment across indexed commodities. Take an indexed commodity, corn, as an example. We compute the "connected" index sentiment by averaging the sentiment measures on other non-Grains indexed commodities (such as energy, metal, etc.). We find that the "connected" index sentiment is negatively related to contemporaneous return on corn, but to predict corn's next-day return positively and significantly. The fact that such a negative correlation reverts on the next day confirms the existence of "non-fundamental" shocks. As index trading propagates such shocks

[5]For example, Goldstein and Yang (2017) argue that hedging demand from financial traders tend to consume liquidity and push prices temporarily away from fundamental values.

across commodities in the same index, it results in synchronized price overshoots and reversals and therefore "excessive" comovement. We confirm that the results are not driven by the 2008-2009 great financial crisis. As a placebo test, we repeat the same tests among non-indexed commodities but do not find evidence for such "non-fundamental" shocks.

To directly measure the extent of index trading in the commodity market, we construct a novel "index exposure" measure by combining data on commodity futures trading volumes with data on index trader positions calculated using the Hamilton and Wu (2015) method. We confirm that the sentiment propagation results are much stronger during periods when the commodity market is more exposed to index trading.

Our second set of tests directly links return autocorrelation to index trading by extending the analysis in Figure 3. We find the daily return reversal among index commodities to be stronger during the high index exposure period. Alternatively, when we regress the daily autocorrelation of an index commodity on its lagged index exposure measure, we find a significantly negative coefficient. Such a return reversal is also economically significant. A trading strategy implemented in real-time to take advantage of the return reversal at index level delivers a Sharpe ratio of 0.69 during the high index exposure period, even after accounting for direct transaction cost. In sharp contrast, non-indexed commodities show a strong positive autocorrelation which gives rise to a profitable time-series momentum trading strategy. It is possible that without newly arriving financial investors, information diffusion remains slow in non-indexed commodities, causing return momentum. Additional analyses ensure that our results are not driven by futures roll dates and are robust to different definitions of our index exposure measure.

Our third test aims at establishing causality from commodity index trading to excessive comovement in the commodity index. We take advantage of the fact that the same commodity can receive very different weights across two popular commodity indices (GSCI and BCOM). The relative weight difference arises in a rather ad-hoc fashion and is determined at the beginning of each year. We find that the negative daily return autocorrelation on commodities overweighted in GSCI (relative to BCOM) correlates more with the measure of index trading based on GSCI (relative to that based on BCOM).

Our paper is related to two strands of literature. First, it contributes to the debate on the price impact of index investments in the commodity market. Henderson, Pearson and Wang (2014) document that the hedging activities of issuers of commodity-linked notes can significantly influence commodity futures prices. Gilbert (2010) and Singleton (2013) show that index investments predict movements of oil prices. Ready and Ready (2019) document that order flows from index traders do influence commodity prices. Chen, Dai and Sorescu (2019) document that the aggregate assets under management of CTAs can predict return correlations between CTAs and the stock market. Mou (2010) and Yan, Irwin and Sanders (2019) document that index rebalancing causes futures prices shift significantly. Using a theoretical model, Basak and Pavlova (2016) show that the excess correlation among commodities can arise if institutional investors care about outperforming a commodity index. Sockin and Xiong (2015) theoretically show that financial inflows and outflows (through index investing) to commodity markets can be misread as a signal about global economic growth if informational frictions exist in the commodity futures markets. Consistent with this paper, the recent empirical paper by Brogaard, Ringgenberg and Sovich (2019) shows inefficient commodity prices can distort real decisions of a firm. However, Büyükşahin and Harris (2011) and Irwin and Sanders (2012) find little evidence that the index position changes link to price movements in futures markets. Hamilton and Wu (2015) presents a mixed result. In a review article, Cheng and Xiong (2014) call for direct tests of price impacts with clear identification strategies. Our paper moves closer to answering their challenge. By focusing on autocorrelations, our empirical setting allows better identification of the price impact of commodity index trading. Particularly, prices of indexed commodities over-shoot and reverse subsequently when reacting to non-fundamental shocks, while non-indexed commodities do not show such a reversal pattern.

Second, our paper also speaks to existing literature that links indexing to side effects, mostly in equity markets. Such side effects include the amplification of fundamental shocks (Hong, Kubik and Fishman, 2012), non-fundamental price changes (Chen, Noronha and Singal, 2004), excessive comovement (Barberis, Shleifer and Wurgler, 2005; Greenwood, 2005, 2008; Da and Shive, 2018), a deterioration of the firms' information environment (Israeli, Lee and Sridharan, 2017), increased non-fundamental volatility in individual stocks (Ben-David, Franzoni and Moussawi, 2018), and reduced welfare of retail investors (Bond and García, 2019). Our results indicate that similar side effects may exist in the commodity market as well.

The remainder of the paper goes as follows. Section 2 describes the data and constructs variables used in this research. Section 3 presents the empirical results. Section 4 conducts some robustness checks and Section 5 concludes.

2 Data and Variable Construction

In this section, we describe the commodities used in our analyses and introduce the two most popular commodity indices and their construction. We then discuss our news database and how we construct a news-based sentiment measure for each commodity. Finally, we describe how we measure the exposure of a commodity to index trading. A summary of our key variables and notations is provided at the end of the paper.

2.1 Commodities and commodity indices

Commodity price data are obtained from Commodity Systems Inc. (CSI). Following Kang, Rouwenhorst and Tang (2019), we compute the daily excess return for each commodity using the nearest-to-maturity (front-month) contract and we roll positions on the 7th calendar day of the maturity month into the next-to-maturity contract.[6] The excess return r_{it} on commodity i on date t is calculated as:

$$r_{it} = \frac{F_i(t,T) - F_i(t-1,T)}{F_i(t-1,T)}. \tag{1}$$

[6] If the 7th is not a business day, we use the next business day as our roll date.

where $F_i(t,T)$ is the futures price on day t for a futures contract maturing on date T. To mitigate the effect of outliers, we winsorize the returns at the top and bottom 1 percentile.

Table 1 lists the 27 commodities we examined. They are categorized into five sectors: Energy, Grains, Livestock, Metals, and Softs. Futures listing exchanges and coverage periods are also provided for each commodity.

[Table 1 is about here.]

The recent financialization makes it easy for institutional investors to trade various commodity indices. A commodity index functions like an equity index, such as the S&P 500, in which its value is derived from the total value of a specified basket of commodities. Currently, the largest two indices by market share are the S&P Goldman Sachs Commodity Index (GSCI) and the Bloomberg Commodity Index (BCOM). These two indices use different selection criteria and weighting schemes: the GSCI is weighted by the world production of each commodity, whereas the BCOM focuses on the relative amount of trading activity of a particular commodity. Importantly, for both indices, the weights are set at the beginning of the year and do not vary during the year. Table 1 provides index membership information for each of the 27 commodities in our sample.

We collect the daily price data of GSCI and BCOM from Yahoo finance and calculate their daily returns as $(P_t - P_{t-1})/P_{t-1}$. We also construct an equal-weighted non-indexed commodities index (NIDX) and calculate its daily returns by simply equally averaging the daily returns across non-indexed commodities. Table 2 provides summary statistics regarding daily returns on individual commodities and commodity indices during our sample period from 2003 to 2018.

[Table 2 is about here.]

Commodities offer attractive annual Sharpe ratios that are comparable to that in the equity market. In addition, their return correlations with the equity market before financialization are fairly low (Tang and Xiong, 2012). Not surprisingly, institutional investors became more willing to invest in commodities as a diversification to the mainstream stock and bond markets, especially since the start of financialization that makes it easy for them to trade commodity indices.

The energy sector, especially crude oil (CL) and natural gas (NG), did not perform well in our sample period. Since both GSCI and BCOM indices place heavy weights in the energy sector, both indices suffered losses in the same period. Non-index commodities, as a group (NIDX), have earned a small positive average daily return of 0.03%.

2.2 Commodity sentiment measure

The news data we use come from the Thomson Reuters News Analytics - Commodities data (TRNA-C). TRNA-C data provides 3 news tones (positive, negative and neutral) for each piece of commodity news and the sample coverage starts from January 2003.[7] By averaging all the news tones on each piece of news in a trading day for each commodity, we obtain a daily panel of 3 news tones for each commodity.

For each commodity, we first regress the *negative* news tone[8] on its first lag and the weekday dummies by running the following regression:[9]

$$Tone_t^{Neg} = \alpha + \beta\, Tone_{t-1}^{Neg} + \phi\, Weekday_t + \epsilon_t. \tag{2}$$

We then treat the residual of the regression ($\widehat{\epsilon}_t$) as the negative sentiment measure for each commodity. The descriptive statistics of our sentiment measure for each commodity are shown in Table 3.

[7] According to the TRNA-C manual, the news tones are calculated base on neural network algorithm and the reported accuracy is around 75%.

[8] Tetlock (2007) pointed out that negative tones are better measured in most of the textual data. We also report our results using net (positive minus negative) news tone in the Appendix A.3 for robustness checks.

[9] Wang, Zhang and Zhu (2018) show that news has a "momentum" effect, i.e. the current news sentiment depends significantly on its lagged level. Several studies (Hafez, 2009, 2011; Healy and Lo, 2011, etc.) have reported strong seasonality in news flows at various sampling frequencies, e.g., intrahour, intraday and intraweek. Therefore, we include the lagged news level and weekday dummies to filter out the potential momentum effect and seasonality in news tones.

[Table 3 is about here.]

As evident in Table 3, crude oil receives more news coverage than other commodities. The sentiment measures have zero means by construction. Their average standard deviation is 0.0618 ranging from 0.0306 for oat (O-) and rough rice (RR) to 0.1038 for soybean (S-).

2.3 Commodity index exposure

Every Friday, the CFTC releases a weekly Commitments of Traders (CoT) report with data collected on previous Tuesday, which includes the total open interest of each commodity and the long/short positions of each type of traders.[10] It also includes a supplemental Commodity Index Trader (CIT) report that shows positions of a set of index traders identified by the CFTC since January 3, 2006. Due to the need for the CIT data, all our regression analyses start in 2006.

According to the manual of CIT, the total open interest in the supplementary CIT report can be recovered from the 9 components that are detailed in the report:

$$2(\mathit{Open\ Interest}^{All}) = \underbrace{(\mathit{Long} + \mathit{Short} + 2\mathit{Spread})}_{\mathit{Non\text{-}commercial}} + \underbrace{(\mathit{Long} + \mathit{Short})}_{\mathit{Commercial}} + \underbrace{(\mathit{Long} + \mathit{Short})}_{\mathit{Index\ Trading}} + \underbrace{(\mathit{Long} + \mathit{Short})}_{\mathit{Non\text{-}reportable}}. \quad (3)$$

Naturally, we can define the index open interest as the average of the long and short positions of index traders: $\mathit{Open\ Interest}^{Idx} = (\mathit{Long}^{Idx} + \mathit{Short}^{Idx})/2$. Based on these data, we can estimate the index trader market share of an indexed commodity i on day t as the ratio of its index open interest to its total open interest during the prior week, i.e.,

$$\mathit{Index\ Market\ Share}_{it} = \frac{\mathit{Open\ Interest}^{Idx}_{i,w(t)}}{\mathit{Open\ Interest}^{All}_{i,w(t)}}, \quad (4)$$

where $w(t)$ denotes the Tuesday immediately before or exactly on day t.

The CIT report only contains 13 agricultural commodities (listed in Table 1) but covers no

[10] The traders are classified into three types: commercial (C), noncommercial (NC), and non-reportables (NR). In CIT report, CFTC separates the index trading positions (Idx) from the positions of the commercial traders.

commodities in the energy and metals sectors. Masters (2008) first introduced an interpolation method to estimate the nonreported indexed commodities' position by taking advantage of the difference in commodity coverages between GSCI and BCOM. Hamilton and Wu (2015) recently refines Masters' approach through a regression setting. We thereby employ Hamilton and Wu (2015) method to obtain each non-reported indexed commodity's estimated index market share. Appendix A.1 provides a detailed description of Masters (2008) method and Hamilton and Wu (2015) method.

Based on the estimated index market share, we obtain each commodity's index trading volume[11] as below

$$\textit{Index Trading Volume}_{it} = \textit{Index Market Share}_{it} \times \textit{Trading Volume}_{it}, \tag{5}$$

and define the index exposure of a commodity i on day t as the standardized version of detrended index trading volume with past 250-day average in the spirit of Campbell, Grossman and Wang (1993), i.e.,

$$\textit{Index Exposure}_{it} = \text{standardize}\left\{\textit{Detrended Index Trading Volume}_{it}\right\}. \tag{6}$$

Detrending is useful because commodity trading volumes have trended up during our sample period with the implementation of the electronic trading system and lower broker charges. The standardization makes it possible to compare trading activities among commodities with different contract sizes. Note that the trading volume is measured in numbers of contract so price information does not enter in our measure of index exposure for an individual indexed commodity.

Finally, the total index exposure for the commodity market is computed as a simple average of

[11]Since the nearest and the second-nearest contracts are most liquid and considering commodity indices' rolling activity (see Stoll and Whaley, 2010; Mou, 2010, etc.), we calculate the total trading volume of each commodity as the sum of trading volume on the nearest and second-nearest contracts.

the index exposures across all I index commodities, i.e.,

$$\textit{Total Index Exposure}_t = \frac{1}{I}\sum_{i=1}^{I} \textit{Index Exposure}_{it}. \tag{7}$$

The total index exposure can be therefore interpreted as the amount of abnormal trading volume on day t that reflects index trading.

3 Empirical Analysis

We conduct three sets of empirical analyses. We first study the propagation of "sentiment" shocks across commodities using our news-based measures. We then focus on daily return autocorrelations for indexed commodities and relate them to measures of their index exposure. Finally, we provide causal evidence that index trading drives negative index return autocorrelations.

3.1 Sentiment spillover

To study the sentiment spillover across the indexed commodities, we construct a "connected" sentiment measure for each commodity. Take corn (C-) for example. To construct its "connected" sentiment on day t, we take a weighted average of sentiment measures on all other indexed commodities from other sectors on that day, i.e.,

$$\textit{Cnn. Sentiment}_{it} = \sum_{S(j)\neq S(i)} W_{jy(t)} \textit{Sentiment}_{jt}, \tag{8}$$

where $S(i)$ is the sector that commodity i belongs to, and the weight $W_{jy(t)}$ is defined as

$$W_{jy(t)} = \frac{E_{y(t)}(\$\textit{Open Interest}^{Idx}_{jt})}{\sum_j E_{y(t)}(\$\textit{Open Interest}^{Idx}_{jt})}, \tag{9}$$

with $E_{y(t)}(\$Open\ Interest^{Idx}_{jw(t)})$ being the average of the weekly dollar-valued open interest on index trading in year $y(t)$. In other words, the weight on "connected" indexed commodity j is determined by its average dollar-valued open interest relative to total dollar-valued open interests across both indices.

In the above definition, the set of indexed commodities "connected" to corn only includes indexed commodities from other sectors such as energy and metals, but not other indexed commodities from the same grains sector such as soybean (S-) and wheat (W-). To the extent that sentiment measure includes commodities from the same sector may still contain fundamental factors,[12] they are more likely to co-move within sector than across sectors. In this sense, our measure alleviates the concerns for fundamental-driven comovements among commodities in the same sector. As a placebo test, we construct the "connected" sentiment measure for non-indexed commodities in the same fashion, except that we use an equal weighting scheme as in the construction of NIDX.

Based on the "connected" sentiment measure, we run the following day/commodity panel regressions to examine both contemporaneous and predictive relations between the "connected" sentiment measure and the commodity returns, for indexed and non-indexed commodities separately,

$$r_{it} = \beta_0 + \beta_1 \cdot Cnn.\ Sentiment_{it} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \tag{10}$$

$$r_{it} = \beta_0 + \beta_1 \cdot Cnn.\ Sentiment_{it-1} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \tag{11}$$

where $\boldsymbol{X}$ is a vector of control variables including lagged (log) basis,[13] lagged Amihud's (2002)

[12] For example, as shown in Casassus, Liu and Tang (2012), different commodities from the same sector are likely to have a fundamental relationship of production, substitution etc.

[13] The log basis is defined as

$$Basis_{it} = \frac{\ln(F_i(t,T_1)) - \ln(F_i(t,T_2))}{T_2 - T_1},$$

where $F_i(t,T_1)$ and $F_i(t,T_2)$ are futures prices of the nearby and second nearby contracts with T_1 and T_2 as their maturities correspondingly.

illiquidity[14], and realized volatility over past 21 trading days for each commodity.[15] Both sector fixed effect and year fixed effect are also controlled in the regression. Note that Szymanowska *et al.* (2014) document that (log) basis, volatility and liquidity might serve as determinants on the risk premium of commodity markets. We thus use these variables as controls. To assess the difference between the coefficients for the indexed and non-indexed commodities, we also run the regressions with an interaction term between connected sentiment and a dummy variable (1 for indexed commodities and 0 for non-indexed commodities). The results are reported in Table 4.

[Table 4 is about here.]

Focusing on Panel A, we find a negative and significant contemporaneous relation between the index commodity return and its "connected" sentiment measure in Column 1. Our "connected" sentiment measure may still contain "fundamental" information that affects all commodities,[16] explaining why its contemporaneous return correlation is also negative and significant for non-indexed commodities in Column 2 where index trading is not possible. Nevertheless, the negative coefficient (-0.1571) is significantly more negative than that for the non-indexed commodity (-0.0931), consistent with the notion that index trading propagates negative sentiment, in addition to fundamental information, across commodities within the same index.

While both sentiment propagation and fundamental information can explain the negative con-

[14]For each commodity, we compute its Amihud's (2002) illiquidity measure as the ratio of the absolute value of the daily return of the nearest contract divided by its corresponding trading volume (in billion dollars) on the same day.

$$Illiquidity_{it} = \frac{|r_{it}|}{(\$billion)\, Trading\ Volume_{it}}.$$

To mitigate the effect of outliers, we first winsorize the illiquidity measure at the top and bottom 1 percentile and then do the standardization.

[15]The *Rvol* is defined as the annualized realized volatility, i.e.,

$$Rvol_{it} = \sqrt{12\sum_{k=0}^{20}[\ln(F_i(t-k,T_1)) - \ln(F_i(t-k-1,T_1))]^2}.$$

[16]For example, business-cycle factors can influence the demand and supply for all commodities.

temporaneous relation we observed in Panel A, only sentiment predicts future return reversals. This is because sentiment-induced trading induces a "non-fundamental" shock in contemporaneous return and such a shock will be reverted in the future. For example, negative sentiment on energy may induce institutional investors to sell the commodity index. Such trading propagates the negative sentiment from the energy sector to other indexed commodities and results in negative price pressure in corn today. As the negative price pressure on corn reverts tomorrow, the "connected" sentiment today should positively predict corn's return tomorrow.[17] In contrast, the "connected" sentiment should not predict the return on non-indexed commodities.

Positive and significant return predictability by "connected" sentiment is exactly what we find in Panel B, but for indexed commodities only. The coefficient on "connected" sentiments is likely to capture the impact of sentiment spillover. For instance, a predictive coefficient of 0.0104 (t-value of 2.99) on the "connected" sentiment implies that a one-standard-deviation increase in the sentiment of "connected" indexed commodities propagates a price pressure that reverts by 2.3 basis points the next day. Column 2 in Panel B does not find any significant return predictability among non-indexed commodities. The difference between the indexed and non-indexed commodities is also quite large (0.0152) and statistically significant as shown in the third column of Panel B.

Turning to the control variables, consistent with Szymanowska *et al.* (2014) and Gorton, Hayashi and Rouwenhorst (2012), lagged basis makes a positive prediction (although insignificant) on commodity returns listed in Table 4. Both liquidity and volatility do not significantly predict commodity returns on daily frequency.

If index trading propagates sentiment and creates price pressure at the index level, we should observe stronger effect during times when index trading exposure is abnormally high. To test this conjecture, we split the sample into two subsamples based on our total index exposure measure defined in the previous section. Specifically, we classify the trading day whose total index exposure

[17]Such trading may also propagate fundamental information specific to energy to the contemporaneous return of corn. The energy-specific information, by definition, should not affect corn and is therefore observationally equivalent to "non-fundamental" shock to corn. We thank Liyan Yang for clarifying this point.

is above (below) zero as "High" of H ("Low" or L) index exposure period. We then re-run the previous regression analyses in the "H" and "L" subperiods separately:

$$r_{it} = \beta_0 + \beta_1 \cdot \textit{Cnn. Sentiment}_{it} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \quad t \in \{H, L\}, \tag{12}$$

$$r_{it} = \beta_0 + \beta_1 \cdot \textit{Cnn. Sentiment}_{it-1} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \quad t \in \{H, L\}. \tag{13}$$

Both sector fixed effect and year fixed effect are controlled in the regression. The results are reported in Table 5.

[Table 5 is about here.]

Focusing on the sentiment return predictability results in Panel B, we find that the return reversal is only significant during the "High" period for the indexed commodities. The coefficient on the sentiment measure is 0.0166 (t-value of 3.55) in trading days with a high amount of index trading. The economic magnitude is large. A coefficient of 0.0166 implies that a one-standard-deviation increase in the sentiment of connected indexed commodities propagates a price pressure of at least 4.0 basis points. Consistent with the notion that index trading results in price overshoot and reversal, when we focus our attention on non-indexed commodities, we observe no return reversals in either"High" or "Low" index exposure period.[18]

So far, our results using news-based sentiment measures provide direct evidence that as index trading propagates "non-fundamental" sentiment shocks across commodities in the same index, and creates correlated price overshoots and reversals at daily frequency. Such excessive comovements will result in negative daily return autocorrelations even at the index level. In the next subsection, we therefore focus our attention on daily return autocorrelation measures.

[18] Note that non-indexed commodities has a significantly negative coefficient in the "Low" index exposure period, indicating a delayed reaction to negative sentiment which results in momentum instead of reversal.

3.2 Return autocorrelation and index exposure

Focusing on the daily return autocorrelation, Figures 2 and 3 confirm the link between index trading and excessive comovement in the commodity market. With a backward rolling window of ten years, we observe a continuing decline in the average index daily return autocorrelations for both GSCI ad BCOM indices in Figure 2. They became significantly negative since 2006 when financialization made index trading easy. A negative return autocorrelation unambiguously signals excessive comovement and price inefficiency at the index level. In sharp contrast, no such decline is observed for the average daily return autocorrelation for a portfolio of non-indexed commodities (NIDX). Instead, it has an increasing trend during the sample period.

Using our index exposure measure, Figure 3 shows a strong negative relation between the index return autocorrelation and the index exposure. In this subsection, we examine and confirm the linkage between index trading and index return autocorrelation at daily frequency, taking advantage of the high-frequency nature of our measure. Following Baltussen, van Bekkum and Da (2019), we directly link autocorrelation measure to (lagged) total index exposure in Table 6. In particular, we regress the commodity return autocorrelation measure, $r_{it}r_{i,t-1}/\sigma_i^2$, on the lagged total index exposure and other controls (log basis, Amihud illiquidity and realized volatility), i.e.,

$$r_{it}r_{i,t-1}/\sigma_i^2 = \beta_0 + \beta_1 \mathit{Total\ Index\ Exposure}_{t-1} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}. \tag{14}$$

Both individual fixed effect and year fixed effect are included in the regression.

[Table 6 is about here.]

We confirm that the return autocorrelation of indexed commodities become more negative when total index exposure is higher. Specifically, a coefficient of -0.0845 means that a one-standard-deviation increase in the total index exposure makes its daily return autocorrelation 3.85% more negative for indexed commodities. On the contrary, non-indexed commodities do not show such a pattern. The different behavior between indexed and non-indexed commodities is significant as evident in Column 3, consistent with the notion that index trading results in the overshooting of prices and reversal in the subsequent period only among indexed commodities.

Next, in Table 7, we evaluate the economic significance associated with these autocorrelation patterns at the index level using several index trading strategies. For example, we study a contrarian strategy based on the short-term return reversal for commodity indices (GSCI and BCOM). Specifically, for the contrarian strategy, we sell (buy) the GSCI or BCOM when their returns on the previous trading day are positive (negative). Following Baltussen, van Bekkum and Da (2019), we take a position r_{t-1} so that the daily return on our strategy is simply $-r_t r_{t-1}$. According to Column 1, the trading strategy has an annual Sharpe ratio of 0.49 for GSCI (Panel A) and 0.38 for BCOM (Panel B) for the 2006-2018 sample period, consistent with Figure 2 which shows a significantly negative daily autocorrelation for both indices since 2006.

[Table 7 is about here.]

Commodity futures contracts are fairly liquid and easy to trade. Nevertheless, to account for trading cost, we use the weighted average of one tick bid-ask spreads for indexed commodities (1.04 bps for GSCI and 1.26 bps for BCOM) and the weighted average of two ticks bid-ask spreads for non-indexed commodities (7.74 bps for NIDX).[19] Column 1 shows sizable annual Sharpe ratios even after transaction costs (0.45 for GSCI and 0.31 for BCOM (Panel B).

Table 5 documents a stronger return reversal for indexed commodities when their index exposure is high. The trading strategy confirms this pattern. Since our index exposure measure is constructed using a full-sample standardization procedure, it is not observable in real-time strictly speaking. To ensure that our trading strategy can be implemented in real-time, in Column 2, we reconstruct a real-time index exposure measure where the standardization procedure is done using a backward

[19] Arzandeh and Frank (2019) show that the bid-ask spreads of large agricultural commodities are about one tick, those of small agricultural commodities are slightly less than two ticks. We take half tick as trading costs of indexed commodities and one tick as trading costs of non-indexed commodities.

250-day rolling window. Using this real-time measure, we find even stronger Sharpe ratios. The annual after-cost Sharpe ratio improves to 0.69 for GSCI and 0.64 for BCOM during the high index exposure periods.

Results in Panels A and B demonstrated that return reversals among indexed commodities are highly significant economically, especially during the high index exposure period. When we focus on non-indexed commodities, we see a different yet robust momentum pattern in Figure 2. To evaluate its economic significance, we consider a momentum trading strategy. Specifically, we buy (sell) the equally-weighted portfolio of non-indexed commodities (NIDX) when its return on the previous trading day is positive (negative). We still take a position r_{t-1} so that the daily return on our strategy is simply $r_t r_{t-1}$. The results are reported in Panel C.

The momentum pattern on NIDX also seems economically significant. Its annual after-cost Sharpe ratio is 0.51 in during the full 2006-2018 sample period. Interestingly, the Sharpe ratio does not change much when we focus on high index exposure periods (0.48 in Column 2). The result is consistent with what we find in Table 6 that total index exposure is not significantly related to the autocorrelation of non-indexed commodities. Overall, the momentum pattern on NIDX serves as a nice placebo. The momentum here could reflect continuing under-reaction to common shocks among non-indexed commodities as they receive little attention from index investors.

Since individual index exposure is not necessarily high when the total index exposure is high, we also conduct the following daily panel regression of each commodity's serial dependence measure on the lagged individual index exposure measure and controls:

$$r_{it} r_{i,t-1}/\sigma_i^2 = \beta_0 + \beta_1 \cdot \mathit{Index\ Exposure}_{i,t-1} + \theta' \boldsymbol{X}_{i,t-1} + \epsilon_{it}, \tag{15}$$

where σ_i^2 is the sample variance of commodity i's returns, and vector $\boldsymbol{X}$ contains each commodity's lagged basis and lagged Amihud illiquidity as control variables motivated by Baltussen, van Bekkum and Da (2019), Nagel (2012) and Bianchi, Drew and Fan (2016). We also control the lagged implied volatility of crude oil options with nearest maturity in the regressions to control for systematic volatility shock in commodity markets in the spirit of Christoffersen and Pan (2018).[20] We run the panel regression for indexed and non-indexed commodities separately, and use the total index exposure as non-indexed commodities' index exposure.

[Table 8 is about here.]

Table 8 reveals two sets of interesting results. First, we observe negative and significant coefficients on the index exposure measure only for the indexed commodities (see Columns 1 and 2). In other words, abnormally high index trading today implies a more negative correlation between the indexed commodity return today and that tomorrow, consistent with the notion that index trading results in price pressure at the index level today and such a price pressure is reverted tomorrow. The economic magnitude of such effect is large. For example, a coefficient of -0.0241 means that a one-standard-deviation increase in the index exposure makes its daily return autocorrelation 2.41% more negative.

Second, to the extent that the reversal on indexed commodities reflects price pressure induced by index trading, we expect the reversal to be stronger when the market liquidity is poor. Columns 3 confirms this conjecture. The coefficient on the interaction term between lagged index exposure and illiquidity measures is negative and highly significant. In other words, when index investors trade during periods of illiquidity, their trading more likely generates negative return autocorrelation for commodities in the index. Column 4 again shows no such interaction among non-indexed commodities.

Since our index exposure measure is a detrended product of total trading volume and index market share, a natural concern is that our results could be completely driven by the total trading volume component rather than the index market share. To address this concern, we rerun regression (15) by separately including the two components of the index exposure measure. The

[20] Christoffersen and Pan (2018) shows that shocks to oil volatility are strongly related to various measures of funding constraints of financial intermediaries, which is arguably a key driver of pricing kernel dynamics.

results are summarized in Table 9. They show that both components are important in driving the autocorrelation of indexed commodity returns. The economic magnitude of both components are significant, e.g., coefficients of -0.4118 and -0.0003 in Column 3 indicate that a one-standard-deviation increase in each component will result in a decrease in daily return autocorrelation by 3.31% and 2.43% respectively. Hence our results are not driven only by index market share nor only by trading volume. Both components, as nicely combined in our index exposure, contribute to our results. Consistent with the previous analysis, both components show no significant impact on the return autocorrelations in non-indexed commodities. This nice placebo result confirms that our analysis is robust to different specifications of index exposure measure.

[Table 9 is about here.]

3.3 Causal evidence

Can some missing factors drive the link between index trading and negative daily return autocorrelation we documented so far? Maybe in the last 15 years, institutional investors simply became more willing to invest in a basket of certain commodities as an asset class. Such an investment demand will result in correlated order flow across these commodities and will result in negative commodity portfolio return autocorrelations, regardless of whether commodities index products have been introduced or not. It is simply a coincidence that part of that correlated order flow is also satisfied through index products (rather than through trading the underlying commodity futures directly). One could even argue that the commodity indexed products were introduced precisely to cater for correlated demand from institutional investors in trading these commodities (that are chosen to be included in GSCI and BCOM indices).

While such a correlated demand story could explain the low-frequency trends displayed in Figure 3, it is harder to explain the high-frequency relation (between index exposure measure and negative daily return autocorrelation) we documented in Tables 6 through 9. It is unlikely a broadly increasing trend to invest in broad commodity baskets should be highly correlated with abnormal trading activities in two specific commodity indices on a day-to-day basis. Nevertheless, in this section, we conduct additional tests, aiming at pinning down the causality from index trading on index return autocorrelation.

The additional causality tests are similar in spirit to those in Greenwood (2008) and Baltussen, van Bekkum and Da (2019) that take advantage of different weighting schemes across two Japanese equity indices. Similar to the case of equity indices, the same commodity can receive different weights across GSCI and BCOM indices. The relative weighting is determined in a fairly ad-hoc fashion, and importantly for our purpose, is determined at the beginning of the year and then held constant throughout the year. A testable implication of index trading therefore goes as follows: for commodities that are overweighted in GSCI index (relative to BCOM index), its daily return autocorrelation should be more negatively correlated with the trading measure on GSCI (relative to that on BCOM).

We implement the test by first calculating each indexed commodity's average weights on GSCI and BCOM over 2004 to 2018. We pick the top 3 commodities traded in both indices that are most overweighted on GSCI (denoted as G3) and BCOM (denoted as B3) separately. This selection rules out the commodities whose weight differences are less than 1% and thus helps us to see the real effect of indexing.

Next, we compute the selected commodities' GSCI exposure and BCOM exposure separately. Similar to the individual index exposure measure introduced in (6), the commodities' exposure on a specific commodity index is defined as each commodity's GSCI/BCOM market share times the total trading volume and then detrended with a 250-day backward rolling window. To compute the market share of a specific commodity index, we first employ Hamilton and Wu (2015) method to estimate the indexed commodities' open interest on that index (see Appendix A.1 for more details). Then, we obtain commodity i's GSCI/BCOM market share as well as its index exposure on GSCI/BCOM as below,

$$\textit{Index Market Share}_{it}^{p} = \frac{\textit{Open Interest}_{it}^{p}}{\textit{Open Interest}_{it}^{All}}, \tag{16}$$

$$\textit{Index Trading Volume}_{it}^{p} = \textit{Index Market Share}_{it}^{p} \times \textit{Trading Volume}_{it}, \tag{17}$$

$$\textit{Index Exposure}_{it}^{p} = \text{standardize}\left\{\textit{Detrended Index Trading Volume}_{it}^{p}\right\}, \tag{18}$$

where $p \in \{GSCI, BCOM\}$.

Finally, we regress the G3 and B3 return serial dependence measures on the lagged GSCI and BCOM exposure measure with controls separately, i.e.,

$$r_{it}r_{i,t-1}/\sigma_i^2 = \beta_0 + \beta_1 \cdot \textit{Index Exposure}_{i,t-1}^{GSCI} + \beta_2 \cdot \textit{Index Exposure}_{i,t-1}^{BCOM} + \theta' \boldsymbol{X}_{i,t-1} + \epsilon_{it}, \tag{19}$$

where $i \in$ G3 or B3 and $\boldsymbol{X}$ is a vector of control variables motivated by Bianchi, Drew and Fan (2016), Nagel (2012) and Christoffersen and Pan (2018), which contains each commodity's lagged basis, lagged Amihud illiquidity, and lagged implied volatility of crude oil options with nearest maturity.

[Table 10 is about here.]

The results in Table 10 strongly support a causal interpretation that index trading drives excessive comovement and negative index return autocorrelation. For commodities that are relatively overweighted in index i, its daily return autocorrelation is indeed more negatively correlated with trading exposure to index i.

4 Robustness Checks

In this section, we perform robustness checks with different data samples (excluding financial crisis, or rolling periods) and using different measures for the news tones.

4.1 Financial crisis

The financial crisis may drive some of the results in our paper. Following Tang and Xiong (2012), we hence choose the time period from September 15, 2008, when is the Lehman Brothers filed bankruptcy, to June 30, 2009, when is the trough of the business cycle identified by the NBER, as the period of the financial crisis. We therefore redo the regression in Table 4, 6 and 8 excluding the financial crisis period, with results reported in Table A1, A2 and A3. Our robustness-check results are consistent with Table 4, 6 and 8. That is, through the index investment, connected news sentiments lead to a price overshoot and a subsequent reversal, and index exposure decreases in futures return autocorrelation. On the contrary, non-indexed commodities do not have such effects.

4.2 Net news tone

In Section 3.1 we use negative news tone in regression, as a robustness check, we redo the regression in Table 4 and 5 using net news tone (positive tone - negative tone). The results are listed in A4 and A5. Again, by using net news tone, we obtain similar results with the ones with negative tones (as shown in Table 4 and 5). Table A6 presents the net news tone results excluding financial crisis, which is still consistent with our main results in Table 4.

4.3 Index rolling activity

Unlike equity index funds that invest directly in the underlying assets, commodity index funds trade futures contracts instead, which requires them to unwind the maturing contracts before they expire and roll their positions to the contracts with later maturity dates. According to the rolling schedule of GSCI and BCOM, both indices shift the basket of contracts from the first to the second nearby contracts at a rate of 20% per day, on the fifth through ninth business days in each month. This routine rolling activity will result in abnormally high index trading volume during the roll period and will likely impact our index exposure measure. Therefore, it would be important to make sure that our results are not driven by these roll dates.

For each commodity, we identify the week containing the roll date of its continuous contract, which is the seventh calendar day of the maturity month.[21] We then re-run the panel regression (14) and (15) on a sample excluding roll weeks and report the results in Table A7 and A8. The results, when excluding roll weeks, are very similar to those using the whole sample, suggesting that index-rolling is not the driver of our findings.

In addition, we reconduct the analyses in Table 9 and the report the results in Table A9. It shows that our results are jointly robust to different index exposure definitions and commodity indices rolling activities.

5 Conclusion

We examine the impact of recent financialization in the commodity market on return serial dependence of indexed commodities. Using news-based sentiment measures, we find that index trading can propagate non-fundamental shocks from some commodities to others in the same index, giving rise to price overshoots and subsequent reversals, or "excessive comovement" at daily frequency. Excessive comovement results in negative daily commodity return autocorrelations even at the index level (but not for non-indexed commodities) and such autocorrelations move with our commodity index exposure measures. Taking advantage of the fact that index weights of the same commodity can vary across different indices in a relatively ad-hoc and pre-determined fashion, we provide causal evidence that index trading drives return serial dependence.

Given the attractive risk-return tradeoff and the diversification benefits associated with commodity index investments, the commodity financialization process can be expected to continue. We do not dispute such benefits. We simply highlight an unexpected side effect to these benefits as negative serial dependence in commodity index return signals excessive price comovements even at the index level. Price overshooting and the subsequent reversal could impose costs on institu-

[21] Note that our setting can cover most of the index roll dates without affecting the return structure of the continuous contract.

tional investors who trade often and individual investors who invest in commodities through those institutions. Our results agree with the theoretical paper of Goldstein and Yang (2017), which proposes that index traders can inject unrelated noise into the futures prices and hurt the market efficiency. They are also consistent with the finding of Henderson, Pearson and Wang (2014), which documents the hedge trades of newly issued commodity-linked notes have a remarkable pressure on commodity futures prices.

Summary of Variables and Notations

Variable/Notation	Definition
$y(t)$, $w(t)$	Year of time t, Week of time t (Tuesday-Tuesday)
$S(i)$, Idx	Sector of commodity i, Set of indexed commodities
$F_i(t,T)$	Futures price of commodity i at time t with maturity T
$Long^{Idx}_{iw(t)}, Short^{Idx}_{iw(t)}$	Index trader's long (short) position of commodity i in week $w(t)$
Return	$r_{it} = \dfrac{F_i(t,T) - F_i(t-1,T)}{F_i(t-1,T)}$
Log Basis	$Basis_{it} = \dfrac{\ln(F_i(t,T_1)) - \ln(F_i(t,T_2))}{T_2 - T_1}$
Amihud Illiquidity	$Illiquidity_{it} = \dfrac{\lvert r_{it}\rvert}{(\$billion)\ Trading\ Volume_{it}}$
Realized Volatility	$Rvol_{it} = \sqrt{12\sum_{k=0}^{20}[\ln(F_i(t-k,T_1)) - \ln(F_i(t-k-1,T_1))]^2}$
Index Open Interest	$Open\ Interest^{Idx}_{iw(t)} = Long^{Idx}_{iw(t)} + Short^{Idx}_{iw(t)}$
Index Market Share	$Index\ Market\ Share_{it} = \dfrac{Open\ Interest^{Idx}_{i,w(t)-1}}{Open\ Interest^{All}_{i,w(t)-1}}$
Index Trading Volume	$Index\ Trading\ Volume_{it} = Index\ Market\ Share_{it} \times Trading\ Volume_{it}$
Detrended Index Trading Volume	$Index\ Trading\ Volume_{it} - \dfrac{1}{250}\sum_{k=1}^{250} Index\ Trading\ Volume_{i,t-k}$
Individual Index Exposure	$Index\ Exposure_{it} = \text{standardize}\{Detrended\ Index\ Trading\ Volume_{it}\}$
Total Index Exposure	$Total\ Index\ Exposure_{it} = \dfrac{1}{I}\sum_{i=1}^{I} Index\ Exposure_{it}$
News Sentiment	The residual from the regression in equation (2)
Value Weight	$W_{jy(t)} = \dfrac{E_{y(t)}(\$Open\ Interest^{Idx}_{jt})}{\sum_j E_{y(t)}(\$Open\ Interest^{Idx}_{jt})}$
Connected Sentiment	$Cnn.\ Sentiment_{it} = \sum_{S(j)\neq S(i)} W_{jy(t)} Sentiment_{jt}$

References

Ai, C., Chatrath, A. and Song, F. (2006). On the comovement of commodity prices. *American Journal of Agricultural Economics*, **88** (3), 574–588.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, **5** (1), 31–56.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, **24** (2), 3–30.

Arzandeh, M. and Frank, J. (2019). Price discovery in agricultural futures markets: Should we look beyond the best bid-ask spread? *American Journal of Agricultural Economics*, **101** (5), 1482–1498.

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

Baltussen, G., van Bekkum, S. and Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics*, **132** (1), 26–48.

Barberis, N., Shleifer, A. and Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, **75** (2), 283–317.

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

Ben-David, I., Franzoni, F. and Moussawi, R. (2018). Do etfs increase volatility? *The Journal of Finance*, **73** (6), 2471–2535.

Bhardwaj, G., Gorton, G. and Rouwenhorst, G. (2015). Facts and fantasies about commodity futures ten years later. *NBER Working Paper No. 21243.*

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

Bond, P. and García, D. (2019). The equilibrium consequences of indexing. *Working Paper (University of Washington).*

Brennan, M. J. (1958). The supply of storage. *The American Economic Review*, **48** (1), 50–72.

Brogaard, J., Ringgenberg, M. and Sovich, D. (2019). The economic impact of index investing. *The Review of Financial Studies*, **32** (9), 3461–3499.

Büyükşahin, B. and Harris, J. H. (2011). Do speculators drive crude oil futures prices? *The Energy Journal*, **32** (2), 167–202.

Campbell, J. Y., Grossman, S. J. and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, **108** (4), 905–939.

Casassus, J., Liu, P. and Tang, K. (2012). Economic linkages, relative scarcity, and commodity futures returns. *The Review of Financial Studies*, **26** (5), 1324–1362.

Chen, H., Noronha, G. and Singal, V. (2004). The price response to s&p 500 index additions and deletions: Evidence of asymmetry and a new explanation. *The Journal of Finance*, **59** (4), 1901–1930.

Chen, Y., Dai, W. and Sorescu, S. M. (2019). Diversification and financialization in commodity markets: Evidence from commodity trading advisors. *Working Paper, Available at SSRN 3287568.*

Cheng, I.-H. and Xiong, W. (2014). Financialization of commodity markets. *Annual Review of Finance and Economics*, **6** (1), 419–441.

Christoffersen, P. and Pan, X. N. (2018). Oil volatility risk and expected stock returns. *Journal of Banking & Finance*, **95**, 5–26.

Da, Z. and Shive, S. (2018). Exchange traded funds and asset return correlations. *European Financial Management*, **24** (1), 136–168.

Gilbert, C. L. (2010). How to understand high food prices. *Journal of Agricultural Economics*, **61** (2), 398–425.

Goldstein, I. and Yang, L. (2017). Commodity financialization and information transmission. *Rotman School of Management Working Paper No. 2555996.*

Gorton, G. B., Hayashi, F. and Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. *Review of Finance*, **17** (1), 35–105.

Greenwood, R. (2008). Excess comovement of stock returns: Evidence from cross-sectional variation in nikkei 225 weights. *The Review of Financial Studies*, **21** (3), 1153–1186.

Greenwood, R. M. (2005). A cross sectional analysis of the excess comovement of stock returns. *HBS Finance Research Paper No. 05-069*.

Hafez, P. A. (2009). *Detection of Seasonality Patterns in Equity News Flows*. Tech. rep., RavenPack.

— (2011). How news events impact market sentiment. *The Handbook of News Analytics in Finance*, pp. 129–146.

Hamilton, J. D. and Wu, J. C. (2015). Effects of index-fund investing on commodity futures prices. *International Economic Review*, **56** (1), 187–205.

Healy, A. D. and Lo, A. W. (2011). Managing real-time risks and returns: The thomson reuters newsscope event indices. *The Handbook of News Analytics in Finance*, pp. 73–109.

Henderson, B. J., Pearson, N. D. and Wang, L. (2014). New evidence on the financialization of commodity markets. *The Review of Financial Studies*, **28** (5), 1285–1311.

Hong, H., Kubik, J. D. and Fishman, T. (2012). Do arbitrageurs amplify economic shocks? *Journal of Financial Economics*, **103** (3), 454–470.

Irwin, S. H. and Sanders, D. R. (2012). Testing the masters hypothesis in commodity futures markets. *Energy Economics*, **34** (1), 256–269.

Israeli, D., Lee, C. M. C. and Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? an information perspective. *Review of Accounting Studies*, **22** (3), 1048–1083.

Kang, W., Rouwenhorst, K. G. and Tang, K. (2019). A tale of two premiums: The role of hedgers and speculators in commodity futures markets. *The Journal of Finance, Forthcoming*.

Marshall, B. R., Nguyen, N. H. and Visaltanachoti, N. (2011). Commodity liquidity measurement and transaction costs. *The Review of Financial Studies*, **25** (2), 599–638.

Masters, M. W. (2008). *Testimony before Committee on Homeland Security and Governmental Affairs of the United States Senate*. Tech. rep., Commodity Futures Trading Commission.

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

Mou, Y. (2010). Limits to arbitrage and commodity index investment: Front-running the goldman roll. *Working Paper, Available at SSRN 1716841*.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, **25** (7), 2005–2039.

Pindyck, R. S. and Rotemberg, J. J. (1990). The excess co-movement of commodity prices. *The Economic Journal*, **100** (403), 1173–1189.

Ready, M. and Ready, R. C. (2019). Order flows and financial investor impacts in commodity futures markets. *Working Paper, Available at SSRN 3164757.*

Sanders, D. R. and Irwin, S. H. (2011). The impact of index funds in commodity futures markets: A systems approach. *Journal of Alternative Investments*, **14** (1), 40–49.

Singleton, K. J. (2013). Investor flows and the 2008 boom/bust in oil prices. *Management Science*, **60** (2), 300–318.

Sockin, M. and Xiong, W. (2015). Informational frictions and commodity markets. *The Journal of Finance*, **70** (5), 2063–2098.

Stoll, H. R. and Whaley, R. E. (2010). Commodity index investing and commodity futures prices. *Journal of Applied Finance*, **20** (1), 7–47.

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

Tang, K. and Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, **68** (5), 54–74.

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, **62** (3), 1139–1168.

Wang, Y., Zhang, B. and Zhu, X. (2018). The momentum of news. *Working Paper, Available at SSRN 3267337.*

Yan, L., Irwin, S. H. and Sanders, D. R. (2019). Is the supply curve for commodity futures contracts upward sloping? *Working Paper, Available at SSRN 3360787.*

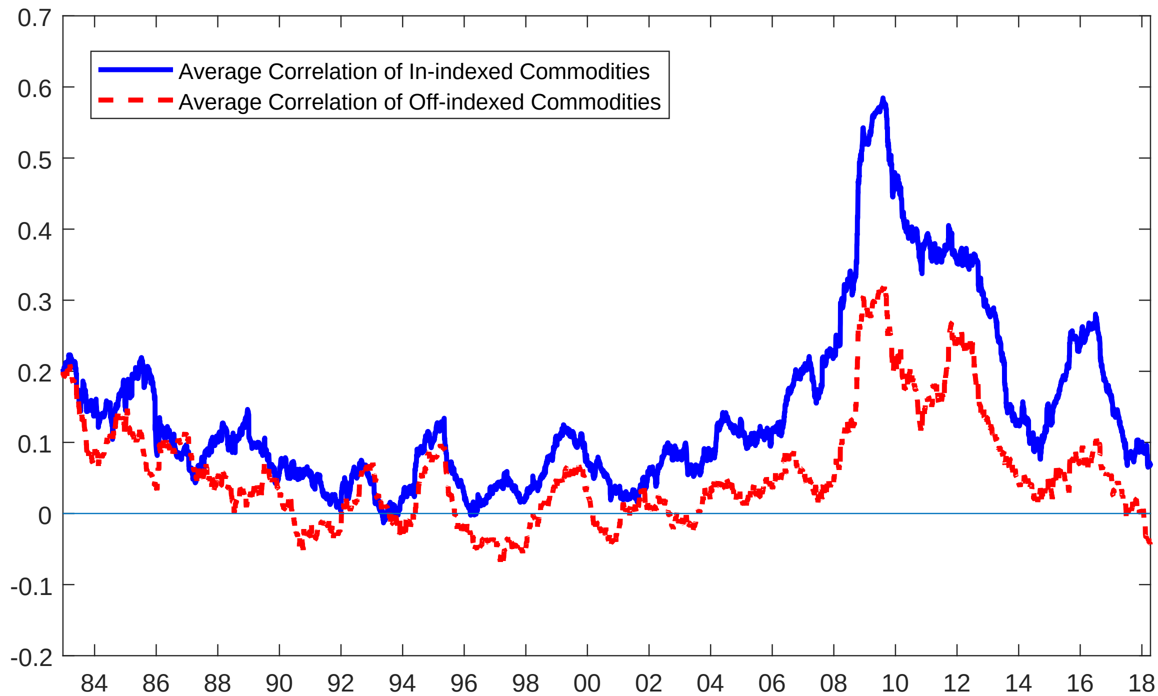

Figure 1: This figure plots the average return correlations of commodities in the GSCI and BCOM indices (indexed commodities) and those not included in these indices (non-indexed commodities). We follow Tang and Xiong (2012) in spirit to compute these correlations. Specifically, we first calculate an equal-weighted index for each sector of indexed and non-indexed commodities, then the average correlation among five sector indices for an annual rolling window. Since there are no non-indexed commodities in energy and live cattle sectors, we take heating oil and RBOB and lean hogs as non-indexed commodities due to their small weights in the index. The sample period is from 1980 to 2018.

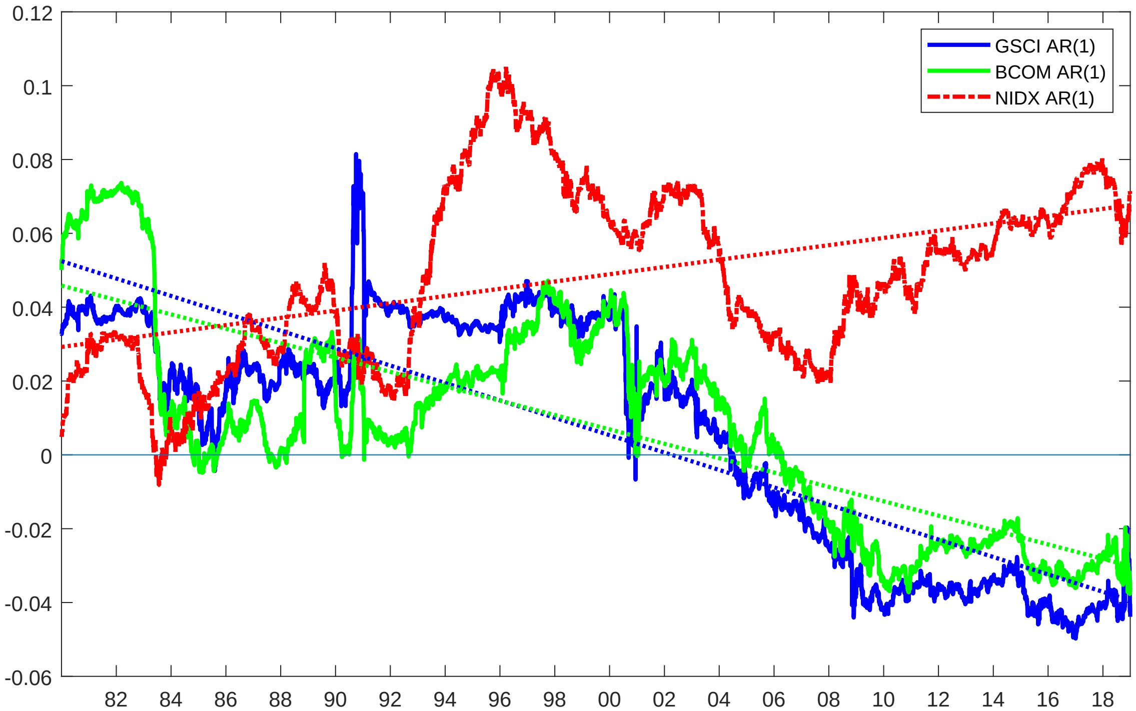

Figure 2: This figure plots the evolution of serial dependence in index returns from 1980 to 2018. Serial dependence is measured by first-order autocorrelation using a 10-year backward rolling window from index returns at the daily frequency. The indices are GSCI, BCOM and an equal-weighted portfolio of non-indexed commodities (NIDX).

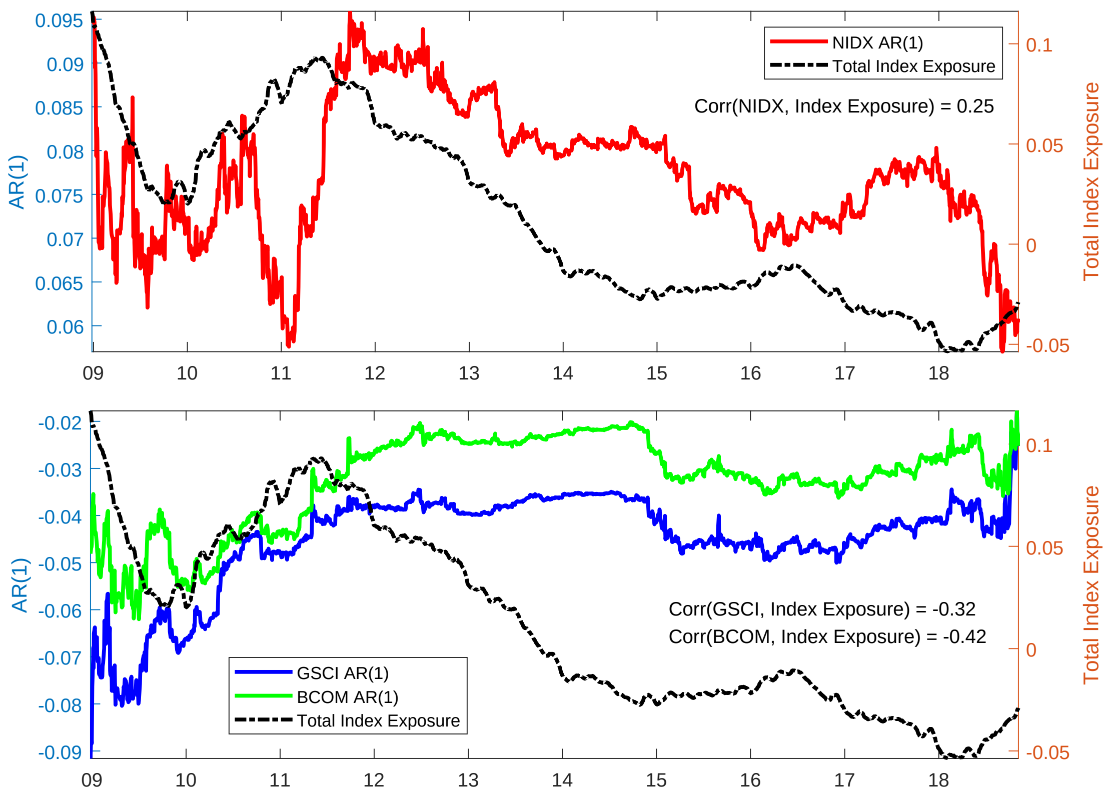

Figure 3: This figure plots the evolution of serial dependence in index returns and total index exposure from 2010 to 2018. Serial dependence is measured by the first-order autocorrelation using a 10-year backward rolling window with at least 2 years of observations from index returns at the daily frequency. The indices are GSCI, BCOM and an equal-weighted portfolio non-indexed commodities (NIDX). The total index exposure is calculated by averaging the individual index exposure in (6). The total index exposure series plotted in the figure is smoothed by taking a moving average using the same 10-year backward rolling window with at least 2 years of observations.

Table 1: Detailed List of Commodities for Analysis

This table provide a detailed list of the commodities studied in this paper and their basic information. The futures contracts of these commodities are all traded in the United States. The commodities that are included in both indices are classified as “Indexed” commodities, while commodities not included in any indices are classified as “Non-indexed” commodities.

Ticker	Name	Full Name	Exchange	Inception	GSCI	BCOM	CIT	Indexed	Non-indexed
Panel A: Energy									
CL	Crude Oil	Crude Oil, WTI / Global Spot	NYMEX	1983/03/30	✓	✓		✓	
HO	Heating Oil	ULSD NY Harbor	NYMEX	1978/11/14	✓	✓		✓	
NG	Natural Gas	Natural Gas, Henry Hub	NYMEX	1990/04/04	✓	✓		✓	
RB	Gasoline	Gasoline, Blendstock	NYMEX	2005/10/03	✓	✓		✓	
Panel B: Grains									
BO	Soybean Oil	Soybean Oil / Crude	CBOT	1959/07/01		✓	✓		
C-	Corn	Corn / No. 2 Yellow	CBOT	1959/07/01	✓	✓	✓	✓	
KW*	KC Wheat	Wheat / No. 2 Hard Winter	CBOT	1970/01/05	✓	*	✓		
MW	Minn Wheat	Wheat / Spring 14% Protein	MGEX	1979/01/02					✓
O-	Oat	Oats / No. 2 White Heavy	CBOT	1959/07/01					✓
RR	Rough Rice	Rough Rice #2	CBOT	1986/08/20					✓
S-	Soybean	Soybeans / No. 1 Yellow	CBOT	1959/07/01	✓	✓	✓	✓	
SM*	Soybean Meal	Soybean Meal / 48% Protein	CBOT	1959/01/07		*	*		✓
W-	Wheat	Wheat / No. 2 Soft Red	CBOT	1959/07/01	✓	✓	✓	✓	
Panel C: Livestock									
FC	Feeder Cattle	Cattle, Feeder / Average	CME	1971/11/30	✓		✓		
LC	Live Cattle	Cattle, Live / Choice Average	CME	1964/11/30	✓	✓	✓	✓	
LH	Lean Hogs	Hogs, Lean / Average Iowa/S Minn	CME	1966/02/28	✓	✓	✓	✓	
Panel D: Metals									
GC	Gold	Gold	NYMEX	1974/12/31	✓	✓		✓	
HG	Copper	Copper High Grade / Scrap No. 2 Wire	NYMEX	1959/01/07	✓	✓		✓	
PA	Palladium	Palladium	NYMEX	1977/01/03					✓
PL	Platinum	Platinum	NYMEX	1968/03/04					✓
SI	Silver	Silver 5,000 Troy Oz.	NYMEX	1963/06/12	✓	✓		✓	
Panel E: Softs									
CC	Cocoa	Cocoa / Ivory Coast	ICE	1959/07/01	✓		✓		
CT	Cotton	Cotton / 1-1/16”	ICE	1959/07/01	✓	✓	✓	✓	
JO	Orange Juice	Orange Juice, Frozen Concentrate	ICE	1967/02/01					✓
KC	Coffee	Coffee 'C' /Colombian	ICE	1972/08/16	✓	✓	✓	✓	
LB	Lumber	Lumber / Spruce-Pine Fir 2x4	CME	1969/10/01					✓
SB	Sugar	Sugar #11/World Raw	ICE	1961/01/04	✓	✓	✓	✓	

*Note: KW and SM are both included in BCOM from 2013. Since SM is included in BCOM from 2013, its position on index trading is reported in CIT report since 2013.

Table 2: Descriptive Statistics of Commodities' Returns

This table provides some descriptive statistics of each commodity/index' daily returns (after winsorization) in columns 2-7. In column 8, we calculate the annualized Sharpe ratio (scaled by $\sqrt{252}$) of each commodity. NIDX denotes the equal-weighted portfolio of non-indexed commodities. The sample is of daily frequency ranging from January 2, 2003 to November 6, 2018

Commodity	Observations	Mean	StDev.	Min	Max	AR(1)	Sharpe Ratio
Panel A: Energy							
CL	3,979	0.003%	0.0207	-0.0577	0.0608	-0.0602	0.0195
HO	3,979	0.024%	0.0193	-0.0497	0.0562	-0.0430	0.1980
NG	3,979	-0.106%	0.0270	-0.0725	0.0786	-0.0542	-0.6204
RB	3,979	0.033%	0.0209	-0.0603	0.0568	-0.0327	0.2506
Panel B: Grains							
BO	3,991	0.002%	0.0142	-0.0385	0.0414	0.0032	0.0237
C-	3,991	-0.002%	0.0169	-0.0481	0.0490	0.0277	-0.0141
KW	3,991	-0.002%	0.0178	-0.0479	0.0528	0.0247	-0.0155
MW	3,991	0.033%	0.0160	-0.0430	0.0494	0.0507	0.3305
O-	3,991	0.042%	0.0197	-0.0526	0.0570	0.0766	0.3355
RR	3,991	0.002%	0.0146	-0.0366	0.0395	0.0748	0.0270
S-	3,991	0.044%	0.0148	-0.0450	0.0410	0.0029	0.4699
SM	3,991	0.072%	0.0167	-0.0479	0.0471	0.0353	0.6865
W-	3,991	-0.019%	0.0193	-0.0499	0.0549	0.0054	-0.1575
Panel C: Livestock							
FC	3,981	0.023%	0.0095	-0.0271	0.0255	0.0797	0.3808
LC	3,981	0.018%	0.0097	-0.0252	0.0253	0.0425	0.2962
LH	3,991	-0.004%	0.0142	-0.0393	0.0369	0.0363	-0.0499
Panel D: Metals							
GC	3,979	0.030%	0.0107	-0.0316	0.0286	-0.0213	0.4443
HG	3,979	0.051%	0.0169	-0.0521	0.0499	-0.0618	0.4744
PA	3,979	0.056%	0.0191	-0.0599	0.0523	0.0589	0.4657
PL	3,979	0.021%	0.0133	-0.0396	0.0354	0.0143	0.2528
SI	3,979	0.047%	0.0188	-0.0610	0.0518	-0.0296	0.3953
Panel E: Softs							
CC	3,971	0.026%	0.0178	-0.0514	0.0484	0.0036	0.2362
CT	3,953	0.003%	0.0170	-0.0478	0.0448	0.0680	0.0257
JO	3,971	0.030%	0.0192	-0.0550	0.0538	0.0853	0.2458
KC	3,971	-0.010%	0.0188	-0.0494	0.0511	-0.0165	-0.0858
LB	3,991	-0.024%	0.0178	-0.0413	0.0451	0.0961	-0.2115
SB	3,971	0.006%	0.0190	-0.0528	0.0503	-0.0023	0.0473
Panel F: Commodity Indices							
GSCI	3,992	-0.003%	0.0146	-0.0829	0.0748	-0.0429	-0.0307
BCOM	3,986	-0.001%	0.0105	-0.0620	0.0581	-0.0309	-0.0213
NIDX	3,992	0.031%	0.0091	-0.0472	0.0441	0.0598	0.5403

Table 3: Descriptive Statistics of Commodities' News Sentiment

This table provides descriptive statistics of each commodity's news sentiment. The news sentiment of each commodity is calculated from the news tones data provided in Thomson Reuters News Analytics. The news sentiment is the residuals from regressing the negative news tone on its first lag and the weekday dummies. The whole sample is of daily frequency ranging from January 2, 2003 to November 6, 2018.

Commodity	Total # of News	Data Range	Observations	StDev.	Min	Max
Panel A: Energy						
CL	1,166,244	2003/01/02 - 2018/11/06	3,989	0.0368	-0.1277	0.1527
HO	204,800	2003/01/08 - 2018/11/06	3,745	0.0739	-0.3533	0.4333
NG	605,270	2003/01/02 - 2018/11/06	3,989	0.0363	-0.2148	0.1372
RB	216,209	2005/12/14 - 2018/11/06	3,245	0.0562	-0.1757	0.2317
Panel B: Grains						
BO	635,167	2003/01/02 - 2018/11/06	3,989	0.0355	-0.1435	0.4230
C-	99,303	2003/01/07 - 2018/11/06	3,532	0.0940	-0.3075	0.5250
KW	92,920	2008/12/05 - 2018/11/06	2,024	0.0726	-0.3121	0.4252
MW	92,920	2008/12/05 - 2018/11/06	2,024	0.0726	-0.3121	0.4252
O-	912,460	2003/01/02 - 2018/11/06	3,989	0.0306	-0.0954	0.1337
RR	912,460	2003/01/02 - 2018/11/06	3,989	0.0306	-0.0954	0.1337
S-	83,866	2003/02/06 - 2018/11/06	2,760	0.1038	-0.3226	0.4644
SM	608,977	2003/01/02 - 2018/11/06	3,987	0.0394	-0.1265	0.1556
W-	92,920	2008/12/05 - 2018/11/06	2,024	0.0726	-0.3121	0.4252
Panel C: Livestocks						
FC	506,675	2003/01/02 - 2018/11/06	3,989	0.0446	-0.1712	0.2854
LC	506,675	2003/01/02 - 2018/11/06	3,989	0.0446	-0.1712	0.2854
LH	506,675	2003/01/02 - 2018/11/06	3,989	0.0446	-0.1712	0.2854
Panel D: Metals						
GC	326,315	2003/01/02 - 2018/11/06	3,989	0.0583	-0.1827	0.2257
HG	59,571	2009/12/18 - 2018/11/06	2,024	0.0930	-0.3198	0.3803
PA	326,315	2003/01/02 - 2018/11/06	3,989	0.0583	-0.1827	0.2257
PL	326,315	2003/01/02 - 2018/11/06	3,989	0.0583	-0.1827	0.2257
SI	326,315	2003/01/02 - 2018/11/06	3,989	0.0583	-0.1827	0.2257
Panel E: Softs						
CC	98,039	2003/01/02 - 2018/11/06	3,985	0.0872	-0.3104	0.3883
CT	119,551	2003/01/02 - 2018/11/06	3,987	0.0708	-0.2558	0.3567
JO	41,362	2003/01/02 - 2018/11/06	3,278	0.1031	-0.4198	0.5498
KC	118,847	2003/01/02 - 2018/11/06	3,987	0.0808	-0.2975	0.3980
LB	325,463	2003/01/02 - 2018/11/06	3,989	0.0434	-0.1702	0.1692
SB	159,840	2003/01/02 - 2018/11/06	3,987	0.0680	-0.2753	0.3273

*Note: As Thomson Reuters only provides some news tones up to sector level, we have to use sector news tones for some commodities. Specifically, (1) GC, SI, PA, and PL use scores for "Gold and Precious Metals"; (2) W-, MW and KW use scores for "Wheat"; (3) FC, LC, and LH use scores for "Livestocks"; (4). O- and RR use scores for "Grains".

Table 4: Spillover Effect of Sentiment on Returns across Indexed/Non-indexed Commodities

This table presents the results of regressing commodities returns on the "connected" sentiment. We first get each commodity's news sentiment as the residuals from regressing the *negative* news tones on its first lag and the weekday dummies. We then obtain the "connected" sentiment for an indexed commodity by taking a value-weighted average of indexed commodities from other sectors. For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Panel A: Contemporaneous			Panel B: Predictive		
	Indexed	Non-indexed	All	Indexed	Non-indexed	All
Cnn. Sentiment	-0.1571***	-0.0931***	-0.0928***			
	(-42.45)	(-18.62)	(-18.71)			
Cnn. Sentiment×Indexed			-0.0604***			
			(-9.84)			
L.Cnn. Sentiment				0.0104***	-0.0053	-0.0051
				(2.99)	(-1.10)	(-1.07)
L.(Cnn. Sentiment×Indexed)						0.0152***
						(2.57)
L.Basis	0.0051	0.0216**	0.0093**	0.0052	0.0198**	0.0090**
	(1.08)	(2.30)	(2.22)	(1.09)	(2.08)	(2.14)
L.Illiquidity	4.81e-05	9.75e-05	4.85e-05	4.46e-05	8.57e-05	6.20e-05
	(0.50)	(0.85)	(0.66)	(0.46)	(0.74)	(0.83)
L.Rvol	-0.0014	-4.86e-05	-0.0010	-0.0022**	-0.0005	-0.0015*
	(-1.42)	(-0.03)	(-1.24)	(-2.25)	(-0.35)	(-1.89)
Intercept	-0.0002	0.0012*	0.0002	0.0014***	0.0001	0.0003
	(-0.63)	(1.74)	(0.64)	(3.94)	(0.23)	(1.04)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	51,784	27,526	79,310	51,770	27,521	79,291
# of Individuals	15	8	23	15	8	23
Overall R-squared	4.16%	1.60%	3.25%	0.18%	0.20%	0.17%

Table 5: Spillover Effect of Sentiment on Returns across Indexed/Non-indexed Commodities under High/Low Total Index Exposure Period

This table presents the results of regressing commodities returns on connected sentiment measures under different levels of total index exposure. The total index exposure is the average of the indexed commodities' individual index exposure. The index trading share is defined as the ratio of indexed open interest to the total open interest for a certain commodity. The index trading volume for a certain commodity is the production of the market trading volume and its corresponding index trading share. The index exposure is thus obtained by detrending the index trading volume with its past 250-day average and then standardizing the time series. We characterize the period when total index exposure is above(below) zero as "High"("Low") exposure period. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The *t*-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Panel A: Contemporaneous				Panel B: Predictive			
Variables	Indexed		Non-indexed		Indexed		Non-indexed	
	High	Low	High	Low	High	Low	High	Low
Cnn. Sentiment	-0.1796***	-0.1298***	-0.1069***	-0.0758***				
	(-35.58)	(-24.73)	(-15.27)	(-10.63)				
L.Cnn. Sentiment					0.0166***	0.0023	0.0057	-0.0167**
					(3.55)	(0.46)	(0.85)	(-2.37)
L.Basis	0.0109	-0.0016	0.0123	0.0285**	0.0009	0.0113*	0.0104	0.0261**
	(1.60)	(-0.23)	(0.88)	(2.25)	(0.15)	(1.68)	(0.79)	(2.00)
L.Illiquidity	-7.60e-05	0.0002	0.0002	-1.46e-06	-6.93e-05	0.0002	-8.70e-06	0.0002
	(-0.60)	(1.63)	(1.10)	(-0.01)	(-0.56)	(1.26)	(-0.05)	(1.12)
L.Rvol	-0.0011	-0.0014	-0.0019	0.0027	-0.0032**	-0.0008	-0.0010	0.0010
	(-0.86)	(-0.97)	(-0.96)	(1.28)	(-2.36)	(-0.55)	(-0.50)	(0.46)
Intercept	0.0015*	0.0016***	0.0003	0.0008	-0.0001	0.0004	0.0003	-0.0001
	(1.90)	(2.67)	(0.23)	(0.87)	(-0.08)	(0.53)	(0.28)	(-0.13)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	28,126	23,658	15,008	12,518	28,122	23,648	14,993	12,528
# of Individuals	15	15	8	8	15	15	8	8
Overall R-squared	5.04%	3.37%	2.12%	1.39%	0.28%	0.43%	0.16%	0.52%

Table 6: Return Serial Dependence and Commodities' Total Index Exposure

This table presents the results of regressing commodities serial dependence measure on commodities' total index exposure. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The total index exposure is the average of the indexed commodities' individual index exposure. The index trading share is defined as the ratio of indexed open interest to the total open interest for a certain commodity. The index trading volume for a certain commodity is the production of the market trading volume and its corresponding index trading share. The index exposure is thus obtained by detrending the index trading volume with its past 250-day average and then standardizing the time series. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The control variables are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Full Sample		
	Indexed	Non-indexed	All
L.Total Index Exposure	-0.0845***	-0.0011	-0.0165
	(-3.53)	(-0.05)	(-0.83)
L.(Total Index Exposure×Indexed)			-0.0589**
			(-2.22)
L.Basis	0.7373*	3.2462***	1.3021***
	(1.69)	(3.76)	(3.25)
L.Illiquidity	-0.0042	0.0136	0.0044
	(-0.35)	(1.15)	(0.51)
L.Oil Implied Volatility	-0.4670***	-0.3538***	-0.4305***
	(-4.98)	(-3.21)	(-5.94)
Intercept	0.2342***	0.1243***	0.1742***
	(3.80)	(3.08)	(5.71)
Individual Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
# of Observations	44,587	23,518	68,105
# of Individuals	15	8	23
Overall R-squared	0.33%	0.50%	0.38%

Table 7: Contrarian (Momentum) Trading Strategy Based on Short-term Return Reversal (Continuation) of GSCI/BCOM (NIDX)

This table presents the descriptive statistics of implementing a time-series contrarian (momentum) strategy based on short-term return reversals (continuation) of commodity indices (non-indexed portfolios). For contrarian (momentum) strategy, we sell (buy) the GSCI/BCOM (NIDX) when the past daily return is positive and buy (sell) the GSCI/BCOM (NIDX) when the past daily return is negative. The daily trading position of each index is $|r^p_{t-1}|$, $p \in \{\text{GSCI, BCOM, NIDX}\}$, respectively. The portfolio is rebalanced on a daily basis. To account for the trading cost, we use the weighted average of one tick bid-ask spreads for indexed commodities (1.04 bps for GSCI and 1.26 bps for BCOM) and the weighted average of two ticks bid-ask spreads for non-indexed commodities (7.74 bps for NIDX). The high index exposure refers to the period when total index exposure is above zero. The real-time index exposure is calculated using a window of past 250 days in the standardization instead of a window of full sample. The averaged daily returns and the standard deviations are reported in basis points.

Panel A: Reverse Portfolio (GSCI)	Full Sample	High Index Exposure (Real-time)
Mean Return (bf. Cost)	0.0927	0.1050
Standard Deviation (bf. Cost)	2.9767	2.3178
Ann. Sharpe Ratio (bf. Cost)	0.4941	0.7188
Mean Return (aft. Cost)	0.0848	0.1003
Standard Deviation (aft. Cost)	2.9762	2.3172
Ann. Sharpe Ratio (aft. Cost)	0.4523	0.6872
Panel B: Reverse Portfolio (BCOM)	**Full Sample**	**High Index Exposure (Real-time)**
Mean Return (bf. Cost)	0.0352	0.0520
Standard Deviation (bf. Cost)	1.4726	1.1868
Ann. Sharpe Ratio (bf. Cost)	0.3799	0.6951
Mean Return (aft. Cost)	0.0283	0.0478
Standard Deviation (aft. Cost)	1.4724	1.1864
Ann. Sharpe Ratio (aft. Cost)	0.3053	0.6400
Panel C: Momentum Portfolio (NIDX)	**Full Sample**	**High Index Exposure (Real-time)**
Mean Return (bf. Cost)	0.0700	0.0454
Standard Deviation (bf. Cost)	1.0214	0.7832
Ann. Sharpe Ratio (bf. Cost)	1.0884	0.9196
Mean Return (aft. Cost)	0.0332	0.0236
Standard Deviation (aft. Cost)	1.0210	0.7822
Ann. Sharpe Ratio (aft. Cost)	0.5161	0.4788

Table 8: Return Serial Dependence and Commodities' Individual Index Exposure

This table presents the results of regressing commodities serial dependence measure on commodities' individual index exposure. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Baseline		Liquidity Provision	
	Indexed	Non-indexed	Indexed	Non-indexed
L.Index Exposure	-0.0241***	-0.0011	-0.0311***	-0.0028
	(-3.16)	(-0.05)	(-3.31)	(-0.13)
L.(Index Exposure×Illiquidity)			-0.0347**	-0.0372
			(-2.10)	(-1.32)
L.Basis	0.8029*	3.2462***	0.7617*	3.2396***
	(1.80)	(3.76)	(1.70)	(3.76)
L.Illiquidity	-0.0057	0.0136	-0.0227	0.0126
	(-0.47)	(1.15)	(-1.42)	(1.06)
L.Oil Implied Volatility	-0.4414***	-0.3538***	-0.4356***	-0.3599***
	(-4.74)	(-3.21)	(-4.69)	(-3.27)
Intercept	0.2081***	0.1243***	0.1995***	0.1243***
	(6.52)	(3.08)	(3.62)	(3.08)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Observations	44,587	23,518	44,587	23,518
# of Individuals	15	8	15	8
Overall R-squared	0.29%	0.50%	0.31%	0.52%

Table 9: Return Serial Dependence and Components of Commodities' Individual Index Exposure

This table presents the results of regressing commodities serial dependence measure on the components of commodities' individual index exposure, i.e., index market share and trading volume. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The index market share is defined as the index open interest divided by the total open interest. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Full Sample					
	Indexed			Non-indexed		
L.Index Market Share	-0.3937**		-0.4118**	-0.2619		-0.2430
	(-2.25)		(-2.34)	(-0.23)		(-0.21)
L.Trading Volume		-0.0003***	-0.0003***		-0.0008	0.0006
		(-2.33)	(-3.11)		(-0.89)	(0.42)
L.Basis	0.5899	0.4365	0.5633	2.8382***	2.7185***	2.9021***
	(1.45)	(1.02)	(1.38)	(3.01)	(3.61)	(3.01)
L.Illiquidity	0.0119	0.0123	0.0091	0.0202	0.0295***	0.0206*
	(0.81)	(0.73)	(0.61)	(1.64)	(2.67)	(1.69)
L.Oil Implied Volatility	-0.4613***	-0.5250***	-0.4594***	-0.4189***	-0.4119***	-0.4188***
	(-4.94)	(-5.02)	(-4.92)	(-3.48)	(-3.23)	(-3.47)
Intercept	0.2741***	0.4779***	0.3045***	0.0790*	0.2217*	0.0361
	(5.20)	(4.18)	(2.77)	(1.87)	(1.80)	(0.16)
Individual Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	48,285	58,658	48,285	23,739	31,405	23,739
# of Individuals	15	15	15	15	15	15
Overall R-squared	0.18%	0.15%	0.19%	0.22%	0.18%	0.23%

Table 10: Serial Dependence and Indexing in Overweighted GSCI/BCOM Commodities

This table presents the results based on top 3 GSCI/BCOM overweighted commodities after regressing return serial dependence measure against index exposure on GSCI and BCOM. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. GSCI(BCOM) index exposure is the standardized version of detrended GSCI(BCOM) index trading volume with the past 250-day average. GSCI(BCOM) index trading volume is estimated by multiplying its total trading volume by the ratio of GSCI(BCOM) index open interest to its total open interest. Each commodity's index open interest on GSCI and BCOM can be obtained by Hamilton and Wu (2015) method in Appendix A.1. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Overweighted in GSCI (CL, HO, RB)		Overweighted in BCOM (GC, NG, S-)	
	Full Sample	Exclude Roll Weeks	Full Sample	Exclude Roll Weeks
L.Index Exposure on GSCI	-0.0420**	-0.0587**	-0.0004	-0.0198
	(-2.18)	(-2.44)	(-0.03)	(-1.02)
L.Index Exposure on BCOM	-0.0199	-0.0175	-0.0389***	-0.0422***
	(-1.26)	(-0.88)	(-2.65)	(-2.64)
L.Basis	-0.2891	0.1748	-0.0722	-0.3452
	(-0.40)	(0.20)	(-0.07)	(-0.28)
L.Illiquidity	-0.1449	-0.1300	-0.0153	0.0227
	(-1.63)	(-1.25)	(-0.47)	(0.40)
L.Oil Implied Volatility	-0.6373**	-0.8283**	-0.1916	-0.3173
	(-2.17)	(-2.17)	(-1.00)	(-1.53)
Intercept	0.2945**	0.4274***	0.0514	-0.0017
	(2.52)	(3.24)	(0.69)	(-0.02)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Observations	8,955	6,864	8,955	7,501
# of Individuals	3	3	3	3
Overall R-squared	1.11%	1.56%	0.34%	0.55%

Appendices

A.1 Details on Estimating the Positions of Non-reported Indexed Commodities in CIT Report

Masters (2008) and Hamilton and Wu (2015) proposed to estimate the unreported index trading positions by making use of the reported data and their weights in each commodity index. Taking crude oil (CL) as an example, the general idea of Masters (2008) is to use the fact that both GSCI and BCOM have their own uniquely included commodities, i.e. soybean oil (BO) and soybean meal (SM) in BCOM[22] and cocoa (CC), feeder cattle (FC) and Kansas wheat[23] (KW) for GSCI. Then, note that index traders replicate the index by allocating capital across commodities according to their known weights[24] $\delta^{(i)}_{jy(t)}, i \in \{G, B\}$, we can separately estimate CL's dollar value long/short positions on index trading, $X_{CL,t}$, on GSCI/BCOM trading as below:

$$\widehat{X}^{B}_{CL,t} = \begin{cases} \dfrac{\delta^{B}_{CL,y(t)}}{\delta^{B}_{BO,y(t)}} X_{BO,t}, & \text{if } y(t) < 2013, \\ \dfrac{1}{2}\left(\dfrac{\delta^{B}_{CL,y(t)}}{\delta^{B}_{BO,y(t)}} X_{BO,t} + \dfrac{\delta^{B}_{CL,y(t)}}{\delta^{B}_{SM,y(t)}} X_{SM,t}\right), & \text{if } y(t) \geq 2013. \end{cases} \tag{20}$$

$$\widehat{X}^{G}_{CL,t} = \begin{cases} \dfrac{1}{3}\left(\dfrac{\delta^{G}_{CL,y(t)}}{\delta^{G}_{CC,y(t)}} X_{CC,t} + \dfrac{\delta^{G}_{CL,y(t)}}{\delta^{G}_{FC,y(t)}} X_{FC,t} + \dfrac{\delta^{G}_{CL,y(t)}}{\delta^{G}_{KW,y(t)}} X_{KW,t}\right), & \text{if } y(t) < 2013, \\ \dfrac{1}{2}\left(\dfrac{\delta^{G}_{CL,y(t)}}{\delta^{G}_{CC,y(t)}} X_{CC,t} + \dfrac{\delta^{G}_{CL,y(t)}}{\delta^{G}_{FC,y(t)}} X_{FC,t}\right), & \text{if } y(t) \geq 2013. \end{cases} \tag{21}$$

where $y(t)$ denotes the year of day t. Note that the weights of commodities in an index are determined at the beginning of a year and stay the same during the year. Thus, the dollar-valued position of index trading for commodity i on day t is estimated as

$$X_{it} = Position_{iw(t)} \times ContractSize_i \times Price_{it}. \tag{22}$$

[22] Soybean meal (SM) is included in BCOM since 2013.
[23] Kansas wheat (KW) is included in BCOM since 2013.
[24] Both weights reported in the GSCI and BCOM manuals are dollar value weights.

Combining the estimates above, Masters (2008) propose to estimate the dollar-valued position of CL on index trading as:

$$\widehat{X}^{Idx}_{CL,t} = \widehat{X}^{B}_{CL,t} + \widehat{X}^{G}_{CL,t}. \tag{23}$$

However, as pointed out by Irwin and Sanders (2012), Masters' estimator is severely biased when there is a huge difference between $\frac{\delta^G_{CL,y(t)}}{\delta^G_{CC,y(t)}} X_{CC,t}$, $\frac{\delta^G_{CL,y(t)}}{\delta^G_{FC,y(t)}} X_{FC,t}$ and $\frac{\delta^G_{CL,y(t)}}{\delta^G_{KW,y(t)}} X_{KW,t}$. To deal with this issue, Hamilton and Wu (2015) propose to generalize Masters' method by using all the reported commodities' positions for estimation. Specifically, they choose $\widehat{X}^G_{it}$ and $\widehat{X}^B_{it}$ to minimize the sum of squared discrepancies in predicting the CIT reported value for X_{it} across 12 commodities. Thus, the estimated dollar value positions on index trading for commodity i in day t is given by

$$\widehat{X}^{Idx}_{it} = \begin{bmatrix} \delta^G_{iy(t)} & \delta^B_{iy(t)} \end{bmatrix} \begin{bmatrix} \sum\limits_{j\in CIT} \left(\delta^G_{jy(t)}\right)^2 & \sum\limits_{j\in CIT} \delta^G_{jy(t)}\delta^B_{jy(t)} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)}\delta^G_{jy(t)} & \sum\limits_{j\in CIT} \left(\delta^B_{jy(t)}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum\limits_{j\in CIT} \delta^G_{jy(t)} X^{Idx}_{jt} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)} X^{Idx}_{jt} \end{bmatrix}, \tag{24}$$

where $\delta_{jy(t)}$ is the weight of a commodity j in a certain index in year $y(t)$, and the superscripts G and B denote the index GSCI and BCOM, respectively. From Equation (24) we obtain both the long and short dollar-valued long/short index positions for unreported commodities, and thus the index open interest.

In addition, we can easily modify Hamilton and Wu (2015) method to estimate the non-reported indexed commodities' dollar-valued GSCI/BCOM trading position as below:

$$\widehat{X}^{G}_{it} = \begin{bmatrix} \delta^G_{iy(t)} & 0 \end{bmatrix} \begin{bmatrix} \sum\limits_{j\in CIT} \left(\delta^G_{jy(t)}\right)^2 & \sum\limits_{j\in CIT} \delta^G_{jy(t)}\delta^B_{jy(t)} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)}\delta^G_{jy(t)} & \sum\limits_{j\in CIT} \left(\delta^B_{jy(t)}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum\limits_{j\in CIT} \delta^G_{jy(t)} X^{Idx}_{jt} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)} X^{Idx}_{jt} \end{bmatrix}, \tag{25}$$

$$\widehat{X}^{B}_{it} = \begin{bmatrix} 0 & \delta^B_{iy(t)} \end{bmatrix} \begin{bmatrix} \sum\limits_{j\in CIT} \left(\delta^G_{jy(t)}\right)^2 & \sum\limits_{j\in CIT} \delta^G_{jy(t)}\delta^B_{jy(t)} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)}\delta^G_{jy(t)} & \sum\limits_{j\in CIT} \left(\delta^B_{jy(t)}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum\limits_{j\in CIT} \delta^G_{jy(t)} X^{Idx}_{jt} \\ \sum\limits_{j\in CIT} \delta^B_{jy(t)} X^{Idx}_{jt} \end{bmatrix}. \tag{26}$$

A.2 Additional Results on Sentiment Propagation and Return Serial Dependence Excluding Financial Crisis Period

Table A1: Spillover Effect of Sentiment on Returns across Indexed/Non-indexed Commodities excluding Financial Crisis Period

This table presents the subperiod results of regressing commodities returns on connected sentiment measures. According to Tang and Xiong (2012), the sample excludes the period September 15, 2008 to June 30, 2009. The connected sentiment measure is constructed in two steps. In the first step, we obtain each commodity's news sentiment as the residuals from regressing its negative news tones on its first lag and weekday dummies. We then obtain the "connected" sentiment for an indexed commodity by taking a value-weighted average of indexed commodities from other sectors. For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Exclude Financial Crisis (2008/09/15 – 2009/06/30)					
Variables	Panel A: Contemporaneous			Panel B: Predictive		
	Indexed	Non-indexed	All	Indexed	Non-indexed	All
Cnn. Sentiment	-0.1401***	-0.0764***	-0.0770***			
	(-38.57)	(-15.42)	(-15.66)			
Cnn. Sentiment×Indexed			-0.0592***			
			(-9.78)			
L.Cnn. Sentiment				0.0080**	-0.0048	-0.0049
				(2.30)	(-0.98)	(-1.02)
L.(Cnn. Sentiment×Indexed)						0.0127**
						(2.15)
L.Basis	0.0063	0.0159	0.0089**	0.0061	0.0136	0.0082*
	(1.27)	(1.63)	(2.02)	(1.22)	(1.38)	(1.84)
L.Illiquidity	-1.90e-05	2.76e-05	-1.89e-05	-3.36e-05	1.43e-05	-1.33e-05
	(-0.20)	(0.23)	(-0.25)	(-0.35)	(0.12)	(-0.18)
L.Rvol	-0.0005	0.0002	-0.0003	-0.0005	6.62e-05	-0.0002
	(-0.49)	(0.14)	(-0.33)	(-0.47)	(0.04)	(-0.24)
Intercept	-0.0002	0.0003	5.08e-05	0.0014**	0.0013*	0.0014***
	(-0.44)	(0.35)	(-0.11)	(2.31)	(1.65)	(2.94)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	48,789	25,952	74,741	48,776	25,947	74,723
# of Individuals	15	8	23	15	8	23
Overall R-squared	3.48%	1.11%	2.63%	0.13%	0.15%	0.13%

Table A2: Return Serial Dependence and Commodities' Total Index Exposure Excluding Financial Crisis Period

This table presents the results of regressing commodities serial dependence measure on commodities' total index exposure in period excluding financial crisis. According to Tang and Xiong (2012), the sample excludes the period September 15, 2008 to June 30, 2009. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The control variables are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Exclude Financial Crisis (2008/09/15 – 2009/06/30)		
	Indexed	Non-indexed	All
L.Total Index Exposure	-0.0534**	0.0049	-0.0003
	(-2.17)	(0.24)	(-0.02)
L.(Total Index Exposure×Indexed)			-0.0501*
			(-1.88)
L.Basis	0.9498**	3.4967***	1.5309***
	(2.06)	(3.85)	(3.61)
L.Illiquidity	-0.0017	0.0125	0.0058
	(-0.15)	(1.03)	(0.51)
L.Oil Implied Volatility	-0.5128***	-0.1654	-0.3843***
	(-4.84)	(-1.28)	(-4.64)
Intercept	0.2854***	0.3049***	0.3636***
	(3.70)	(3.95)	(7.49)
Individual Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
# of Observations	41,593	21,944	63,537
# of Individuals	15	8	23
Overall R-squared	0.30%	0.61%	0.39%

Table A3: Return Serial Dependence and Commodities' Individual Index Exposure Excluding Financial Crisis Period

This table presents the results of regressing commodities serial dependence measure on commodities' individual index exposure in period excluding financial crisis. According to Tang and Xiong (2012), the sample excludes the period September 15, 2008 to June 30, 2009. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Baseline		Liquidity Provision	
	Indexed	Non-indexed	Indexed	Non-indexed
L.Index Exposure	-0.0166**	0.0049	-0.0220**	0.0031
	(-2.18)	(0.24)	(-2.32)	(0.31)
L.(Index Exposure×Illiquidity)			-0.0240*	-0.0287
			(-1.69)	(-1.12)
L.Basis	0.9931**	3.4967***	0.9665**	3.4819***
	(2.10)	(3.85)	(2.04)	(3.83)
L.Illiquidity	-0.0032	0.0125	-0.0147	0.0123
	(-0.28)	(1.03)	(-1.02)	(1.01)
L.Oil Implied Volatility	-0.5413***	-0.1654	-0.5329***	-0.1667
	(-4.92)	(-1.28)	(-4.85)	(-1.29)
Intercept	0.3008***	0.3049***	0.1986***	0.3040***
	(3.59)	(3.95)	(3.34)	(3.94)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Observations	41,593	21,944	41,593	21,944
# of Individuals	15	8	15	8
Overall R-squared	0.29%	0.61%	0.30%	0.62%

A.3 Additional Results on Sentiment Propagation with Net News Tone

Table A4: Spillover Effect of Sentiment (Net News Tone) on Returns across Indexed/Non-indexed Commodities

This table presents the results of regressing commodities returns on the "connected" sentiment measure. The "connected" sentiment measure is constructed in two steps. We first obtain each commodity's net news sentiment as the residuals from regressing the net news tones on its first lag and the weekday dummies. We then obtain the "connected" sentiment for an indexed commodity by taking a value-weighted average of indexed commodities from other sectors. For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Panel A: Contemporaneous			Panel B: Predictive		
	Indexed	Non-indexed	All	Indexed	Non-indexed	All
Cnn. Sentiment	0.0860***	0.0519***	0.0517***			
	(43.65)	(19.31)	(19.64)			
Cnn. Sentiment×Indexed			0.0321***			
			(9.92)			
L.Cnn. Sentiment				-0.0052***	0.0028	0.0024
				(-2.84)	(1.09)	(0.91)
L.(Cnn. Sentiment×Indexed)						-0.0073**
						(-2.33)
L.Basis	0.0059	0.0217*	0.0099**	0.0052	0.0198**	0.0090**
	(1.24)	(2.31)	(2.35)	(1.08)	(2.08)	(2.13)
L.Illiquidity	5.48e-05	8.81e-05	4.72e-05	4.37e-05	8.53e-05	6.09e-05
	(0.57)	(0.77)	(0.64)	(0.45)	(0.74)	(0.82)
L.Rvol	-0.0012	4.33e-05	-0.0008	-0.0022**	-0.0005	-0.0015*
	(-1.23)	(0.03)	(-1.03)	(-2.24)	(-0.34)	(-1.90)
Intercept	-4.47e-05	0.0008	0.0004	0.0014***	0.0001	0.0003
	(-0.11)	(1.16)	(1.27)	(3.52)	(0.25)	(1.00)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	51,784	27,526	79,310	51,770	27,521	79,291
# of Individuals	15	8	23	15	8	23
Overall R-squared	4.43%	1.72%	3.47%	0.18%	0.20%	0.17%

Table A5: Spillover Effect of Sentiment (Net News Tone) on Returns across Indexed/Non-indexed Commodities under High/Low Total Index Exposure Period

This table presents the results of regressing commodities returns on connected sentiment measures and controls under different levels of total index exposure. The total index exposure is the average of the indexed commodities' individual index exposure. The index trading share is defined as the ratio of indexed open interest to the total open interest for a certain commodity. The index trading volume for a certain commodity is the production of the market trading volume and its corresponding index trading share. The index exposure is thus obtained by detrending the index trading volume with its past 250-day average and then standardizing the time series. We characterize the period when total index exposure is above(below) zero as "High" ("Low") exposure period. The connected sentiment measure is constructed using net news tone. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Panel A: Contemporaneous				Panel B: Predictive			
	Indexed		Non-indexed		Indexed		Non-indexed	
	High	Low	High	Low	High	Low	High	Low
Cnn. Sentiment	0.0951***	0.0748***	0.0573***	0.0447***				
	(35.58)	(26.54)	(15.51)	(11.43)				
L.Cnn. Sentiment					-0.0085***	-0.0012	-0.0020	0.0078**
					(-3.43)	(-0.45)	(-0.56)	(2.01)
L.Basis	0.0114*	-0.0009	0.0125	0.0284**	0.0009	0.0112*	0.0104	0.0260**
	(1.69)	(-0.13)	(0.90)	(2.24)	(0.14)	(1.68)	(0.79)	(1.99)
L.Illiquidity	-6.18e-05	0.0002	0.0002	-1.09e-05	-6.98e-05	0.0002	-7.63e-06	0.0002
	(-0.48)	(1.54)	(1.02)	(-0.07)	(-0.56)	(1.25)	(-0.05)	(1.11)
L.Rvol	-0.0010	-0.0012	-0.0017	0.0027	-0.0032**	-0.0008	-0.0010	0.0010
	(-0.78)	(-0.82)	(-0.87)	(1.29)	(-2.36)	(-0.55)	(-0.50)	(0.45)
Intercept	0.0013**	0.0013**	0.0005	0.0004	-7.06e-06	0.0004	0.0003	-0.0002
	(2.16)	(2.11)	(0.48)	(0.45)	(-0.01)	(0.53)	(0.27)	(-0.17)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	28,126	23,658	15,008	12,518	28,122	23,648	14,993	12,528
# of Individuals	15	15	8	8	15	15	8	8
Overall R-squared	5.08%	3.88%	2.17%	1.58%	0.28%	0.43%	0.16%	0.50%

Table A6: Spillover Effect of Sentiment (Net News Tone) on Returns across Indexed/Non-indexed Commodities Excluding Financial Crisis Period

This table presents the subperiod results of regressing commodities returns on connected sentiment measures. According to Tang and Xiong (2012), the sample excludes the period September 15, 2008 to June 30, 2009. The connected sentiment measure is constructed in two steps. We first obtain each commodity's news sentiment as the residuals from regressing the net news tones on its first lag and the weekday dummies. We then obtain the "connected" sentiment for an indexed commodity by taking a value-weighted average of indexed commodities from other sectors. For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The news tones and controls are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The *t*-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Exclude Financial Crisis (2008/09/15 – 2009/06/30)					
	Panel A: Contemporaneous			Panel B: Predictive		
	Indexed	Non-indexed	All	Indexed	Non-indexed	All
Cnn. Sentiment	0.0770***	0.0430***	0.0432***			
	(39.87)	(16.04)	(16.45)			
Cnn. Sentiment×Indexed			0.0317***			
			(9.88)			
L.Cnn. Sentiment				-0.0042**	0.0023	0.0019
				(-2.25)	(0.86)	(0.71)
L.(Cnn. Sentiment×Indexed)						-0.0058*
						(-1.84)
L.Basis	0.0070	0.0162*	0.0094**	0.0061	0.0136	0.0082*
	(1.42)	(1.66)	(2.15)	(1.22)	(1.37)	(1.84)
L.Illiquidity	-1.17e-05	2.01e-05	-1.88e-05	-3.45e-05	1.41e-05	-1.43e-05
	(-0.12)	(0.17)	(-0.25)	(-0.36)	(0.12)	(-0.19)
L.Rvol	-0.0003	0.0002	-0.0002	-0.0005	0.0001	-0.0002
	(-0.35)	(0.16)	(-0.18)	(-0.48)	(0.04)	(-0.25)
Intercept	-0.0002	-2.05e-06	-0.0003	0.0014**	0.0013	0.0014***
	(-0.47)	(-0.00)	(-0.72)	(2.32)	(1.64)	(2.93)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	48,789	25,952	74,741	48,776	25,947	74,723
# of Individuals	15	8	23	15	8	23
Overall R-squared	3.73%	1.21%	2.83%	0.13%	0.15%	0.13%

A.4 Additional Results on Return Serial Dependence Excluding Roll Weeks

Table A7: Return Serial Dependence and Commodities' Total Index Exposure Excluding Roll Weeks

This table presents the results of regressing commodities serial dependence measure on commodities' total index exposure in period excluding rolling weeks. The roll week of a commodity is the corresponding week of the roll date which is the seventh calendar day of the maturity month. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. "Indexed" is a dummy variable, which equals 1 when the commodity is indexed and 0 otherwise. The control variables are of daily frequency ranging from January 2, 2003 to November 6, 2018. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Exclude Roll Weeks		
	Indexed	Non-indexed	All
L.Total Index Exposure	-0.0985***	-0.0073	-0.0237
	(-3.63)	(-0.35)	(-1.16)
L.(Total Index Exposure×Indexed)			-0.0653**
			(-2.27)
L.Basis	0.7545	3.6028***	1.4128***
	(1.46)	(3.72)	(2.99)
L.Illiquidity	0.0159	0.0439***	0.0273**
	(0.91)	(2.64)	(2.20)
L.Oil Implied Volatility	-0.4877***	-0.3245***	-0.4317***
	(-4.80)	(-2.79)	(-5.56)
Intercept	0.0863***	-0.0014	0.0630**
	(2.67)	(-0.03)	(2.10)
Individual Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
# of Observations	38,334	21,065	59,399
# of Individuals	15	8	23
Overall R-squared	0.36%	0.61%	0.43%

Table A8: Return Serial Dependence and Commodities' Individual Index Exposure Excluding Roll Weeks

This table presents the results of regressing commodities serial dependence measure on commodities' individual index exposure in period excluding rolling weeks. The roll week of a commodity is the corresponding week of the roll date which is the seventh calendar day of the maturity month. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Baseline		Liquidity Provision	
	Indexed	Non-indexed	Indexed	Non-indexed
L.Index Exposure	-0.0290***	-0.0073	-0.0478***	-0.0091
	(-3.41)	(-0.35)	(-4.06)	(-0.41)
L.(Index Exposure×Illiquidity)			-0.0804***	-0.0129
			(-3.39)	(-0.36)
L.Basis	0.8333	3.6028***	0.7776	3.6039***
	(1.57)	(3.72)	(1.47)	(3.72)
L.Illiquidity	0.0150	0.0439***	-0.0215	0.0425**
	(0.84)	(2.64)	(-0.95)	(2.49)
L.Oil Implied Volatility	-0.4602***	-0.3245***	-0.4494***	-0.3260***
	(-4.56)	(-2.79)	(-4.48)	(-2.80)
Intercept	0.0619*	-0.0014	0.0431	-0.0017
	(1.83)	(-0.03)	(1.21)	(-0.04)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Observations	38,334	21,065	38,334	21,065
# of Individuals	15	8	15	8
Overall R-squared	0.31%	0.61%	0.40%	0.61%

Table A9: Return Serial Dependence and Components of Commodities' Individual Index Exposure Excluding Roll Weeks

This table presents the results of regressing commodities serial dependence measure on the components of commodities' individual index exposure, i.e., index market share and trading volume in period excluding rolling weeks. The roll week of a commodity is the corresponding week of the roll date which is the seventh calendar day of the maturity month. The serial dependence measure is defined as $(r_{it}r_{it-1})/\sigma_i^2$. The index market share is defined as the index open interest divided by the total open interest. The CIT data is of weekly frequency ranging from January 3, 2006 to November 6, 2018. The t-statistics reported in the parenthesis are based on Newey-West standard errors with 4 lags. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Exclude Roll Weeks					
	Indexed			Non-indexed		
L.Index Market Share	-0.2407*		-0.2601*	0.0459		0.1071
	(-1.68)		(-1.69)	(0.04)		(0.08)
L.Trading Volume		-0.0003**	-0.0003***		-0.0008	0.0014
		(-2.38)	(-3.33)		(-0.82)	(1.07)
L.Basis	0.5112	0.3114	0.4888	2.5931**	2.8778***	2.7329**
	(1.08)	(0.62)	(1.03)	(2.52)	(3.47)	(2.56)
L.Illiquidity	0.0244	0.0296	0.0217	0.0417***	0.0495***	0.0428***
	(1.57)	(1.27)	(1.38)	(2.63)	(3.42)	(2.68)
L.Oil Implied Volatility	-0.4517***	-0.5616***	-0.4504***	-0.3532***	-0.3432***	-0.3531***
	(-4.54)	(-5.02)	(-4.53)	(-2.87)	(-2.57)	(-2.87)
Intercept	0.1006**	0.3408***	0.2205***	0.0386	0.2445*	-0.0570
	(2.55)	(2.83)	(3.85)	(0.85)	(1.81)	(-0.25)
Individual Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	41,504	50,444	41,504	21,342	28,086	21,342
# of Individuals	15	15	15	15	15	15
Overall R-squared	0.27%	0.19%	0.29%	0.57%	0.29%	0.58%