Financialization and Commodity Market Serial Dependence*

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

This Draft: January 2019

Abstract

Recent financialization in the commodity market 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, or "excessive co-movement" at daily frequency. Excessive co-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 excessive co-movement. Overall, our paper adds value to the understanding of price discovery and market efficiency of financialized commodity futures markets.

JEL Classification: G12, G40, Q02.

Keywords: Return autocorrelation; Index trading; News sentiment; Non-fundamental shocks.

*We have benefited from comments and suggestions from Li Guo, Jing Wu and seminar participants at Notre Dame University and the 3rd Australian Commodity Markets 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@mail.tsinghua.edu.cn.

§Singapore Management University, School of Economics, 90 Stamford Road, 178903, Singapore. Email: yubo.tao.2014@phdecons.smu.edu.sg

1 Introduction

The last two decades witnessed the financialization of the commodity markets. According to the estimates from the Commodity Futures Trading Commission (CFTC), investment flows to various commodity indices exceeded $600 billion during the period from 2000 to 2017. 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 co-movement 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. [1]

[Figure 1 is about here.]

Co-movement 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. In a review article, Cheng and Xiong (2014) write "direct tests of price impacts and impacts on correlations should incorporate clear identification strategies in the spirit of Angrist and Pischke (2010)."

Our paper fills this gap. Our main variable of interest is the daily return autocorrelation instead of return correlation 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. With a backward rolling window of ten years, we do not observe a clear trend in the past 38 years in the daily autocorrelation in returns of the non-indexed commodity portfolio (NIDX). 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.[2] They entered the negative territory around

[1] 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.

[2] GSCI was originally developed in 1991, by Goldman Sachs. In 2007, ownership was transferred to Standard &

2005 and became significantly negative since 2006. While a declining index return autocorrelation 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 of price discovery.* It suggests that prices across multiple indexed commodities can overshoot and subsequently revert at the same time, resulting in "excessive co-movement," even at the index level. Our paper hence provides empirical evidence to the theoretical hypothesis by Goldstein and Yang (2017), which writes "The trading of index traders can inject both fundamental information and unrelated noise into the futures prices. Therefore, market efficiency can either increase or decrease with financialization." The negative autocorrelation certainly demonstrates a negative side effect of the financialization, likely due to unrelated noise injected into the futures markets.

[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 some commodities in the index to the rest. Figure 3 provides some supporting evidence. We plot rolling average daily return autocorrelations with a shorter backward window of three years against a measure of exposure to index trading, in a shorter sample starting from 2006 (due to the availability of the indexing exposure measure). We see a clear negative relation between the index autocorrelation and the index exposure measure. In other words, when institutional investors are trading commodity index more actively, the commodity index return (on both GSCI and BCOM) becomes more negatively autocorrelated. No such relation is observed for the portfolio

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.

of non-indexed commodities (NIDX).

[Figure 3 is about here.]

The rest of the paper provides additional evidence that pins down the link between financialization and excessive return co-movement among indexed commodities.

We ran three sets of tests. In the first, we directly measure daily sentiment on a commodity derived from its news articles. Specifically, the sentiment measure is constructed as the residual from orthogonalizing news tone of articles about a commodity on fundamental factors of the same commodity. 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.) using institutional investors' total exposure to that commodity as the weight. We find that the "connected" index sentiment is related to contemporaneous return on corn positively and significantly, but to negatively and significantly predict corn's return tomorrow. The positive contemporaneous correlation could suggest that sentiment is propagated from some commodities to others in the same index. It could also suggest that our sentiment measure may still contain common fundamental factors. Nevertheless, the fact that such a positive correlation reverts on the next day confirms the existence of "non-fundamental" shocks. As index trading propagates such "non-fundamental" shocks across commodities in the same index, it results in synchronized price overshoots and reversals and therefore "excessive" co-movement. We confirm that the results are not driven by the 2008-2009 great financial crisis. In fact, the results are stronger after excluding the financial crisis period. As a placebo test, we repeat the same tests among non-indexed commodities but do not find evidence for such "non-fundamental" shocks.

Our second set of tests directly link "excessive" co-movement to index trading. We first confirm that the index sentiment propagation results are much stronger during periods when the index is more exposed to institutional trading. More formally, we also regress daily autocorrelation of indexed commodities on the abnormal index exposure measure and find a significantly negative coefficient. Abnormal index exposure does not affect the autocorrelation on the return of non-indexed commodities at all.

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 different weights across two popular commodity indices (GSCI and BCOM). The relative weight difference arises in a rather ad-hoc fashion and is determined in 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 excessive exposure to ETFs based on GSCI (relative to that based on BCOM).

Our paper links to three strands of literatures. First, it adds value to the debates of price impact of the index investments. 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 future markets. Singleton (2013) and Gilbert (2010) show that index investments do predict movements of commodity prices. Mou (2011) and Yan *et al.* (2019) document that index rebalancing causes futures prices shift significantly. Henderson *et al.* (2014) document that the hedging activities of issuers of commodity-linked notes can significantly influence commodity futures prices. Brogaard *et al.* (2018) documents that firms using commodity indices have relatively worse performance, and hence index investing distorts the price signal in commodity markets. However, Büyükşahin and Harris (2011), Irwin and Sanders (2012) and Sanders and Irwin (2011) find little evidence that the index position changes link to price movements in futures markets. Hamilton and Wu (2015) presents a mixed result. Our paper adds value to this debate by presenting a supporting evidence on the price pressure from the index investment. Particularly, prices of indexed commodities over-shoot and reverse subsequently when reacting to non-fundamental sentiment shocks, while non-indexed commodities do not show such a pattern.

Second, our paper relates to the study of co-movements among different commodities. For example, Pindyck and Rotemberg (1990) document a co-movement of unrelated commodities and attribute it to common effects of macroeconomic variables. However, Ai *et al.* (2006) argues that the co-movement among commodities is not excessive, and can be explained by common tendencies in demand and supply factors. Different with these studies, our results suggest that commodity index trading can propagate price pressure across commodities in the same index. To clarify, while such a price pressures results in "excessive" co-movement at the index level, we do not claim it to drive the boom and bust commodity cycles entirely. Indeed, our results seem to be stronger after excluding the greater financial crisis from our analysis.

Third, 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 *et al.*, 2012), non-fundamental price changes (Chen *et al.*, 2004), excessive comovement (Barberis *et al.*, 2005; Greenwood, 2005, 2008; Da and Shive, 2018), a deterioration of the firms information environment (Israeli *et al.*, 2017), increased non-fundamental volatility in individual stocks (Ben-David *et al.*, 2017), and reduced welfare of retail investors (Bond and García, 2017). 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, and section 4 concludes.

2 Data and Variable Construction

In this section, we describe the commodities used in our analyses and introduce two most popular commodity indices and their construction. We then describe how we measure the exposure of a commodity to indexing. Finally, we discuss our news database and how we construct a news-based sentiment measure for each commodity.

2.1 Commodities and commodity indices

Commodity price data are obtained from Pinnacle Corporation. Following Kang *et al.* (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.[3] The excess return r_{it} on commodity i on date t is calculated as:

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

where $F_i(t,T)$ is the futures price on day t for a futures contract maturing on date T.

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 in 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.

Investors can use three types of financial instruments to gain exposure to a commodity index: commodity index swaps, exchange-traded funds (ETF), and exchange-traded notes (ETN). We

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

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.

[Table 2 is about here.]

Table 2 highlights the benefit of investing in the commodity market. Commodities offer attractive annual Sharpe ratios that are comparable to that in the equity market. More importantly, their return correlations with the equity market (proxied using the S&P 500 index) are fairly low with an average correlation of 0.16, bringing in additional diversification benefit. Take Gold for example, its annual Sharpe ratio is 0.47 and its return correlation with the equity market is almost zero in our sample period. Not surprisingly, given these attractive features, institutional investors became more willing to invest in commodities, especially since the start of financialization that makes it easy for them to trade commodity indices.[4]

Energy sector, especially crude oil (CL) and natural gas (NG), did not perform well in our sample period from 2003 to 2015. Since both GSCI and BCOM indices place heavy weights in the energy sector, both indices suffered losses in the same period.

2.2 Commodity index exposure

The exposure of a commodity to index trading by institutional investors (or index exposure in short), is defined as the total market cap of index trading on that commodity as the percentage of total market cap of all trading on that commodity. Each Tuesday, the CFTC releases a weekly

[4] Tang and Xiong (2012) argue that the capital inflow into commodity futures markets integrates the segmented commodity futures markets with mainstream financial markets, for example the equity markets; particularly they show an increasing correlation between commodity and equity indexes especially during the financial crisis. However, the correlation declines in recent years (Bhardwaj *et al.*, 2015) likely caused by the capital outflow from the commodity markets. The overall correlation between GSCI and SP500 is around 0.3 in our sample.

Commitments of Traders (CoT) report, which includes the total open interest of each commodity and the long/short positions of each type of traders.[5] It also includes a supplemental Commodity Index Trader (CIT) report that shows the positions of a set of index traders identified by the CFTC since January 3, 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:

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

In this paper, we define the index open interest as the average of the long and short positions of index traders: $\mathit{Open\ Interest}^{Idx} = (\mathit{Long}^{Idx} + \mathit{Short}^{Idx})/2$. Therefore, we can derive the dollar value open interests (Market Cap) on index/total trading as:

$$\mathit{Market\ Cap}_t^{Idx} = \sum_{i=1} \mathit{Open\ Interest}_{it}^{Idx} \times \mathit{Contract\ Size}_i \times \mathit{Price}_{it}, \quad (3)$$

$$\mathit{Market\ Cap}_t^{All} = \sum_{i=1} \mathit{Open\ Interest}_{it}^{All} \times \mathit{Contract\ Size}_i \times \mathit{Price}_{it}. \quad (4)$$

Unfortunately, the CIT only reports 13 agricultural commodities (listed in Table 1) and it covers no commodities in the energy and metals sectors. 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[6] and cocoa (CC), feeder cattle (FC) and Kansas wheat (KW) for GSCI. Then, note that index traders replicate the index by allocating across the commodities according to the known weights[7] $\delta_{jy(t)}^{(i)}, i \in \{G, B\}$, we can separately estimate CL's dollar value long/short positions on index trading, $X_{CL,t}$, on GSCI/BCOM

[5] 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.

[6] Note that soybean meal (SM) was added to BCOM in 2013.

[7] Both weights reported in the GSCI and BCOM manuals are dollar value weights.

trading as below:

$$\widehat{X}^B_{CL,t} = \begin{cases} \dfrac{\delta^B_{CL,y(t)}}{\delta^B_{BO,y(t)}} X_{BO,t}, & \text{if } y(t) < 2013, \\ \dfrac{1}{2}\left(\dfrac{\delta^B_{CL,y(t)}}{\delta^B_{BO,y(t)}} X_{BO,t} + \dfrac{\delta^B_{CL,y(t)}}{\delta^B_{SM,y(t)}} X_{SM,t}\right), & \text{if } y(t) \geq 2013. \end{cases} \tag{5}$$

$$\widehat{X}^G_{CL,t} = \frac{1}{3}\left(\frac{\delta^G_{CL,y(t)}}{\delta^G_{CC,y(t)}} X_{CC,t} + \frac{\delta^G_{CL,y(t)}}{\delta^G_{FC,y(t)}} X_{FC,t} + \frac{\delta^G_{CL,y(t)}}{\delta^G_{KW,y(t)}} X_{KW,t}\right). \tag{6}$$

where $y(t)$ denotes the year of 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 value of index trading for commodity i at time t is estimated as

$$X_{it} = Position_{it} \times ContractSize_i \times Price_{it}. \tag{7}$$

Combining the estimates above, Masters (2008) propose to estimate the total market cap of CL on index trading as:

$$\widehat{Market\ Cap}^{Idx}_{CL,t} = \widehat{Market\ Cap}^{B}_{CL,t} + \widehat{Market\ Cap}^{G}_{CL,t}. \tag{8}$$

However, as pointed out by Irwin and Sanders (2011), 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

$$\widehat{X}^{Idx}_{it} = \begin{bmatrix} \delta^G_{iy(t)} & \delta^B_{iy(t)} \end{bmatrix} \begin{bmatrix} \sum_{j \in CIT} \left(\delta^G_{jy(t)}\right)^2 & \sum_{j \in CIT} \delta^G_{jy(t)} \delta^B_{jy(t)} \\ \sum_{j \in CIT} \delta^B_{jy(t)} \delta^G_{jy(t)} & \sum_{j \in CIT} \left(\delta^B_{jy(t)}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j \in CIT} \delta^G_{jy(t)} X^{Idx}_{jt} \\ \sum_{j \in CIT} \delta^B_{jy(t)} X^{Idx}_{jt} \end{bmatrix}, \tag{9}$$

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 (9) we obtain both the long and short dollar value positions for unreported commodities, and thus the total market cap. Combining the CIT-reported open interest on index trading data, we can estimate the daily market cap of the total (index) trading by

$$\widehat{Market\ Cap}_t^{Idx} = \sum_{j \in Idx} \widehat{Market\ Cap}_{jt}. \tag{10}$$

Figure 4 plots the market capital of index traders. The figure shows that before the financial crisis and around 2011, the market capital of index trading reached its highest level. It trended down afterwards.

[Figure 4 is about here.]

Then, the index exposure is defined as

$$Indexing_t = Market\ Cap_{it}^{Idx} \Big/ Market\ Cap_{it}^{All}, \tag{11}$$

To study the impact of index trading on commodities returns, we finally propose a flow measure, abnormal index exposure, by evaluating how much incremental market cap on index trading contributes to the total market cap traded, i.e.,

$$Abn.\ Indexing_t = \frac{Market\ Cap_t^{Idx} - Market\ Cap_{t-1}^{Idx}}{Market\ Cap_t^{All}}. \tag{12}$$

2.3 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.[8] 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 compute the net tone as the difference between the positive tone and the negative tone. We then calculate the abnormal net tone as the residual of regressing the net tones on its first lag and the month dummies. Finally, to extract news sentiment, we then orthogonalize the abnormal net tones on commodity fundamentals such as returns, basis and liquidity. The reasons to include those controls are as follows: as shown in Brennan (1958) and Gorton *et al.* (2012) basis mainly represents the level of inventory, which can be considered as the mismatch between demand and supply of a certain commodity; Szymanowska *et al.* (2014) have shown that basis is a determinant of commodity risk premium. Since hedging activity from production firms may cause extra trading in futures markets, we include Amihud illiquidity as a control in our regression, which is considered as the best liquidity measure in commodity markets (Marshall *et al.* (2011)). Specifically, for each commodity, we run the following regression:

$$Abn.\ Net\ Tone_t = \alpha + \beta' \begin{bmatrix} r_t \\ r_{t-1} \end{bmatrix} + \theta' \begin{bmatrix} Basis_t \\ Basis_{t-1} \end{bmatrix} + \phi' \begin{bmatrix} Illiquidity_t \\ Illiquidity_{t-1} \end{bmatrix} + \epsilon_t, \tag{13}$$

where *Basis* is the log basis[9] and *Illiquidity* is the Amihud (2002) illiquidity measure.[10] We then treat the residual of the regression as the sentiment measure for each commodity. The descriptive statistics of our sentiment measure for each commodity are shown in Table 3.

[Table 3 is about here.]

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

[9]The log basis is defined as $\frac{\ln(F_i(t,T_1))-\ln(F_i(t,T_2))}{T_2-T_1}$, where $F_i(t,T_1)$ and $F_i(t,T_2)$ are the prices of the closest and next closest to maturity contracts for commodity i.

[10]For each commodity, we compute its Amihud's (2002) illiquidity measure as the ratio of the absolute value of its daily return divided by its dollar trading volume in the same day.

As evident in Table 3, crude oil receives more news coverage than other commodities.[11] The sentiment measures have zero means by construction. Their average standard deviations is 0.1069 ranging from 0.0522 for oat (O-) to 0.1953 for soybean (S-).

3 Empirical Analysis

We conduct three sets of empirical analyses. We first study the propagation of "non-fundamental" shocks across commodities using our news-based sentiment measure. We then examine daily return autocorrelations for commodity indices 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.,

$$Cnn.\ Sentiment_{it} = \sum_{S(j)\neq S(i)} W_{jy(t)} Sentiment_{jt}, \tag{14}$$

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

$$W_{jy(t)} = \frac{E_{y(t)}(\$Open\ Interest_{jt}^{Idx})}{\sum_j E_{y(t)}(\$Open\ Interest_{jt}^{Idx})}, \tag{15}$$

with $E_{y(t)}(\$Open\ Interest_{jw(t)}^{Idx})$ 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

[11] Since oat and rough rice are close substitutes, the news tone in our dataset treats them as one commodity; we hence use identical news tone for both oat and rough rice.

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, they are more likely to co-move within sector than across sectors. In this sense, our measure alleviates the concerns for fundamental-driven co-movement. 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,

$$r_{it} = \beta_0 + \beta_1 \cdot \mathit{Cnn.\ Sentiment}_{it} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \tag{16}$$

$$r_{it} = \beta_0 + \beta_1 \cdot \mathit{Cnn.\ Sentiment}_{it-1} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \tag{17}$$

where $\boldsymbol{X}$ is a vector of control variables including lagged returns, lagged log basis, and lagged Amihud's (2002) illiquidity measure for each commodity. We also include the lagged change in 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).[12] Both individual/sector fixed effect and year fixed effect are controlled in the regression. The results are reported in Table 4.

[12] 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.

[Table 4 is about here.]

Focusing on Panel A, we find a positive and significant contemporaneous relation between the indexed commodity return and its "connected" sentiment measure. The relation is consistent with the notion that index trading could propagate sentiment across commodities within the same index. It may also reflect fundamental-driven co-movement that are not fully controlled for using return, basis and liquidity measures in our sentiment measure construction. Indeed, we also observe positive and significant coefficients on "connected" sentiment measures for non-indexed commodities even though index trading is not possible here.

While both sentiment spillover and common fundamental factor can explain the positive contemporaneous relation we observed in Panel A, only sentiment predicts future return negatively. This is because sentiment-induced trading induces a "non-fundamental" shock or price pressure in contemporaneous return and such a shock will be reverted in the future. For example, positive sentiment on energy may induce institutional investors to buy the commodity index. Such trading propagates the sentiment from energy sector to other indexed commodities and results in a positive price pressure in corn today, explaining the positive contemporaneous relation between the return on corn and its "connected" sentiment. As the positive price pressure on corn reverts tomorrow, the "connected" sentiment today should negatively predict corn's return tomorrow.

Such a negative return predictability by "connected" sentiment is exactly what we find in Panel B for indexed commodities. The coefficient on "connected" sentiments is likely to capture the impact of sentiment spillover. For instance, a coefficient of -0.0052 (t-value of -1.86) on the "connected" sentiment implies that a one-standard-deviation increase in the sentiment of "connected" indexed commodities propagate a price pressure of 1.9 basis points. The last column in Panel B does not find significant return predictability among non-indexed commodities, again consistent with the notion that sentiment is mainly propagated via index trading.

Figure 1 highlights the well-known commodity boom and bust cycles during the 2008-2009 financial crisis. A natural concern is whether our results so far are completely driven by such market wide fluctuations. To that end, we repeat in Table 5 our previous analyses, but after excluding the great financial crisis (2008/09/15 to 2009/06/30 according to the bankruptcy of Lehman Brothers and NBER recession dating as in Tang and Xiong (2012)). We find our results to be similar, if not stronger, after removing the great financial crisis. For example, the "connected" sentiment measure is now associated with a bigger coefficient (in absolute term) of -0.0072 in predicting the return reversal on the next day. In other words, our results so far are more consistent with high-frequency (daily) price overshoots and reversals than low-frequency (business cycle) variation in fundamental risk factors.

[Table 5 is about here.]

Turning to the control variables, consistent with Szymanowska *et al.* (2014), lagged basis makes a positive prediction (although insignificant) on commodity returns listed in Table 4. The coefficient of illiquidity also agrees with Kang *et al.* (2019), i.e. in an illiquid market hedgers have to pay a higher liquidity premium to speculators in order to execute their hedging. Past returns positively predict futures returns for non-indexed commodities, likely caused by under-reaction, since non-indexed commodities are relatively illiquid, and hence their futures markets are not efficient in transmitting fundamental information. This can be seen from Table 2, where non-indexed commodities normally have positive autocorrelations. However, indexed commodities such as energy and metals often have negative autocorrelation, which explains the negative coefficient of lagged returns in Table 4 and 5. Implied volatility shocks of crude oil significantly predict commodity returns of indexed commodities, but with less strength for non-indexed commodities.

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 high. To test this conjecture, we split the sample into two subsamples based on our abnormal index exposure measure defined in the previous section. Specifically, we first calculate each week's average abnormal index exposure.

“High” of H (“Low” or L) index exposure period includes weeks whose average abnormal index exposure measure is above (below) the median of the full sample. We then re-run the previous regression analyses in the “H” and “L” subperiods separately:

$$r_{it} = \beta_0 + \beta_1 \cdot \mathit{Cnn.\ Sentiment}_{it} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it}, \quad t \in \{H, L\} \tag{18}$$

$$r_{it} = \beta_0 + \beta_1 \cdot \mathit{Cnn.\ Sentiment}_{it-1} + \theta' \boldsymbol{X}_{it-1} + \epsilon_{it},, \quad t \in \{H, L\} \tag{19}$$

where $w(t)$ denotes the week of date t. Both individual/sector fixed effect and year fixed effect are controlled in the regression. The results are reported in Table 6.

[Table 6 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 -0.0146 (t-value of -3.55) during months with abnormally high amount of index trading. The economic magnitude is large. A coefficient of -0.0146 implies that a one-standard-deviation increase in the sentiment of “connected” indexed commodities propagate a price pressure of 5.5 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.

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

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 co-movement 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 co-movement 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). It is almost always positive throughout the sample period.

Using our index exposure measure, Figure 3 shows that the gradual decline in the index return autocorrelation is accompanied with a rising exposure to index trading during the same sample period since 2006. Of course, "missing" factors can drive the common trend in the low-frequency variations of both daily index return autocorrelation and 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 measures. In the first test, we check if the autocorrelation coefficient of the returns will be decreasing when abnormal index exposure increases. In particular, we include an interaction term between lagged return and abnormal index exposure in the autoregressive model of commodity returns:

$$r_{it} = \beta_0 + \beta_1 \cdot r_{i,t-1} + \beta_2 \cdot \mathit{Abn.\ Indexing}_{t-1} + \gamma \cdot (r_{i,t-1} \times \mathit{Abn.\ Indexing}_{t-1}) + \theta' \boldsymbol{X}_{i,t-1} + \epsilon_{it}, \quad (20)$$

where the vector $\boldsymbol{X}$ contains each commodity's lagged basis, lagged Amihud illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity as control variables in the spirit of Nagel (2012) and Bianchi *et al.* (2016). The key parameter of interest is γ and the results are presented in Table 7.

[Table 7 is about here.]

Table 7 reveals three sets of interesting results. First, we observe negative and significant coefficients on the interaction term only for the indexed commodities. In other words, abnormally high index trading today implies a 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. Second, we find index exposure measure to have nothing to do with the return autocorrelation of non-indexed commodities, consistent with the pattern in Figure 3. This result serves as a nice placebo. Finally, the result is still significant, when excluding the financial crisis period.

Complementary to interaction results, we use abnormal index exposure to directly predict the return serial dependence following Baltussen *et al.* (2018). Specifically, we run the following daily panel regression of each commodity's serial dependence measure on the lagged abnormal index exposure measure and controls:

$$(r_{it} \cdot r_{i,t-1})/2\sigma_i^2 = \beta_0 + \beta_1 \cdot \textit{Abn. Indexing}_{t-1} + \theta' \boldsymbol{X}_{i,t-1} + \epsilon_{it}, \tag{21}$$

where σ_i^2 is the sample variance of commodity i's returns, and vector $\boldsymbol{X}$ contains each commodity's lagged serial dependence measure, lagged basis, lagged Amihud illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity as control variables in the spirit of Baltussen *et al.* (2018), Nagel (2012) and Bianchi *et al.* (2016). We run the panel regression for indexed and non-indexed commodities separately. The results are presented in Table 8.

[Table 8 is about here.]

Consistent with Table 7, we observe negative and significant coefficients on the index exposure measure only for the indexed commodities. The economic magnitude of such effect is large. For example, coefficient of -6.2068 means that a one-standard-deviation increase in the abnormal index exposure makes its daily return autocorrelation -2.55% more negative. Lastly, the result remains significant after excluding the great financial crisis from our analysis.

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 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 was 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 trading and negative daily return autocorrelation) we documented in Table 7 and 8. This is no reason to believe that a broadly increasing trend to invest in a general commodity basket should be correlated with abnormal trading activities in two specific commodity indices on a day-to-day basis. Nevertheless, in this subsection, 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 *et al.* (2018) 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 a portfolio of commodities that are overweighted in GSCI index (relative to BCOM index), its daily return autocorrelation should be more negatively correlated with trading measure on GSCI (relative to that on BCOM).

We implement the test by constructing zero-investment portfolio ("GSCI/BCOM OW portfolio"). Take GSCI OW portfolio as an example, we first compare commodity i's weight in GSCI, $w_{jy(t)}^{GSCI}$, to its weight in BCOM, $w_{jy(t)}^{BCOM}$:

$$OW_{jy(t)}^{GSCI} = \begin{cases} w_{jy(t)}^{GSCI} - w_{jy(t)}^{BCOM}, & \text{if } w_{jy(t)}^{GSCI} > 0, \\ 0, & \text{if } w_{jy(t)}^{GSCI} = 0, \end{cases} \tag{22}$$

where $y(t)$ is the year of date t. Then, we assign weight ϖ_{jt}^{GSCI} on each commodity j as

$$\varpi_{jt}^{GSCI} = \left(OW_{jy(t)}^{GSCI} - \frac{1}{N} \sum_{j=1}^{N} OW_{jy(t)}^{GSCI} \right) r_{t-1}^{GSCI}, \tag{23}$$

and calculate the portfolio return $R_t^{GSCI} = \sum_{j=1}^{N} \varpi_{jt}^{GSCI} r_{jt}$, where r_{t-1}^{GSCI} is the return on GSCI index and r_{jt} is the return on commodity j. A portfolio that bases on the commodities that are overweighted in BCOM can be constructed in the same way.

Next, we compute the ETF indexing measure for each commodity index. Specifically, we first compute the total market cap ($\mathit{Shares\ Outstanding}_t \times NAV_t$) of each commodity index's ETF/ETNs (i.e., GSG and GSP for GSCI and DJCI and DJP for BCOM). In the second step, we calculate the total market cap of commodities on GSCI/BCOM trading by estimating each commodity's dollar value trading positions on each index with modified Hamilton and Wu (2015) method as below:

$$\widehat{X}_{it}^{G} = \begin{bmatrix} \delta_{iy(t)}^{G} & 0 \end{bmatrix} \begin{bmatrix} \sum_{j\in CIT} \left(\delta_{jy(t)}^{G}\right)^2 & \sum_{j\in CIT} \delta_{jy(t)}^{G}\delta_{jy(t)}^{B} \\ \sum_{j\in CIT} \delta_{jy(t)}^{B}\delta_{jy(t)}^{G} & \sum_{j\in CIT} \left(\delta_{jy(t)}^{B}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j\in CIT} \delta_{jy(t)}^{G} X_{jt}^{Idx} \\ \sum_{j\in CIT} \delta_{jy(t)}^{B} X_{jt}^{Idx} \end{bmatrix}, \tag{24}$$

$$\widehat{X}_{it}^{B} = \begin{bmatrix} 0 & \delta_{iy(t)}^{B} \end{bmatrix} \begin{bmatrix} \sum_{j\in CIT} \left(\delta_{jy(t)}^{G}\right)^2 & \sum_{j\in CIT} \delta_{jy(t)}^{G}\delta_{jy(t)}^{B} \\ \sum_{j\in CIT} \delta_{jy(t)}^{B}\delta_{jy(t)}^{G} & \sum_{j\in CIT} \left(\delta_{jy(t)}^{B}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j\in CIT} \delta_{jy(t)}^{G} X_{jt}^{Idx} \\ \sum_{j\in CIT} \delta_{jy(t)}^{B} X_{jt}^{Idx} \end{bmatrix}. \tag{25}$$

The abnormal ETF indexing is then defined as a ratio of the index-related ETF/ETNs' total market cap changes to the total market cap of the index, i.e.,

$$Abn.\ ETF\ Indexing_t^{(i)} = \frac{ETF\ Market\ Cap_t^{(i)} - ETF\ Market\ Cap_{t-1}^{(i)}}{Market\ Cap_t^{(i)}}, \quad i \in \{GSCI, BCOM\}. \tag{26}$$

Then, we define the GSCI/BCOM relative ETF indexing measure as

$$Relative\ ETF\ Indexing_t^{(i)} = Abn.\ ETF\ Indexing_t^{(i)} - Abn.\ ETF\ Indexing_t^{(j)}, \tag{27}$$

where $i = GSCI, j = BCOM$ or $i = BCOM, j = GSCI$.

Finally, we regress the GSCI/BCOM OW portfolio's returns on the lagged GSCI/BCOM relative ETF indexing measure with controls, i.e.,

$$R_t^{(i)} = \beta_0 + \beta_1 \cdot Relative\ ETF\ Indexing_{t-1}^{(i)} + \theta' \boldsymbol{X}_{t-1} + \epsilon_t, \quad i \in \{GSCI, BCOM\}. \tag{28}$$

where $\boldsymbol{X}$ is a vector of control variables motivated by Bianchi *et al.* (2016), Nagel (2012) and Baltussen *et al.* (2018), which contains the lagged GSCI (BCOM) average return over the past 21 trading days, lagged realized GSCI (BCOM) volatility over the past 250 trading days, the lagged log GSCI (BCOM) related ETF/ETNs' trading volume detrended with one-year average log trading volume, and lagged implied volatility of crude oil options with nearest maturity. The results using the full sample and the sample excluding financial crisis for both portfolios are shown in Table 9.

[Table 9 is about here.]

The results in Table 9 strongly support a causal interpretation that index trading drives excessive co-movement and negative index return autocorrelation. For a portfolio of commodities that are relatively overweighted in index i, its daily return autocorrelation is indeed more negatively correlated with relative trading exposure to index i. The results hold for both GSCI and BCOM indices and after excluding the great financial crisis.

4 Conclusion

We examine the impact of recent financialization in the commodity market on excessive co-movement among indexed commodities. Using news-based sentiment measures, we find that index trading enabled by financialization can propagate non-fundamental shocks from some commodities to others in the same index, giving rise to price overshoots and subsequent reversals, or "excessive co-movement" at daily frequency. Excessive co-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 excessive co-movement. Such excessive co-movement could contribute to the boom-and-bust cycles observed during the recent financial crisis, even though it does not drive such cycles.

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 co-movements even at the index level. Excessive price movement could impose costs on institutional investors who trade often and individual investors who invest in commodities through those institutions. In addition, as the recent paper by Brogaard *et al.* (2018) argues, inefficient commodity prices could even distort real decisions of a firm.

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.

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. (2018). Indexing and stock market serial dependence around the world. *Journal of Financial Economics, Forthcoming.*

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. (2017). Do etfs increase volatility? *Journal of Finance, Forthcoming.*

Bhardwaj, G., Gorton, G. and Rouwenhorst, G. (2015). Facts and fantasies about commodity futures ten years later. *NBER Working Paper.*

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. (2017). Informed trading, indexing, and welfare. *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. (2018). The economic impact of index investing. *The Review of Financial Studies, Forthcoming.*

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

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.

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*.

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

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. (2011). Index funds, financialization, and commodity futures markets. *Applied Economic Perspectives and Policy*, **33** (1), 1–31.

— and — (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. *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. (2011). Limits to arbitrage and commodity index investment: front-running the goldman roll. *Working Paper, Columbia University.*

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.

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.

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.

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

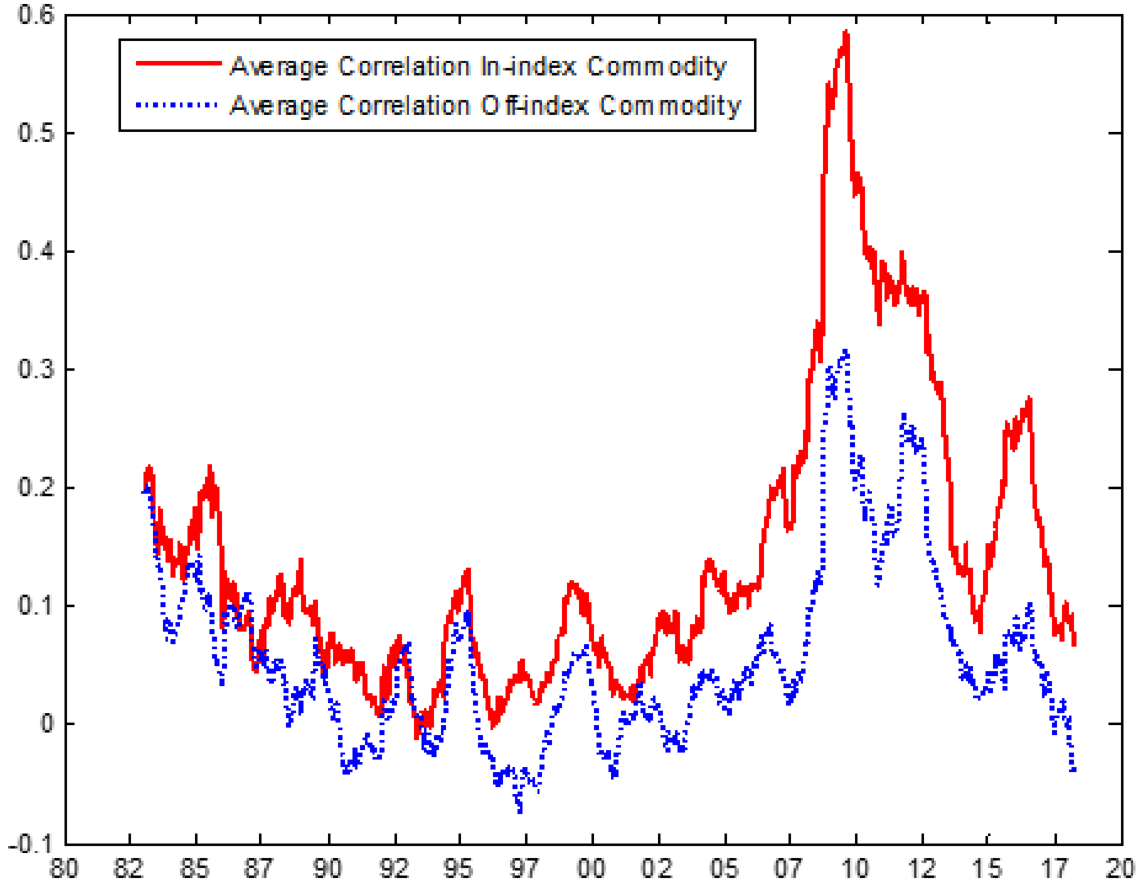

Figure 1: This figure plots the average return correlations of commodities in the GSCI and BCOM indices (indexed commodities) and those not included in these indices (non-indexed commodities). We follow the spirit of Tang and Xiong (2012) in computing these correlations. Specifically, we first calculate 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. The sample period is from 1980 to 2018.

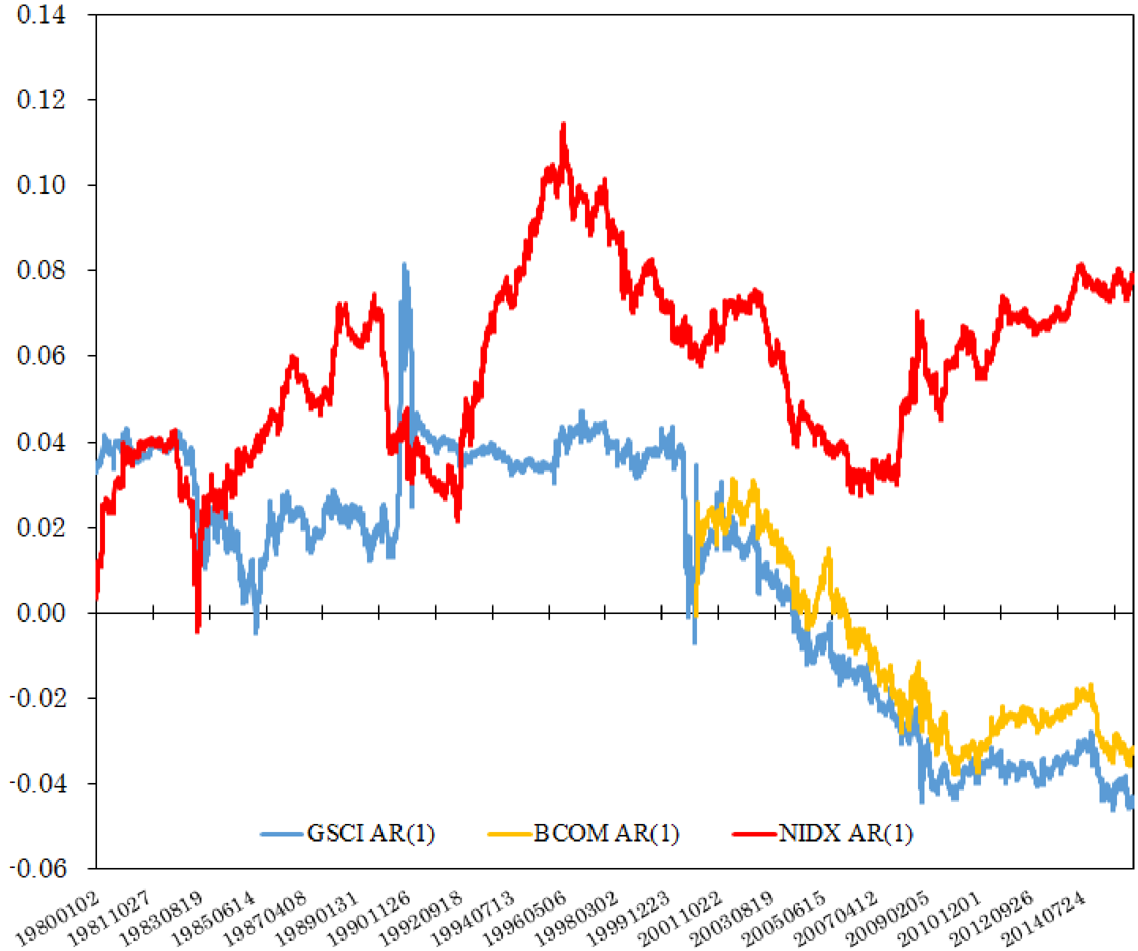

Figure 2: This figure plots the evolution of serial dependence in index returns from 1980 to 2018. Serial dependence is measured by first-order autocorrelation from index returns at the daily frequency. The indices are GSCI, BCOM and an equal-weighted portfolio non-indexed commodities (NIDX). Plotted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

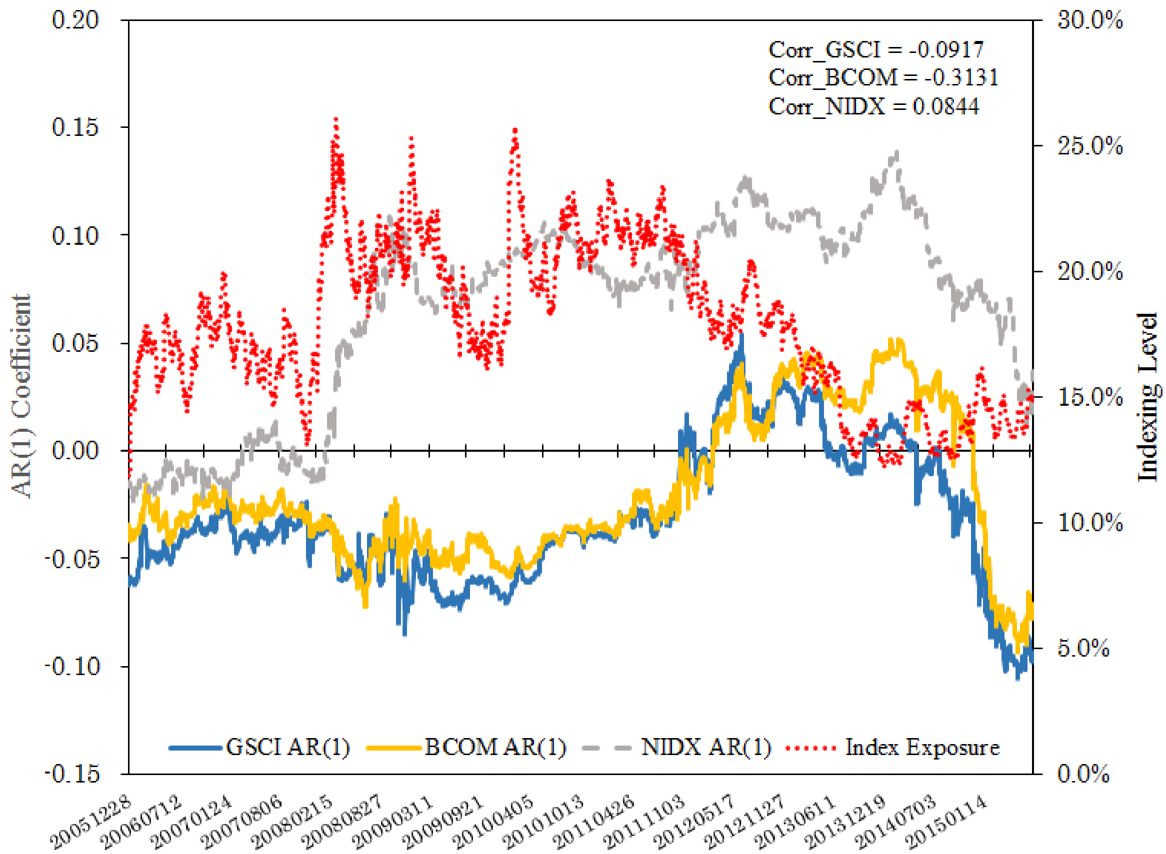

Figure 3: This figure plots the evolution of serial dependence in index returns and index exposure measure from 2006 to 2015. Serial dependence is measured by first-order autocorrelation from index returns at the daily frequency. The indices are GSCI, BCOM and an equal-weighted portfolio non-indexed commodities (NIDX). The index exposure is defined in equation (11). Plotted series are moving averages of these measures using a 3-year backward rolling window.

Figure 4: This figure plots the estimated market capital of index trading from 2006 to 2015 using Hamilton and Wu (2015) method.

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 GSCI and BCOM also include commodities traded in London, which are not included in our analysis. The commodities that are included in both indices are classified as "Indexed" commodities while these are not included in both 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' returns in columns 2-7. In column 8, we calculate the annualized Sharpe ratio (scaled by $\sqrt{252}$) of each commodity. In the last column, we show the correlation coefficient between the returns on each commodity/index and the return on S&P 500 composite index. NIDX denotes the equal weighted portfolio of non-indexed commodities. The sample is of daily frequency ranging from January 2, 2003 to December 29, 2015.

Commodity	Obs.	Mean	StDev.	Min	Max	AR(1)	Sharpe Ratio	ρ_{SP500}
Panel A: Energy								
CL	3266	-0.006%	0.0221	-0.1225	0.1427	-0.0650	-0.0455	0.2710
HO	3266	0.010%	0.0205	-0.0923	0.1041	-0.0401	0.0808	0.2267
NG	3266	-0.111%	0.0295	-0.1362	0.2064	-0.0583	-0.5974	0.0521
RB	3266	0.041%	0.0227	-0.1066	0.1385	-0.0395	0.2842	0.2291
Panel B: Grains								
BO	3272	0.009%	0.0157	-0.0698	0.0837	0.0094	0.0939	0.2122
C-	3272	0.008%	0.0186	-0.0989	0.0905	0.0316	0.0644	0.1491
KW	3272	0.009%	0.0188	-0.0860	0.0843	0.0082	0.0786	0.1419
MW	3272	0.037%	0.0180	-0.1071	0.2468	0.0379	0.3273	0.1187
O-	3272	0.045%	0.0209	-0.1125	0.1109	0.0744	0.3441	0.1279
RR	3272	0.017%	0.0154	-0.0595	0.0971	0.0777	0.1725	0.1272
S-	3272	0.055%	0.0162	-0.0782	0.0670	0.0081	0.5389	0.1588
SM	3272	0.084%	0.0182	-0.0805	0.0971	0.0374	0.7329	0.0971
W-	3272	-0.015%	0.0209	-0.0949	0.0919	-0.0049	-0.1145	0.1370
Panel C: Livestock								
FC	3264	0.024%	0.0090	-0.0583	0.0454	0.1277	0.4232	0.1298
LC	3264	0.014%	0.0096	-0.0616	0.0378	0.0797	0.2368	0.1347
LH	3272	-0.003%	0.0138	-0.0547	0.0543	0.0313	-0.0362	0.0407
Panel D: Metals								
GC	3266	0.036%	0.0122	-0.0934	0.0901	-0.0066	0.4667	0.0047
HG	3266	0.060%	0.0191	-0.1105	0.1235	-0.0826	0.4952	0.2942
PA	3266	0.040%	0.0209	-0.1237	0.1054	0.0721	0.3003	0.2210
PL	3266	0.022%	0.0144	-0.0916	0.0774	0.0509	0.2482	0.1769
SI	3266	0.049%	0.0217	-0.1771	0.1328	-0.0208	0.3593	0.1196
Panel E: Softs								
CC	3262	0.034%	0.0187	-0.0931	0.0974	0.0172	0.2900	0.1382
CT	3254	-0.006%	0.0188	-0.1236	0.0906	0.0592	-0.0477	0.1849
JO	3262	0.040%	0.0206	-0.1277	0.1669	0.0787	0.3064	0.0864
KC	3262	0.005%	0.0205	-0.1064	0.1385	-0.0237	0.0423	0.1433
LB	3272	-0.040%	0.0188	-0.0669	0.1064	0.0831	-0.3416	0.1233
SB	3262	0.004%	0.0201	-0.1163	0.0853	-0.0117	0.0291	0.1292
Panel F: Commodity Indices								
GSCI	3272	-0.010%	0.0152	-0.0829	0.0748	-0.0442	-0.1007	0.2869
BCOM	3266	-0.004%	0.0111	-0.0620	0.0581	-0.0308	-0.0584	0.2859
NIDX	3272	0.027%	0.0096	-0.0472	0.0441	0.0643	0.4531	0.2607

Table 3: Descriptive Statistics of Commodities' News Sentiment

This table provides some 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 are calculated in two steps. In the first step, we obtain the abnormal net tones by regressing the net tone (positive - negative) on its first lag and the month dummies. Then, in the second step, we orthogonalize the abnormal net tones (residuals in the first step regression) of each commodity on the its fundamentals (log basis and Amihud illiquidity). The whole sample is of daily frequency ranging from January 2, 2003 to December 29, 2015.

Commodity	Total # of News	Data Range	Obs.	StDev.	Min	Max
Panel A: Energy						
CL	994888	2003/01/02 - 2015/12/29	3257	0.0636	-0.2468	0.2184
HO	168937	2003/01/08 - 2015/12/29	3019	0.1343	-0.7825	0.8574
NG	492454	2003/01/02 - 2015/12/29	3257	0.0670	-0.2682	0.2752
RB	165861	2005/12/14 - 2015/12/29	2523	0.0959	-0.3365	0.3422
Panel B: Grains						
BO	559769	2003/01/02 - 2015/12/29	3257	0.0556	-0.6330	0.1941
C-	61972	2003/01/07 - 2015/12/29	2808	0.1666	-0.7501	0.6574
KW	54229	2008/12/05 - 2015/12/29	1308	0.1450	-0.7189	0.6274
MW	54229	2008/12/05 - 2015/12/29	1308	0.1460	-0.7012	0.6223
O-	806511	2003/01/02 - 2015/12/29	3257	0.0522	-0.2221	0.1697
RR	806511	2003/01/02 - 2015/12/29	3257	0.0544	-0.2426	0.1754
S-	60180	2003/02/06 - 2015/12/29	2040	0.1953	-0.8654	0.7206
SM	551181	2003/01/02 - 2015/12/29	3255	0.0617	-0.3280	0.2322
W-	54229	2008/12/05 - 2015/12/29	1308	0.1436	-0.7200	0.6036
Panel C: Livestocks						
FC	465815	2003/01/02 - 2015/12/29	3257	0.0691	-0.3273	0.2344
LC	465815	2003/01/02 - 2015/12/29	3257	0.0678	-0.3210	0.3229
LH	465815	2003/01/02 - 2015/12/29	3257	0.0695	-0.3290	0.2539
Panel D: Metals						
GC	282645	2003/01/02 - 2015/12/29	3257	0.1002	-0.4176	0.3684
HG	35100	2009/12/18 - 2015/12/29	1304	0.1747	-0.7621	0.7534
PA	282645	2003/01/02 - 2015/12/29	3257	0.1050	-0.4134	0.3545
PL	282645	2003/01/02 - 2015/12/29	3257	0.1009	-0.3438	0.3178
SI	282645	2003/01/02 - 2015/12/29	3257	0.1016	-0.4014	0.3273
Panel E: Softs						
CC	88814	2003/01/02 - 2015/12/29	3253	0.1346	-0.5673	0.4840
CT	105907	2003/01/02 - 2015/12/29	3255	0.1043	-0.6056	0.3229
JO	36098	2003/01/02 - 2015/12/29	2943	0.1638	-0.8048	0.7349
KC	107244	2003/01/02 - 2015/12/29	3255	0.1239	-0.5875	0.4784
LB	251274	2003/01/02 - 2015/12/29	3257	0.0841	-0.4554	0.4819
SB	149012	2003/01/02 - 2015/12/29	3255	0.1058	-0.4753	0.3903

*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 meausure and controls. The "connected" sentiment measure is constructed in two steps. In the first step, we obtain each commodity's new sentiment by orthogonalizing the abnormal news tones on its fundamentals (i.e., basis and Amihud illiquidity). Then, we aggregate the sentiment of indexed commodities that belong to other sectors. For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. The control variables include lagged returns, lagged basis, lagged Amihud Illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity. The news tones and controls are of daily frequency ranging from January 2, 2003 to December 29, 2015. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. 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
Cnn. Sentiment	0.0605***	0.0507***		
	(21.47)	(13.21)		
L.Cnn. Sentiment			-0.0052*	-0.0015
			(-1.86)	(-0.41)
L.Return	-0.0121*	0.0722***	-0.0116	0.0721***
	(-1.69)	(7.61)	(-1.60)	(7.53)
L.Basis	0.0039	0.0055	0.0037	0.0048
	(0.61)	(0.40)	(0.58)	(0.35)
L.Illiquidity	1.58e-05***	1.08e-07	1.56e-05***	1.11e-07
	(2.66)	(1.23)	(2.59)	(1.30)
L.ΔOil ImVol	0.0001***	1.06e-05	0.0001***	2.32e-05
	(4.06)	(0.23)	(4.12)	(0.49)
Intercept	-0.0006*	0.0004	0.0010**	0.0004
	(-1.64)	(0.88)	(2.41)	(0.87)
Sector Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Obs.	38,165	19,312	38,149	19,305
# of Individuals	16	8	16	8
Overall R-squared	1.50%	1.69%	0.29%	0.71%

Table 5: 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 individual/sectoral "connected" sentiment meausures and controls. According to NBER's record of expansions and troughs, the sample excludes the period December 2007 to June 2009. The "connected" sentiment measure is constructed in two steps. In the first step, we obtain each commodity's new sentiment by orthogonalizing the abnormal news tones on its fundamentals (i.e., basis and Amihud illiquidity) and take the residuals as the sentiment. Then, we aggregate the sentiment of indexed commodities that belong to other sectors according to the weights based on estimated index trader's open interests (Hamilton and Wu, 2015). For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. The control variables include lagged returns, lagged basis, lagged Amihud Illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity. The news tones and controls are of daily frequency ranging from January 2, 2003 to December 29, 2015. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. 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	Excluding Financial Crisis (2008/09/15 – 2009/06/30)			
	Panel A: Contemporaneous		Panel B: Predictive	
	Indexed	Non-indexed	Indexed	Non-indexed
Cnn. Sentiment	0.0600***	0.0375***		
	(21.92)	(9.98)		
L.Cnn. Sentiment			-0.0072***	-0.0028
			(-2.62)	(-0.76)
L.Return	-0.0047	0.0813***	-0.0032	0.0819***
	(-0.64)	(8.09)	(-0.43)	(8.11)
L.Basis	0.0049	-0.0055	0.0038	-0.0073
	(0.73)	(-0.39)	(0.46)	(-0.51)
L.Illiquidity	1.00e-05**	9.49e-08	9.07e-06*	9.77e-08
	(1.96)	(1.05)	(1.76)	(1.09)
LD.Oil ImVol	6.06e-05*	-2.95e-05	5.18e-05	-2.76e-05
	(1.69)	(-0.56)	(1.45)	(-0.53)
Intercept	-0.0003	-5.31e-06	5.55e-05	0.0015**
	(-0.53)	(-0.01)	(0.10)	(2.05)
Sector Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Obs.	35,465	17,872	35,449	17,865
# of Individuals	16	8	16	8
Overall R-squared	1.47%	1.38%	0.17%	0.82%

Table 6: Spillover Effect of Sentiment on Returns across Indexed/Non-indexed Commodities under High/Low Abnormal Index Exposure

This table presents the results of regressing commodities returns on individual/sectoral "connected" sentiment measures and controls under different levels of abnormal index exposure. The abnormal index exposure is measured as the ratio of index investment flow to the total market cap. Based on this measure, we calculate the median of the weekly average abnormal index exposure. Then, we classify the daily sample within the week whose average exposure is above/below the median into "High/Low" exposure groups. The "connected" sentiment measure is constructed in two steps. In the first step, we obtain each commodity's new sentiment by orthogonalizing the abnormal news tones on its fundamentals (i.e., basis and Amihud illiquidity) and take the residuals as sentiment. Then, we aggregate the sentiment of indexed commodities that belong to other sectors according to estimated index trader's open interests (Hamilton and Wu, 2015). For "connected" sentiment of non-indexed commodities, we take a simple average on the sentiment of non-indexed commodities from other sectors. The control variables include lagged returns, lagged basis, lagged Amihud Illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity. The news tones and controls are of daily frequency ranging from January 2, 2003 to December 29, 2015. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. 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	0.0493***	0.0697***	0.0406***	0.0582***				
	(12.08)	(16.90)	(7.60)	(10.20)				
L. Cnn. Sentiment					-0.0146***	0.0022	-0.0046	0.0008
					(-3.55)	(0.54)	(-0.86)	(0.15)
L.Return	-0.0343***	-0.0220**	0.0431***	0.0838***	-0.0462***	-0.0091	0.0434***	0.0805***
	(-3.28)	(-2.06)	(2.83)	(5.90)	(-4.19)	(-0.84)	(2.96)	(5.69)
L.Basis	0.0017	0.0093	0.0162	0.0104	0.0009	0.0084	0.0170	0.0085
	(0.17)	(1.17)	(0.87)	(0.53)	(0.09)	(1.02)	(0.88)	(0.44)
L.Illiquidity	9.91e-06	2.09e-05***	-1.02e-07	6.23e-07**	7.55e-06	2.20e-05***	-8.41e-08	5.29e-07***
	(1.57)	(3.56)	(-1.52)	(2.27)	(1.15)	(4.44)	(-1.37)	(2.67)
L.ΔOil ImVol	0.0002***	0.0001**	-6.22e-05	0.0001*	6.50e-05	0.0003***	-8.41e-05	0.0002***
	(3.04)	(2.28)	(-1.01)	(1.68)	(1.36)	(4.76)	(-1.23)	(2.62)
Intercept	0.0067***	-0.0045***	0.0021	-0.0007	0.0066**	-0.0033***	0.0034	-0.0006
	(9.07)	(-7.37)	(0.82)	(-1.11)	(2.02)	(-5.53)	(1.46)	(-0.95)
Sector Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	17,433	17,387	8,787	8,741	17,460	17,344	8,806	8,715
# of Individuals	16	16	8	8	16	16	8	8
Overall R-squared	2.11%	2.64%	1.69%	2.77%	1.14%	0.87%	0.79%	1.33%

Table 7: Return Serial Dependence and Commodity Indexing: Interaction Effect

This table presents the results of regressing commodities returns on the interaction between commodities' lagged returns and abnormal index exposure. The abnormal index exposure is measured as the ratio of change in market cap on index trading to the total market cap. In the regression, we control for each commodity's lagged change in returns, lagged change in basis, lagged change in Amihud illiquidity, and lagged change in implied volatility of crude oil options with nearest maturity. The financial crisis period is 2008/09/15 – 2009/06/30. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. 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		Exclude Financial Crisis	
	Indexed	Non-indexed	Indexed	Non-indexed
L.(Abn. Index Exposure × Return)	-4.3122***	-0.8201	-5.7340***	0.2486
	(-2.83)	(-0.35)	(-3.42)	(0.10)
L.Abn. Index Exposure	0.1186***	0.0747*	0.1162***	0.0547
	(3.87)	(1.74)	(3.75)	(1.21)
L.Return	-0.0223***	0.0703***	-0.0129	0.0825***
	(-2.83)	(6.80)	(-1.62)	(7.55)
L.Basis	0.0004	0.0001	0.0013	-0.0129
	(0.05)	(0.01)	(0.18)	(-0.79)
L.Illiquidity	1.78e-05***	9.43e-08	1.12e-05*	7.57e-08
	(3.01)	(1.26)	(1.87)	(0.99)
LD.Oil ImVol	0.0002***	4.78e-05	6.92e-05*	-6.89e-06
	(4.78)	(0.97)	(1.86)	(-0.12)
Intercept	0.0081***	0.0079***	0.0080**	0.0078***
	(15.25)	(2.84)	(2.17)	(2.78)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Obs.	34,789	17,513	32,089	16,073
# of Individuals	16	8	16	8
Overall R-squared	0.45%	0.85%	0.32%	0.98%

Table 8: Return Serial Dependence and Commodity Indexing: Predict Correlation

This table presents the results of regressing commodities serial dependence measure on commodities' abnormal index exposure. The serial dependence measure is defined as $(r_{it}r_{it-1})/2\sigma_i^2$. The abnormal index exposure is measured as the ratio of change in market cap on index trading to the total market cap. In the regression, we control for each commodity's lagged change in returns, lagged change in basis, lagged change in Amihud illiquidity, and lagged change in implied volatility of crude oil options with maturity less than one month. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. The financial crisis period is 2008/09/15 – 2009/06/30. 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		Exclude Financial Crisis	
	Indexed	Non-indexed	Indexed	Non-indexed
L.Abn. Index Exposure	-6.2068***	-1.3423	-4.8122**	1.3344
	(-2.80)	(-0.44)	(-2.44)	(0.52)
L.Serial Dependence	-0.0175	-0.1116**	-0.0072	-0.1200*
	(-0.74)	(-2.13)	(-0.30)	(-1.94)
L.Basis	-0.1601	2.3330**	-0.2128	2.3660**
	(-0.33)	(2.14)	(-0.43)	(2.02)
L.Illiquidity	-0.0016**	-5.96e-06*	-0.0003	-6.67e-06*
	(-2.13)	(-1.71)	(-0.51)	(-1.74)
LD.Oil ImVol	-0.0072**	0.0007	0.0031	0.0057**
	(-2.42)	(0.25)	(1.50)	(2.11)
Intercept	0.2251*	0.0063	0.2798***	0.0067
	(1.87)	(0.09)	(8.13)	(0.15)
Individual Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
# of Obs.	34,789	17,513	32,089	16,073
# of Individuals	16	8	16	8
Overall R-squared	0.34%	1.51%	0.22%	1.87%

Table 9: Serial Dependence and Indexing in Overweighted GSCI/BCOM Commodities

This table presents the results based on relative index overweighting in GSCI/BCOM commodities after regressing a zero-investment trading strategies against GSCI (BCOM) indexing relative to BCOM (GSCI) indexing ("Relative indexing"). Relative indexing is the change in GSCI (BCOM) related ETF/ETN exposure minus BCOM (GSCI) related ETF/ETN exposure, where an index's ETF/ETN exposure is defined as the change in market cap of ETF/ETNs following GSCI/BCOM divided by the estimated GSCI/BCOM market cap. The market cap of each ETF is calculated by multiplying its shares outstanding by its NAV. The GSCI/BCOM market cap is obtained by summing up each indexed commodity's dollar value positions with CIT reported data using Hamilton and Wu (2015). The GSCI/BCOM OW portfolio returns are defined as $R_t^p := \sum_{j=1}^{N} \varpi_{jt}^p r_{jt}$, where $p \in \{GSCI, BCOM\}$ and ϖ_{jt}^p is a weight assigned to return r_{jt} on commodity j based on the overweighting measure

$$OW_{jy(t)}^p = \begin{cases} w_{jy(t)}^p - w_{jy(t)}^q, & \text{if } w_{jy(t)}^p > 0, \\ 0, & \text{if } w_{jy(t)}^p = 0. \end{cases}$$

with $p = GSCI$, $q = BCOM$ or vice versa, and $w_{jy(t)}^p$ being the commodity j's weight in GSCI/BCOM for year $y(t)$ respectively. We then calculate strategy weight $\varpi_{jt}^p = (OW_{jy(t)} - N^{-1}\sum_{j=1}^{N} OW_{jy(t)})r_{t-1}^p$. In the regression for GSCI (BCOM) OW portfolio, we control for the lagged relative (GSCI-BCOM) return, lagged relative realized volatility over the past 250 trading days, the lagged relative log ETF/ETNs' trading volume detrended with one-year average log trading volume, and lagged implied volatility of crude oil options with maturity less than 1 month. The CIT data is of weekly frequency ranging from January 3, 2006 to January 5, 2016. The ETF trading data is of daily frequency starting from June 7, 2006 to December 29, 2015. 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	GSCI OW Portfolio		BCOM OW Portfolio	
	Full Sample	Exclude Financial Crisis	Full Sample	Exclude Financial Crisis
L.Relative ETF Indexing	-8.82e-05*	-0.0001**	-9.80e-05**	-0.0001**
	(-1.79)	(-2.01)	(-2.44)	(-2.50)
L.Relative Return	0.0012	0.0003	-0.0007	-0.0004
	(1.26)	(0.56)	(-1.25)	(-1.08)
L.Relative RVOL	-0.0001	0.0002**	9.21e-06	-0.0002**
	(-1.00)	(2.27)	(0.09)	(-2.35)
L.Relative Volume	-1.15e-06	-2.42e-06	-6.35e-08	1.15e-06
	(-0.38)	(-1.10)	(0.00)	(0.82)
LD.Oil ImVol	-2.50e-06*	-3.42e-07	1.73e-06*	7.73e-08
	(-1.68)	(-0.36)	(1.79)	(0.13)
Intercept	-8.83e-07	-4.46e-06**	1.98e-06	4.10e-06***
	(-0.33)	(-2.36)	(1.02)	(2.74)
# of Obs.	1,930	1,750	1,930	1,750
Adj. R-squared	1.20%	0.17%	1.01%	0.38%