Order Flows and Retail Investor Impacts in Commodity Futures Markets

April 14, 2018

Mark J. Ready and Robert C. Ready*

Abstract: We examine signed order flows and price formation in six commodity futures markets and find that trading in futures markets plays an important role in price discovery. We use these results to investigate the impacts of retail investors in these markets. We examine potential retail order flow from two sources: changes in the positions of commodity index traders and issuances of commodity-linked notes (CLNs). We find strong evidence of order flows and price impacts associated with changes in the positions of commodity index traders. We also find that these effects are concentrated in the minutes just prior to daily futures settlement, when the price impact of trades is generally lowest. We confirm the positive returns around CLN issuance documented by Henderson, Pearson, and Wang (2014), but we find that these returns are too large to be explained by uninformed price pressure from hedging trades in futures markets. We also find no evidence that the CLN-related positive returns are driven by abnormal order flows.

---

*Mark Ready is at the University of Wisconsin-Madison. Robert Ready is at the University of Oregon. We would like to thank participants at the 2017 Pacific Northwest Finance Conference for their helpful comments.

1 Introduction

Increasing retail investment in commodity futures over the last decade has generated substantial interest in the impact of this investment on commodity markets. While theoretical work provides mechanisms through which trading from uninformed retail investors can create price impacts in futures markets, the empirical evidence for these effects is mixed.[1] However most of this empirical work focuses on futures returns and thus ignores a basic question: How large must a trade be to materially impact the futures market?

In this paper we contribute to this discussion by examining the impact of minute-by-minute order flow imbalance on prices for six major commodity futures markets: WTI Crude, Brent Crude, Gold, Copper, Wheat, and Corn. The commodities are respectively the largest and most liquid markets in the four major commodity classes: energy, precious metals, industrials, and agricultural commodities.

We find that signed trading volume in futures markets has a large explanatory power for changes in prices, indicating that these markets play a central role in price discovery. We also find substantial heterogeneity across the trading day. In particular, in the minutes around the determination of the daily futures settlement price, we see large increases in volume and a substantial decrease in the price impact of order flow. This finding is particularly relevant for retail investment vehicles, such as index funds, which track changes in daily settlement prices, as they may be able to trade around these settlements and reduce the impacts of their trades.

To test this hypothesis, we examine the impact of retail order flow coming from weekly changes in the positions of commodity-index traders. The CFTC provides this data only for agricultural futures, so we restrict this analysis to corn and wheat futures. Here we find evidence largely consistent with theoretical models of retail investors. When positions of index traders increase (decrease), we see high levels of buy (sell) volume and positive (negative) returns. Both the volume and return impacts are highly concentrated around the minutes around the daily settlement.

We also use our results to examine the findings of Henderson et al. (2014), who show that issuances of Commodity-Linked Notes (CLNs) are associated with positive price changes on

[1] See for example, Sockin and Xiong (2015), Hamilton and Wu (2015), Baker (2014), and Basak and Pavlova (2016).

[1] These data are only provided for agricultural contracts, and is therefore not available for the

the day the notes are priced. Their primary analysis considers 486 CLNs across various commodities. The average notional value of these notes is approximately $15 million, and they find an associated price increase of approximately 30 basis points on the pricing day of the notes. They attribute these changes to the price impact of the hedging trades made by the issuers of these notes, and suggest that this result supports the theory that uninformed retail trading can have a strong impact on commodity prices.

Our data allow us to examine whether or not the size of these notes is large enough for hedging trades to generate the observed price changes. Here we find that the observed price impacts are too large. We examine the trading activity associated with the largest notes in our six commodity analysis, restricting our analysis to those which have greater than $10 Million of notional, giving us approximately 200 notes with an average notional of approximately $30 Million. Consistent with Henderson et al. (2014) we find a significantly positive daily return on the pricing days for these notes, with an average daily return of approximately 30 basis points. However, our estimates suggest that if the full value of these notes were traded in the minute of settlement, the average impact would be approximately 5 basis points. We also find no evidence that there is any signed trading volume associated with these notes, either throughout the day or in the period around the daily settlement. Moreover, nearly all of the price impact is related to positive returns prior to 12pm on the settlement day. CLN issuers have flexibility to determine the specific date that the notes are priced and issued, so association between issue and return prior to noon opens the possibility that CLN issuers prefer days with rising prices.

1.1 Related Literature

To our knowledge, our paper is the first to systematically examine trade imbalances in commodity futures markets, and thus the first to document price impacts of order flows, as well as to examine the intraday behavior of signed order flow in these markets.

The study of the impact of retail investors on commodity markets is motivated by a growing theoretical literature. Hamilton and Wu (2015), Sockin and Xiong (2015), Baker (2014), and Basak and Pavlova (2016) and others derive theoretical models by which uninformed retail investors can create price impacts in commodity futures markets.

Our work is therefore mostly closely related to the empirical studies of "financialization" in commodity prices. For instance, Irwin and Sanders (2012) examine retail investor flows coming from daily purchases of futures contracts by the United States Oil Fund (USO) and find no impact on the prices or returns of oil futures, and Hamilton and Wu (2015) find no evidence that index-fund investment can predict the returns to commodity futures. In contrast, Henderson, Pearson, and Wang (2014) examine issuances of commodity-linked notes (CLNs) and find evidence of sizeable positive price impacts on the pricing days of these notes. Other work finds evidence for and against the impacts of retail investors including, Tang and Xiong (2012), Silvennoinen and Thorp (2012), Buyuksahin and Robe (2011), Fattouh, Kilian, and Mahadeva (2013), Aliquist and Gervais (2013), and Singleton (2013).

There is a small set of papers which study intraday trading and liquidity in commodity markets. Bessembinder, H., Carrion, A., Tuttle, L., & Venkataraman (2016) study liquidity around the predictable roll of the futures in the United States Oil Fund, and are therefore interested in the calendar spread trades which are distinct from the price level effects we study here. Elder, Miao, and Ramchander (2012) study intraday price patterns in Brent and WTI futures, and Marshall, Nguyen, and Visaltanachoti (2012) study liquidity proxies in commodity prices. However, these papers do not study signed volume and price formation in futures markets.

2 Data

Our data sources include:

- Intraday futures data from Thomson Reuters Tick History from January of 2007 through March of 2014.
- A sample of commodity-linked notes obtained from 424b filings obtained from the SEC's EDGAR database.
- Positions of index traders in corn and wheat futures from the CFTC "Supplementary Positions of Traders" reports

Our data cover six major exchange-traded futures contracts. We include two energy contracts, both the West Texas Intermediate (WTI) contract traded on the NYMEX (now owned by the CME) and the Brent contract traded on the ICE. We also include the gold, corn, wheat, and copper contracts from the CME. In terms of open interest and volume, these contracts are generally largest in their respective commodity classes. Moreover, the gold, corn and wheat contracts on the CME are the dominant futures markets for each commodity. The copper contract on the CME rivals the contract traded on the London Metal Exchange, but generally has slightly lower volume. Nevertheless, even in copper, we find that CME volume plays an important role in price discovery.

Our primary analysis uses 1-minute returns and order imbalance for the nearest-to-maturity high volume contracts in each market. As an illustration for how we construct these measures, we first describe them in detail for the WTI crude oil futures.

2.1 Volume Patterns for WTI

WTI futures contracts are available for every month going out five years and for June and December delivery months going out an additional four years. Unlike index futures, where nearly all of the trading is in the contract with the nearest delivery dates, there is substantial trading and open interest in longer-dated WTI futures contracts. However, most of this trading in the longer-dated contracts is through calendar spread trades, wherein traders agree to simultaneously buy one maturity and sell another. Most of the trading in a single contract is concentrated in the nearer months.

We use data starting in January 2007 and we calculate our imbalance measures using trades and quotes from the Globex platform that are obtained from Thomson Reuters. The NYMEX adopted the CME Globex platform for electronic trading of the WTI contracts in June of 2006 (the CME announced its acquisition of the NYMEX in March of 2008). The Thompson Reuters data include some floor trades over the earlier part of our sample, and evidently includes most or all of the floor trades starting in March of 2013. Starting in March of 2013, the data also include calendar spread trades, but we are able to identify them separately. In order to illustrate the typical pattern in trading volumes, Table 1 shows the WTI contract volumes (in thousands of contracts, each for 1,000 barrels of oil) for the trading days in June 2013.

Table 1: Daily WTI futures volumes for June 2013 (in thousands of contracts)

	July 2013 Contract			August 2013 Contract			All other contracts		
	Globex			Globex			Globex		
Trade Date	Single Month	Cal. Spread	Floor	Single Month	Cal. Spread	Floor	Single Month	Cal. Spread	Floor
20130603	214.2	55.4	2.0	13.6	61.2	4.1	15.9	235.8	17.2
20130604	226.7	56.7	8.4	13.2	58.5	7.5	18.2	269.1	37.3
20130605	189.4	56.7	12.3	11.7	40.5	3.0	13.2	219.7	23.0
20130606	178.4	68.3	5.8	15.3	71.7	3.6	21.8	277.4	20.6
20130607	219.4	75.3	17.8	19.2	76.6	9.4	31.4	366.1	26.4
20130610	124.9	67.9	18.1	14.7	69.6	10.7	12.3	214.5	25.1
20130611	174.0	59.4	6.7	23.5	57.7	5.7	14.7	191.0	20.6
20130612	170.0	53.1	9.2	26.7	71.4	9.3	14.3	177.1	6.2
20130613	144.6	57.7	8.3	38.7	61.6	6.0	18.0	186.5	18.7
20130614	161.8	51.1	14.3	48.8	66.5	5.3	42.4	307.5	34.0
20130617	150.1	71.7	7.0	54.0	78.7	6.7	26.2	186.5	21.1
20130618	81.9	50.6	6.7	65.7	75.5	4.9	15.3	191.9	12.1
20130619	31.7	45.8	11.3	144.8	92.1	4.0	26.9	271.1	15.8
20130620	7.1	13.9	0.1	282.9	81.5	3.3	45.5	343.9	19.0
20130621	-	-	-	267.4	52.6	-	93.6	261.7	-
20130624	-	-	-	223.9	75.5	4.9	39.5	336.2	31.4
20130625	-	-	-	176.4	78.9	5.1	29.5	445.8	43.6
20130626	-	-	-	221.1	59.4	1.7	33.0	255.7	12.2
20130627	-	-	-	188.4	67.5	2.4	33.5	255.3	16.0
20130628	-	-	-	177.4	52.7	1.8	36.1	257.4	18.5

Note that the July 2013 contract last traded on June 20, but most of the trading volume had moved to the August 2013 contract the day before that. The table also shows that calendar spread trading makes up a fairly substantial portion of the front and next month volume, and it constitutes the vast majority of trading in the remaining months. Finally, the table shows that floor trading volume is much smaller than Globex volume, which is a feature common to most futures contracts. In fact, the NYMEX suspended floor trading in WTI futures and many other futures products in July of 2015.

We exclude floor trades because they are executed manually, making it impossible to accurately align them in time with the GLOBEX quotes, and therefore impossible to assign trade direction. We also exclude calendar spread trades from our imbalance measure, motivated in part by results from supplemental tests where we found that the imbalance in the calendar spread trades has little impact on the level of front and next month futures prices.

We classify each Globex single-month trade as a buy or sell by comparing the price to the current quote for that contract, and we aggregate buying and selling volume by minute. We also measure the (logged) return over each minute using quote midpoints as of the end of each minute. We define our futures return as the return on the contract with the highest total volume for that day. Thus, referring back to Table 1, our return data June 18, 2013 use the July 2013 contract and our return data for June 19, 2013 use the August 2013 contract.

Globex trading in WTI futures runs from Sunday night at 6:00 p.m. to Friday night at 5:00 p.m. with one-hour breaks at 5:00 p.m. each day. The bulk of the trading occurs during the day, so we limit our analysis to the time periods from 7:30 a.m. to 4:00 p.m. each day. This time window captures 88% of the total WTI volume in the front and next month contracts.

2.3 Definition of Near Month Imbalance

Most of the trading activity in the contracts that we consider takes place in contracts that have only a few months to expiration. Many users of commodity futures maintain positions in these high volume contracts and roll their positions into later contract months as their contracts near expiration. While this general description applies to all six of our commodities, the specific trading patterns differ.

The WTI and Brent contracts are the easiest to understand. Contracts are available for every calendar month out through 5 years. Trading continues until three business days before the 25th calendar day of the month before the delivery month. As illustrated in Table 1, the nearest contract to expiration, which called the front month contract, has the highest trading volume until a few days prior to expiration. The contract expiring in the next calendar month has the next highest volume across all contracts, and it becomes the highest volume contract as the front month contract nears expiration.

The CME procedures for determining daily settlement prices begin by focusing the contract that generally has the highest volume. This is called the "Active Month" for WTI, gold and copper, and is called the "Lead Month" for corn and wheat. We measure returns using the quote midpoints for the Active/Lead Month contracts. We measure imbalances using the difference between buy and sell volume for trades in all months from the front month through the month that is currently the Active/Lead month or is within three weeks of becoming the Active/Lead month. Although we exclude trades that are part of explicit calendar spreads, we recognize that some traders may roll their position using separate individual trades in the two contract months. Our definition of imbalance effectively nets out any trades that are a result of a trader rolling between contract months. For example, if a WTI trader uses market orders to sell the front month and buy the next month (within three weeks of the front month expiration), our measure will reflect zero net imbalance for those trades. .As a robustness check, we also repeat some of our tests using imbalance based on just the trades in the Active/Lead month.

The Active Month in the WTI futures is the nearest month contract, except for the last two trading days prior to expiration, at which point the next month contract becomes the Active Month. The volume patterns in the other commodities are more complex. Gold futures contracts are available for the nearest three calendar months and for all even calendar months (February, April, June, etc.) for the next two years. Although some trading occurs in odd calendar months that are close to expiration, the volume in odd expiration months is much lower than in the nearby even calendar months. In addition, volume for October tends to be lower that for the other even months. The Active Months in gold are the even months, except for October. The current Active Month is the nearest of these contracts that is not in the final calendar month of trade. For example, on February 1 the April contract becomes the Active Month.. The active months in copper are March, May, July, September and December, and the current active month works the same way it does in gold. So for example, on March 1 the May contract becomes the Active Month.

Corn and wheat futures contracts are available for expirations March, May, July, September and December. Trading occurs through the business day prior to the 15th calendar day of the expiration month. For wheat, each of these months is the Lead month until the 12th business day of the calendar month prior to expiration. For example, on the 12th business day of November, the lead month changes from December to March. Corn is very similar to wheat, except September is never considered the Lead month in corn.

3 The Price Impact of Trade Imbalances

As a first step to understanding the impact of order flows in this market, we follow the Vector Autoregression approach developed in Hasbrouck (1991). Specifically, assume that the (log) quote midpoint for the commodity evolves according to:

$$q_t = m_t + s_t$$

Where m_t is the "efficient price" based on all relevant information, including public announcements and order flow up to time *t*, and the s_t component captures transient market microstructure effects. The efficient price evolves according to:

$$m_t = m_{t-1} + w_t$$

where the increments w_t are mean zero, have variance σ_w^2, and are serially independent at all lags. The s_t process has zero unconditional mean and is jointly covariance stationary with w_t.

We observe the evolution of quote midpoints, $r_t = q_t - q_{t-1}$, and the signed order flow x_t, and following Hasbrouck (1991) we assume these evolve according to the following VAR:

$$r_t = a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \cdots + v_{1,t} \quad (1)$$

$$x_t = c r_{t-1} + c_2 r_{t-2} + \cdots \qquad + d_1 x_{t-1} + d_2 x_{t-2} + \cdots + v_{2,t}$$

In the above VAR, $v_{1,t}$ denotes the impact of public announcements in period t and $v_{2,t}$ denotes the surprise in current period order flow, and these have variances σ_1^2 and σ_2^2, respectively. The assumption that the current period order flow does not depend on the current period public announcement allows the above VAR to be recast in the following VMA representation:

$$r_t = v_{1,t} + \boldsymbol{a}_1^* v_{1,t-1} + \boldsymbol{a}_2^* v_{1,t-2} + \cdots + \boldsymbol{b}_0^* v_{2,t} + \boldsymbol{b}_1^* v_{2,t-1} + \boldsymbol{b}_2^* v_{2,t-2} + \cdots \quad (2)$$

$$x_t = \qquad \boldsymbol{c}_1^* v_{1,t-1} + \boldsymbol{c}_2^* v_{1,t-2} + \cdots + \quad v_{2,t} + \boldsymbol{d}_1^* v_{2,t-1} + \boldsymbol{d}_2^* v_{2,t-2} + \cdots$$

The system in (1) is estimated using OLS, giving the coefficients as well as estimates for σ_1^2 and σ_2^2. Then a Cholesky decomposition recovers the coefficients in (2). This VMA representation allows for the calculation of impulse response functions. Hasbrouck shows that the fraction of the variance of the efficient price innovations w_t that is due to the innovations in the order flow is given by:

$$R_w^2 = (\textstyle\sum_{t=0}^{\infty} \boldsymbol{b}_t^*)^2 \sigma_2^2 / \{(\sum_{t=0}^{\infty} \boldsymbol{b}_t^*)^2 \sigma_2^2 + (1 + \sum_{t=1}^{\infty} \boldsymbol{a}_t^*)^2 \sigma_1^2\}$$

When examining equity data, Hasbrouck applies the approach to trade-by-trade data, although trades within 5 seconds of each other are aggregated into a single observation. In contrast, we aggregate data into one minute time intervals. We measure the quote midpoint returns using the quotes at the end of each interval for either the near month or the next month contract, whichever has the highest volume for that day. As discussed above, we measure imbalance by signing each trade for both the near and next month contracts and totaling the signed buys and sells across both the front and next month contracts. This approach should leave our imbalance measure unaffected by traders rolling from the near month to the next month contract, provided they use a similar mix of market and limit orders for their orders in the two contracts.

As in Hasbrouck (1991), we set the lagged values returns and imbalances to zero at the start of each trading day.

We examine three primary dimensions of liquidity based on the VAR, including:

$\boldsymbol{b}_0$, the initial price impact of the innovation in order flow (higher values suggest either a higher fraction of trades come from the informed or the information held by informed traders is more valuable)

$\sum \boldsymbol{b}^*$, the permanent price impact of an innovation in order flow. We illustrate this with impulse response functions to test if the impact of order flow is reversed in subsequent minutes.

R_w^2, the fraction of the efficient price variance explained by order flow innovations (as with $\boldsymbol{b}_0^*$, a higher value implies more informed trades, but this measure is relative to the amount of information that arrives through public announcements).

3.1 Summary of Near Month Volume Imbalance and Returns

Table 2 shows summary statistics for our six futures contracts. We measure returns in percent, and express both volumes and imbalances as the number of contracts and and as millions of dollars of futures notional. For instance, oil futures contracts are for 1,000 barrels, and the average oil price over our sample was approximately \$100 per barrel. Gold futures contracts are for 100 troy ounces and the average gold price was a bit more than \$1,000 per ounce. Copper futures are for 25,000 pounds and the average copper price was about \$3 per pound. Accordingly, for oil, gold and copper, a single contract roughly corresponds to \$100 thousand notional value (gold notional value a bit higher and copper notional value a bit lower). Corn futures contracts are for 5,000 bushels and the average corn price was about \$5 per bushel, so one contracts corresponds to \$25 thousand of notional value. As the table shows, trade volumes are large and, trade volumes, imbalances, and returns are quite volatile over the period. Average one-minute volume ranges from approximately \$32 Million of notional for WTI to approximately \$3.7 Million of notional for Copper. Average imbalances are near zero, but they are quite volatile with standard deviations of near \$20 Million per minute for gold, Brent, and the WTI, and close to \$3 and \$7 Million per minute for copper and corn respectively.

We are interested in how this imbalance translates into prices, so we first estimate the VAR described in the previous section for each of our five commodities.

Table 2: Summary Data for Near Month Contracts by Minute

The table shows summary statistics (means and standard deviations) for minute-by-minute returns, trading volume, and signed trading volume (imbalance). Volume and imbalance shown both as number of contracts and millions of dollars calculated using the previous minute's price. Left hand panels show statistics for all trading minutes in the sample, while the right hand panels show statistics in only the minute prior to daily futures settlement. The sample is January 1st, 2017 to April 1st 2014.

	All Minutes					Settlement Minute				
	Ret.	**Vol.**	**Imb.**	**Vol.**	**Imb.**	**Ret.**	**Vol.**	**Imb.**	**Vol.**	**Imb.**
	(%)	(# Contracts)		(Mil $)		(%)	(# Contracts)		(Mil $)	
CME WTI Crude Oil										
Mean	0.00	38,852	-2.0	33.8	-0.2	-0.01	104,031	-75.8	90.0	-6.6
St. Dev.	0.09	46,926	172.2	42.7	15.2	0.09	46,964	253.0	45.5	22.2
# of Min					920,396					1,823
ICE Brent Crude Oil										
Mean	0.00	29,211	1.2	28.4	0.1	-0.02	80,073	-17.6	78.4	-2.0
St. Dev.	0.08	44,411	217.9	46.2	21.8	0.16	61,751	388.1	69.0	39.1
# of Min					912,869					1,823
CME Gold										
Mean	0.00	20,654	-1.5	24.8	-0.2	0.00	60,666	18.8	76.1	2.2
St. Dev.	0.05	146,112	125.6	180.2	15.4	0.07	45,604	184.8	65.8	23.3
# of Min					934,832					1,845
CME Copper										
Mean	0.00	4,687	-0.2	3.4	0.0	0.00	31,652	10.9	22.9	0.8
St. Dev.	0.09	36,122	35.9	27.3	2.8	0.12	28,454	119.7	25.2	9.3
# of Min					891,977					1,859
CBOT Corn										
Mean	0.00	30,588	-7.1	8.7	-0.2	0.01	423,144	101.4	116.6	3.7
St. Dev.	0.12	63,963	262.1	17.6	7.3	0.25	298,302	1026.9	93.8	29.5
# of Min					514,404					1,808
CBOT Wheat										
Mean	0.00	13,012	-2.7	4.7	-0.1	-0.04	223,638	-61.5	77.7	-2.0
St. Dev.	0.15	29,807	108.3	9.7	3.9	0.36	171,898	614.5	66.1	22.0
# of Min					487,299					1,808

Table 3: Full Sample Price Impact VARs

The table shows the results from vector autoregressions of the form described in equation (1) in the text. R_w^2 is the percentage of variation in returns explained by unexpected innovations in order flow, calculated from a vector moving average representation of the VAR. Return is measured in percent, while imbalance is measured in 100s of contracts.

	Independent Variables										
	Imb (t)	Imb(t-1)	Imb(t-2)	Imb(t-3)	Ret (t-1)	Ret (t-2)	Ret(t-3)	Cons	Obs	R^2	R^2_w
WTI Crude											
Return	0.032	-0.003	-0.001	-0.001	-0.058	-0.015	-0.007	0.001	939,382	0.4	0.38
	[787.4]	[-62.87]	[-25.12]	[-22.43]	[-56.26]	[-14.19]	[-7.13]	[7.339]			
Imbalance		0.088	0.029	0.029	1.701	0.458	0.124	-0.016		0.03	
		[66.178]	[21.535]	[22.371]	[65.29]	[17.545]	[4.802]	[-9.439]			
Brent Crude											
Return	0.004	-0.001	-0.001	-0.001	-0.023	-0.012	-0.01	-0.001	928,740	0.01	0.01
	[112.7]	[-5.303]	[-4.192]	[-1.671]	[-21.91]	[-10.90]	[-9.21]	[-0.978]			
Imbalance		0.067	0.035	0.025	0.739	0.305	0.203	0.011		0.01	
		[63.973]	[33.173]	[24.078]	[25.750]	[10.625]	[7.081]	[4.788]			
Gold											
Return	0.03	-0.002	-0.001	-0.001	-0.055	-0.021	-0.011	0.000	940,232	0.33	0.31
	[827.8]	[-41.09]	[-19.98]	[-17.97]	[-64.47]	[-24.64]	[-13.2]	[8.025]			
Imbalance		0.068	0.026	0.024	1.531	0.53	0.221	-0.008		0.02	
		[66.23]	[25.68]	[23.60]	[78.01]	[26.97]	[11.27]	[-11.39]			
Copper											
Return	0.098	-0.003	-0.003	-0.002	-0.024	-0.004	-0.006	0.001	945,859	0.18	0.16
	[457.6]	[-13.77]	[-11.86]	[-7.409]	[-24.01]	[-3.819]	[-7.007]	[3.563]			
Imbalance		0.054	0.028	0.023	0.088	0.044	0.008	-0.002		0.01	
		[47.936]	[24.791]	[20.804]	[17.96]	[9.097]	[1.891]	[-4.875]			
Corn											
Return	0.02	0.000	-0.001	-0.001	-0.097	-0.021	-0.012	0.001	544,468	0.24	0.20
	[407.1]	[-2.967]	[-15.70]	[-16.07]	[-74.65]	[-16.67]	[-9.62]	[10.751]			
Imbalance		0.1	0.049	0.036	1.441	0.45	0.117	-0.06		0.03	
		[65.375]	[32.010]	[23.985]	[39.42]	[12.405]	[3.465]	[-17.45]			
Wheat											
Return	0.051	-0.001	-0.002	-0.002	-0.061	-0.019	-0.01	0.001	523,585	0.18	0.15
	[331.6]	[-7.65]	[-10.52]	[-8.92]	[-44.16]	[-12.68]	[-7.20]	[5.37]			
Imbalance		0.087	0.04	0.027	0.314	0.119	0.038	-0.023		0.02	
		[57.15]	[25.34]	[17.36]	[25.65]	[9.11]	[2.98]	[-16.17]			

3.2 Full Sample Price Impact VAR

Table 3 shows the results of the VAR specification in Equation (2) for the full sample. Imbalance is measured in 100s of contracts, and return is expressed in percentage to facilitate interpretation. Again, for most of the commodities, 100 contracts translates into roughly $10 million of notional (with the exception of Corn and Wheat, which translates into approximately $2.5 Million of notional over the sample).

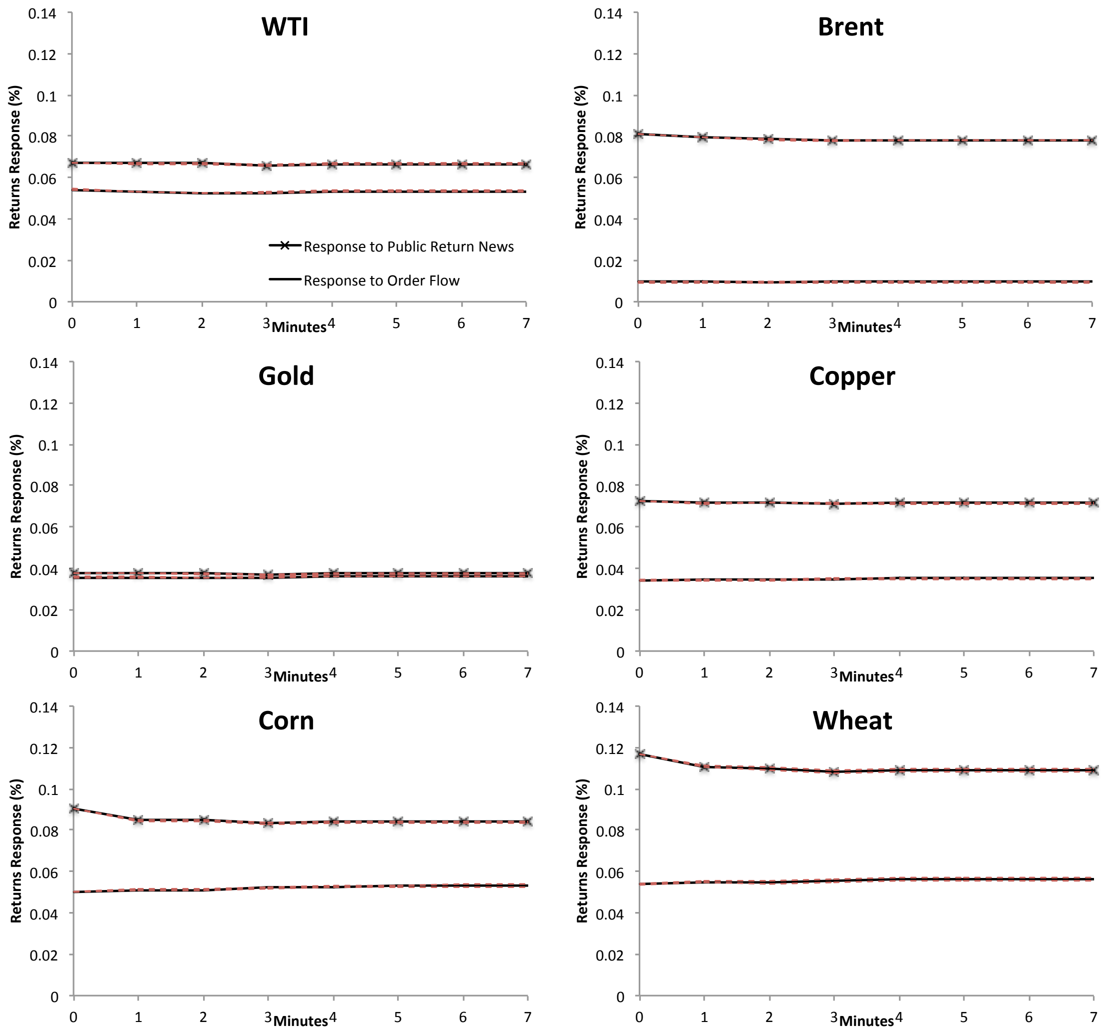
The most important parameter, $\boldsymbol{b}_0$, is shown in the first column of the return rows. The value of 0.032 for WTI futures shows that a minute with 100 contracts of buy (sell) imbalance will create a same-minute price increase (decrease) of 3.22 basis points. The entries in the imbalance column suggest that there is some herding in imbalance, with both lagged imbalance and returns positively predicting positive future imbalance. However, the low R^2 (2.9%) of this imbalance regression suggests that most of the observed imbalance is unpredictable.

This pattern is generally repeated across the commodities but there are differences in the amount of impact. A roughly $10 Million dollar flow yields an impact of approximately 3 basis for gold, similar to the WTI, but a similar size trade will move copper and corn prices approximately 10 basis points. For all four of these commodities, the R^2 of these return regressions is relatively large, and results in a correspondingly high value of R_w^2 from the VMA representation, both results suggesting that order flow imbalance in these markets is playing a major role in price discovery. The one notable exception is the Brent futures. The ICE Brent futures is a highly traded benchmark, but we find that trading in this market has much more limited price impacts.

To ascertain whether or not these price impacts from order flow reverse in subsequent minutes. We use the VMA representation to calculate impulse response functions. These impulse responses are shown in Figure 1. This figure plots impulse response functions for returns in response to a one standard deviation innovation in order flow and in public price news for the six commodities.. The primary takeaway from these plots is that the price impacts of both order flow and public return news are mostly permanent at one-minute horizons. For oil, gold, and copper there is essentially no reversal or continued trend in prices. For corn, wheat, and brent there is a small reversal after a movement in prices unrelated to order flow, but for impacts of order flow we see very little reversal.

Figure 1: Return Impulse Response Functions for VARs

The figure shows the impulse response of returns to innovations in order flow and public news from the vector moving average representation (2) in the text applied to the vector autoregession specification estimated in Table 3. Plots show impulse responses to one standard deviations innovations in public return news and order flow.



3.3 Intraday Patterns in Volume and Trade Impact

The results from the VAR suggest that trade in financial futures markets plays a major role in price discovery. Moreover, we find that the impacts from trade are largely permanent, and do not substantially reverse at 1-minute horizons. However these full sample tests obscure substantial variation in intraday trading patterns, so we now turn to the intraday patterns in volumes and trade impacts. We do this to understand how a sophisticated investor might implement hedging positions associated with a retail investment. Since most retail products are benchmarked to the daily futures settlement price, it is intuitive that the hedging trades would take place near the settlement, which occurs at various times for the different contracts.

To estimate how trading impacts change through the day, we utilize the insight from our VAR analysis that price impacts are mostly permanent, and simply estimate a univariate OLS regression of current minute return on current period imbalance. To facilitate comparison across commodities, we also use imbalance measured in $10s of millions of notional rather than the number of contracts as the independent variable in this regression. We perform this regression for each minute of the trading day. For each contract we consider the interval from 7:30 a.m. through 5:15 p.m. (with the exception of corn and wheat, which stop trading at 2:20 p.m.) New York time, which is a total of 511 minutes. Thus, there are 511 regression estimates, each with approximately 1800 observations (the number of days in the sample). Figure 1 shows the results for these regressions, along with summary statistics for volume, for each of the 6 commodities.

The first panel shows the minute-by-minute average volume and trade impacts throughout the trading day for the WTI futures. The volume rises on the open of pit-trading at 9 AM, and then spikes at times of various announcements, including the EIAs weekly energy outlook published each Wednesday at 10:30 AM. The largest spike however occurs at 2:30 PM in New York when the daily futuressettlement price is set.

The fall in price impact immediately before the WTI settlement suggests these trades have lower information content. The average impact throughout the day is relatively stable around 3 basis points per 100 contracts of imbalance, but this drops drastically in the minutes just around the settlement to roughly 1.6 basis points per 100 contracts.

The implication of this finding is that even for reasonably large trades, say one necessary to hedge a $30 Million CLN, would only have an impact of roughly 5 basis points if traded with a market order in the last minute before settlement.

This pattern is repeated for each of the 6 commodities. For all of the commodities volume spikes and trade impacts fall around the futures settlement, which occurs respectively at 2:30 PM, 1:00 PM, 1:30 PM, and 1:15 PM New York time for Brent, copper, gold, and both corn and wheat respectively. The drop in price impact is most notable for the WTI and gold, but is apparent in all five commodities. While impacts do vary during the day, the high amount of volume at the settlement means that the impacts in these minutes are estimated with high levels of statistical accuracy. Table 4 illustrates this and presents the univariate regressions estimated for the whole sample and for the settlement minute. In all cases, the settlement minute has significantly lower price impact than the full sample estimate.

Table 4: Univariate Price Impact Regressions

The table shows the results from univariate regressions of one-minute returns on one-minute imbalances for each of the five commodities. Return is measured in percentage points and imbalance is measured in $10s of Millions. The left column for each commodity shows the univariate regression run over all minutes in the sample, while the right columns shows the regression run using only the settlement minute for each day. T-statistics are shown in brackets.

	WTI Crude		Brent Crude		Gold	
	Return		Return		Return	
	All Minutes	Settle Minute	All Minutes	Settle Minute	All Minutes	Settle Minute
Imbalance	0.033***	0.016***	0.004***	0.001**	0.022***	0.01***
	[279.61]	[21.81]	[70.11]	[2.29]	[94.18]	[10.71]
Cons	0.00***	0.00	-0.00	-0.02***	0.00***	0.00
	[7.29]	[0.00]	[-1.00]	[-4.27]	[8.04]	[0.04]
Obs	929,814	1,814	919,881	1,809	944,092	1,845
R_sq	0.34	0.23	0.01	0.00	0.28	0.18

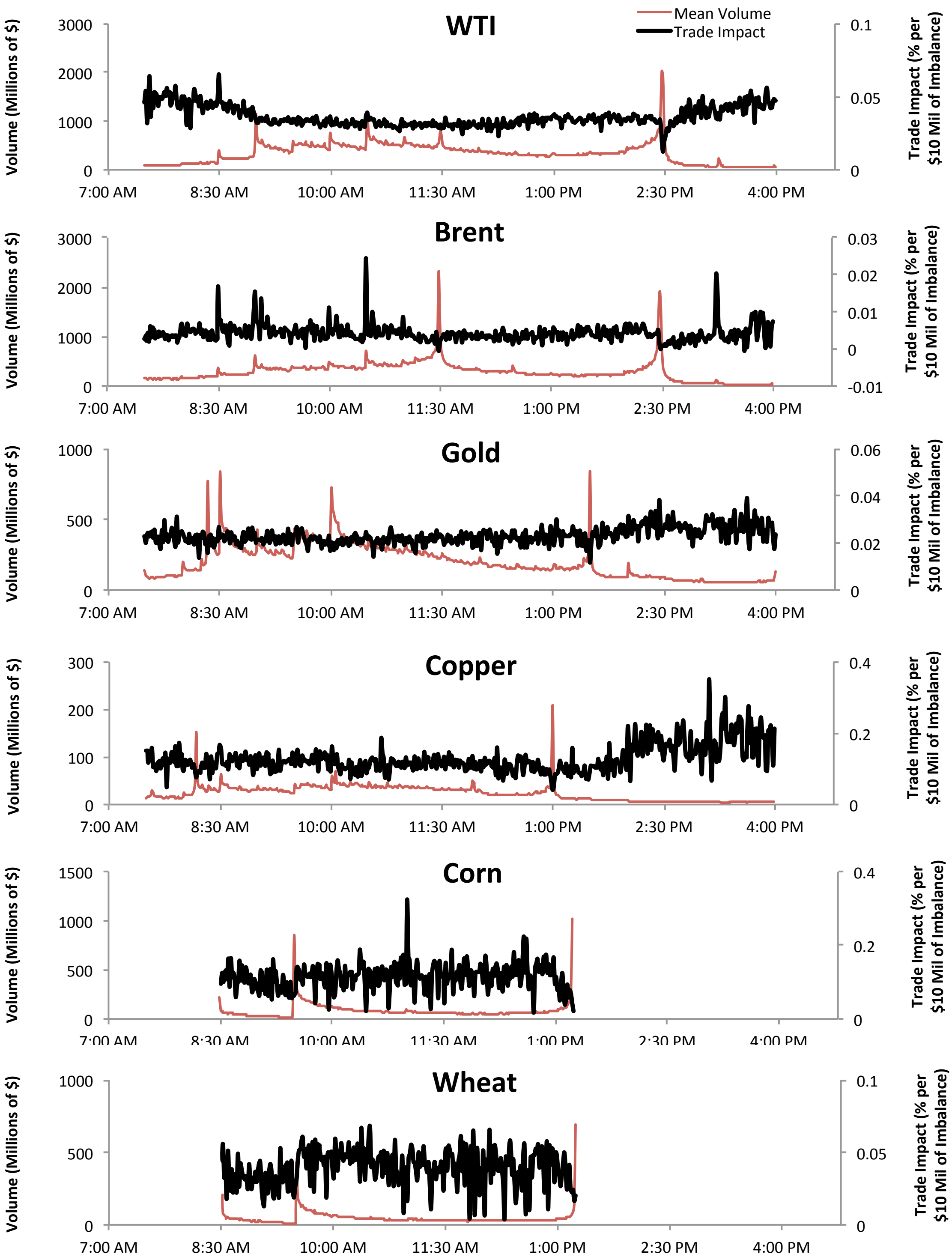
	Copper		Corn		Wheat	
	Return		Return		Return	
	All Minutes	Settle Minute	All Minutes	Settle Minute	All Minutes	Settle Minute
Imbalance	0.12***	0.04***	0.063**	0.042***	0.135***	0.08***
	[45.28]	[12.47]	[33.63]	[17.65]	[33.03]	[11.91]
Cons	0.00***	-0.00	0.00***	-0.00	0.00***	-0.02***
	[4.14]	[-0.72]	[8.32]	[-0.70]	[5.22]	[-2.66]
Obs	902,559	1,859	521,403	1,808	493,952	1,808
R_sq	0.16	0.11	0.17	0.27	0.15	0.25

Robust t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Figure 2: Intraday Volume and Trade Impacts

The figure shows the average intraday volume by minute for each commodity as well as the minute-by-minute trade impact. The trade impact is measured as the slope in a univariate regression of return (%) on trade imbalance ($10s of millions) estimated using each minute's returns for all of the sample days.



4 Retail Investor Flows and Futures Trading

In this section we investigate the market impacts of two sources of retail investor flows. We follow the procedure of Henderson et. al (2014) we collect and process the universe of 424b filings for issuers of CLNs from the SEC's Edgar website. Using data from the CFTC's Supplemental Positions of Traders report, we calculate the change in index-fund holdings for corn and wheat futures.[1] Table 5 presents summary data on these sources of retail flows.

Table 5: Summary Data for Retail Investor Flows

Table shows summary statistics for retail investor flows associated with changes in corn and wheat index holdings and issuance of Commodity-Linked Notes.

Panel A: Weekly Changes in Positions of Commodity Index Traders

	Statistics (Millions of $)			
Commodity	Mean	Std. Dev.	Min	Max
Corn	-6.8	220.3	-958.6	1306.0
Wheat	-7.7	140.7	-1308.6	437.1

Panel B: CLN Notional value

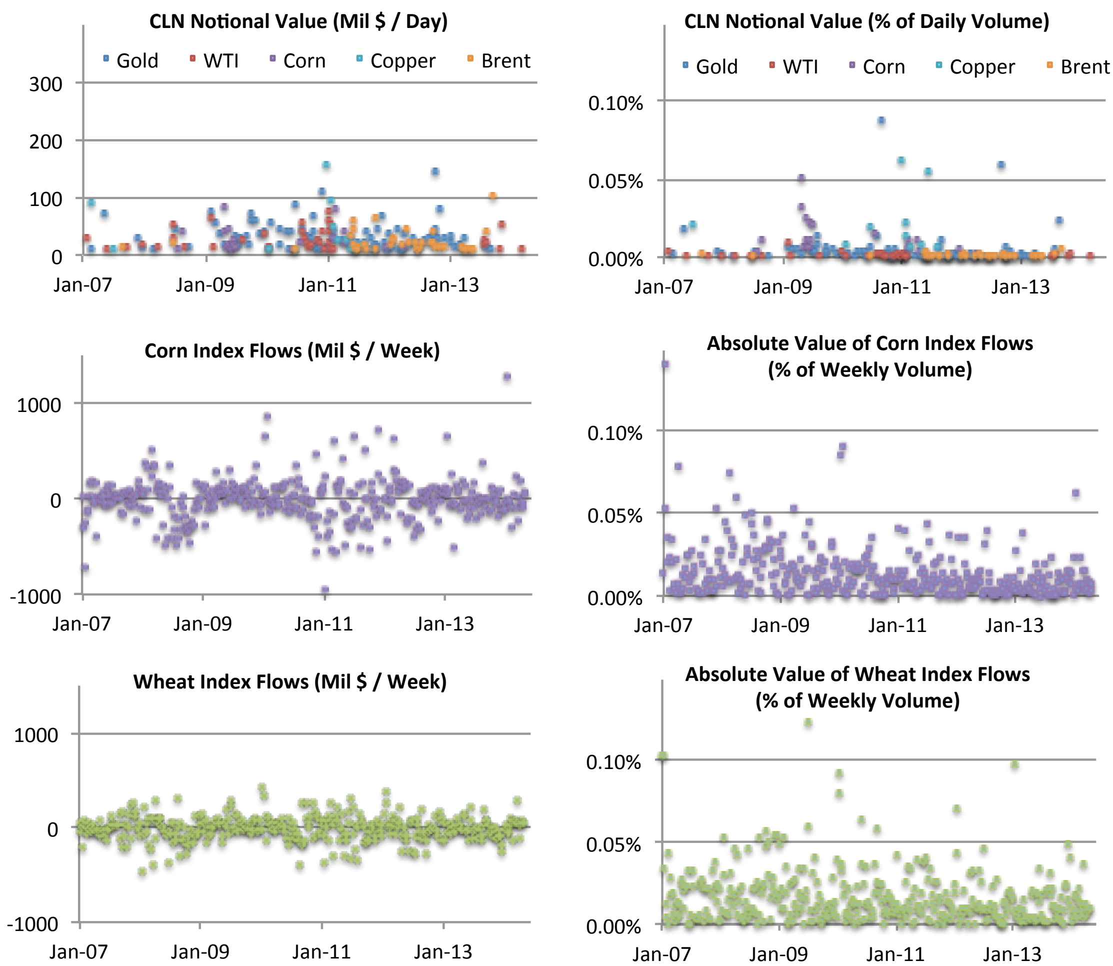
	Number	Statistics (Millions of $)				
Commodity	of Notes	Mean	Std. Dev.	Min	Max
Gold	91	33.0	23.2	10	143.2
Copper	12	42.7	46.9	10	155.5
WTI	39	27.9	17.7	10	75.9
Brent	38	23.7	19.4	10	103.8
Corn	20	28.8	20.8	10.5	81.6

We find 200 notes linked solely to our five commodities with face values of greater than $10 Million, and the average size of these notes is approximately $30 Million. We exclude the issuance of Exchange Traded Notes (ETN), which can be very large (ie.. greater than a $1 Billion). We do this following Henderson et. al, who note that these funds are marketed differently than the linked-notes, and find no price impact around the pricing of these ETNs. Our sample of large notes appears to closely track the set captured by Henderson et al. in terms of number and magnitude.

[1] These data are only provided for agricultural contracts, and is therefore not available for the other four commodities.

Figure 3: Sources of Retail Investor Flow

The figure plots notional values of large commodity-linked notes on their pricing day and weekly changes in the positions of index traders for corn and wheat futures. The left hand plots show these values in millions of dollars. The right hand plots show the notional divided by the futures volume for the associated commodity over the period (daily for CLNs and weekly for changes in index trader positions).



While these 200 notes represent the very largest commodity-linked notes, they are much smaller in magnitude and frequency than weekly changes in index-fund positions from the CFTC data. While this can be seen from Table 5, it is perhaps easier to see visually, so we plot these flows in Figure 3. The left hand side of the figure plots the weekly changes in positions of index traders in the notional values of the CLNs in Millions of dollars. The right hand side of the figure plots the absolute values of the flows to index traders as the fraction of weekly volume, and the CLN notional as a fraction of daily volume in the commodity. The variations in the flows from the wheat and corn index-funds are an order of magnitude larger than those from the CLNs. As we will see however, the patterns associated with the two sources of retail flows are quite different. For the CLNs we see a strong return response but no abnormal order flow, while for index traders we see abnormal order flow in futures, along with a price impact, but it is concentrated around the daily futures settlement.

4.1 Retail Investor Flows, Commodity Returns, and Trade Imbalance

To provide a first look at the potential impacts of these flows, we examine the returns associated with each of the three variables. To do this we estimate the following regressions

$$Imbalance_t = \alpha + \beta\ DailyRetailFlow_t$$

$$Return_t = \alpha + \beta\ DailyRetailFlow_t$$

Here *DailyRetailFlow*$_t$ is the amount of contracts associated with the daily measure of flows. We choose these measures to facilitate economic interpretation. For index fund flows, when performing the regression of imbalance, we regress weekly imbalance on the total change in index trader positions measured in contracts. Therefore, the slope coefficient can be interpreted as the percentage of the change in index trader position reflected in abnormal trade imbalance. For the return regressions, we standardize the index flow so it has a standard deviation of 1, so that the slope can be interpreted as the weekly return impact of a one standard deviation change in index trade positions. For CLNs we follow Henderson et al. a use a dummy which is equal to 1 if there was a CLN issued on that day and 0 otherwise. For the CLN the dependent variables are the daily return or imbalance for the associated commodity, and it is the weekly return or imbalance for the corn or wheat contract for the changes in index-fund positions. We estimate this regression for each source of flow using the daily return and imbalance, as well as the return and imbalance in the settlement minute. Table 6 shows the results.

In Panel A of Table 6 we see that the main result of Henderson et al. is confirmed. The days with CLN issuances are associated with a 27 basis point higher return to the commodity. Intriguingly however, this return does not seem to be associated with abnormally positive trade imbalance, nor does it seem to be at all concentrated around the settlement minute when the trade price of the CLN is being set.

In Panels B and C we repeat this exercise for the flows of index traders. For both wheat and corn we see similar results. When we look at the full week of returns and imbalance, we see a relation to both returns and imbalance, but the relation to returns is quite weak, and not statistically significant at the 5% of level. The relationship with imbalance is stronger and for wheat is quite significant. For corn we see that about 28% of the total weekly change in index trader positions is reflected in imbalance, while for wheat we see approximately 51%.

For both wheat and corn, the results are particularly strong in the minute prior to the daily futures settlement. Here we see that across the week 5% of the total corn flow and 12% of the total wheat flow show up as abnormal order flow in this single minute. This order flow is associated with a significant price impact, with a one standard deviation change in index trader positions associated with a 5 basis point return in this minute for corn, and a 12 basis point return in corn futures.

To visualize these patterns and to help understand if the return impacts in the settlement minute reverse, Figures 4 and 5 plot the regression slopes from expanding windows of cumulative returns and trading imbalance across the trading day for each of the regression specification.

Figure 4 shows the returns associated with retail order flow. The left hand plots show slope estimates where the dependent variable is the cumulative return up to each minute in the trading day. For example, the 10:00 am point on the plot shows the estimated slope and 95% confidence interval for a regression where the dependent variable is the cumulative return (including the overnight return) prior to 10:00 AM for the day of a CLN, or across the week for the index trader regression. For the CLN we only show same trading period (9:15 AM to 1:15 PM) as the corn and wheat index traders. However, the 9:15 AM return captures the overnight and morning return for the commodities in which trading opens prior to 9:15 AM. For the corn and wheat estimates we exclude an 11-month period from May of 2012 to April of 2013 during which the close was moved to 2:00 pm, so the estimates do not perfectly correspond with those in Table 6.

The right hand panels repeat this exercise, but instead focuses on the 30 minutes prior to the daily settlement. Again the plots show regression coefficients for expanding windows. So for the 15 minute point on the plot, the dependent variable is the cumulative return from 30 minutes before to 15 minutes before the daily settle. For the corn and wheat plots this 30 minute period corresponds to 12:45 to 1:15 PM, expected for the 11-month period when the close was moved to 2:00 PM. For the CLN regressions this 30 minute period is for whichever commodity is associated with the current note.

The first two rows of the plot show the results for corn and wheat index traders. We see that the positive return association with increases in index fund traders slowly increases across the day, and then spikes at the closing minute. The confidence intervals show that the result is not statistically significant until the settlement minute is included, and even then the result is only marginally significant. Looking at the right hand side of the top two rows, we see that the relation is much stronger near the settlement minute. Here we see that the positive return relation is already significant across the first half of the 30 minutes prior to settlement, and then strongly significant once the settlement minute is included.

The pattern in the first two rows is very different than what we see in CLNs. Looking at the bottom left panel, we see that the positive return impact in response to the pricing of CLNs is already significant looking at just the overnight return and morning return prior to 9:15 AM, and this return accounts for more than half of the total response. This impact rises slowly through the morning, and by 12:30 PM more than 90% of the 27 basis point daily impact is incorporated in prices. Looking at the right hand panel, we see no impact on returns around the settlement minute. This suggests that the price impact of CLNs is not associated with hedging trades, which would likely take place during the increased period of liquidity around the closing minute.

Figure 5 repeats the exercise but with imbalance as the dependent variable. Here we see patterns in the first two rows similar to those we see in the return plots. Imbalance associated with index trader positions rises through the day, and then spikes before the settlement. In contrast, for CLNs we see no abnormal order flow associated with CLNs, either during the day or near the settlement.

Table 6: Returns and Imbalances around Retail Investor Flows

The table shows regressions of daily returns and imbalance on sources of retail investor flows. Retail investor flow and imbalance are measured in 100s of contracts. Standard errors are clustered by week for the weekly changes in corn index-fund positions. T-statistics are shown in parentheses.

Panel A: Commodity-Linked Notes and Futures

	Daily Ret	Daily Imb	Settlement Minute Ret	Settlement Minute Imb
CLN Day Dummy	0.27***	-3.16	0.01	0.06
	[2.85]	[-0.71]	[1.24]	[0.15]
Cons	-0.01	-4.02***	-0.00**	-0.01
	[-1.47]	[-11.94]	[-2.00]	[-0.35]
Obs	9,215	9,215	9,215	9,215
R-sq	0.00	0.00	0.00	0.00

Panel B: Changes in Corn Index Positions and Corn Futures

	Weekly Ret	Weekly Imb	Settlement Minute Ret	Settlement Minute Imb
Δ Corn Index-fund	0.425*	0.279**	0.0882***	0.0535***
Positions	[1.948]	[2.222]	[3.899]	[5.282]
Constant	-0.23	-103.3***	0.112***	2.687***
	[-2.093]	[-10.41]	[5.062]	[2.852]
Obs	382	382	382	382
R-sq	0.011	0.015	0.041	0.059

Panel C: Changes in Wheat Index Positions and Wheat Futures

	Weekly Ret	Weekly Imb	Settlement Minute Ret	Settlement Minute Imb
Δ Wheat Index-fund	0.465*	0.512***	0.242***	0.125***
Positions	[1.708]	[4.358]	[4.197]	[4.345]
Constant	-0.292	-36.81***	-0.156***	-2.614***
	[-1.288]	[-9.852]	[-3.895]	[-3.815]
Obs	382	382	382	382
R-sq	0.010	0.072	0.086	0.120

Robust t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

4.2 Commodity-linked notes

The finding that the CLNs have the largest associations with daily returns is somewhat at odds with their very small size relative to the other sources of flows. Moreover, it does not appear that these returns are driven by abnormal trade imbalance. To try to get a sense of the magnitude of this effect, we use our estimates of dollar imbalance impact estimated in Table 3, and ask the following question: If futures trades with notional equal to the size of the note were executed entirely in the settlement minute, what impact would we expect to see? Table 7 shows the calculation.

Table 7: Daily Returns and Predicted Daily Returns around CLN Issuance

The table shows the mean notional and daily returns on days with CLN Issuance. The predicted impact is the slope from the settlement minute regression described in Table 3.

Commodity	# of Notes	Mean Notional ($ Mil)	Mean Daily Return (%)	Predicted Impact (%) / $10 Mil	Mean Predicted Impact (%)
Gold	91	33.0	0.206	0.012	0.040
Copper	12	42.7	0.616	0.041	0.175
WTI	39	27.9	-0.175	0.02	0.056
Brent	38	23.7	0.535	0.004	0.009
Corn	20	28.8	0.530	0.049	0.141
Total	200	30.4	0.251		0.055

The table shows the mean notional amount for the notes in each commodity, along with the average daily returns. The table also shows the predicted impact in percentage points per $10 Million of imbalance in the settlement minute from Table 3, and the mean predicted impact calculated using the notional amount of each note. As the table shows, these predicted impacts are considerably smaller than the observed daily returns, with an average predicted impact of 5 basis points relative to the observed 25 basis point average return. Moreover, this estimate of predicted impact should be viewed as an extremely conservative upper bound. A sophisticated trader should be able to put such a trade on with much smaller impacts, as evidenced by the ability of the USO to trade much larger values with minimal impact. While it is possible that the CLN traders are not as sophisticated as the dedicated traders who work for the USO, the CLNs were issued by large financial institutions (ie. Barclays and Goldman Sachs among others) which presumably have access to considerable trading expertise.

Figure 4: Retail Investment Impacts on Returns Across the Trading Day

The figure plots the slope coefficient and 95% confidence interval from regressions where the dependent variables are cumulative futures returns for expanding windows across the trading day, and the independent variables are sources of retail investment. The sources of retail investment are weekly changes in the positions of index traders for corn and wheat (standardized to have a standard deviation of one) for the first two rows of plots, and a dummy variable which takes a value of one if a CLN for a given commodity priced on a given trading day for the last row. The left hand plots shows the entire trading day, where the 9:15 AM return includes the overnight return prior to 9:15 AM for each commodity. The right hand plots show the 30 minutes prior to the daily futures settlement.

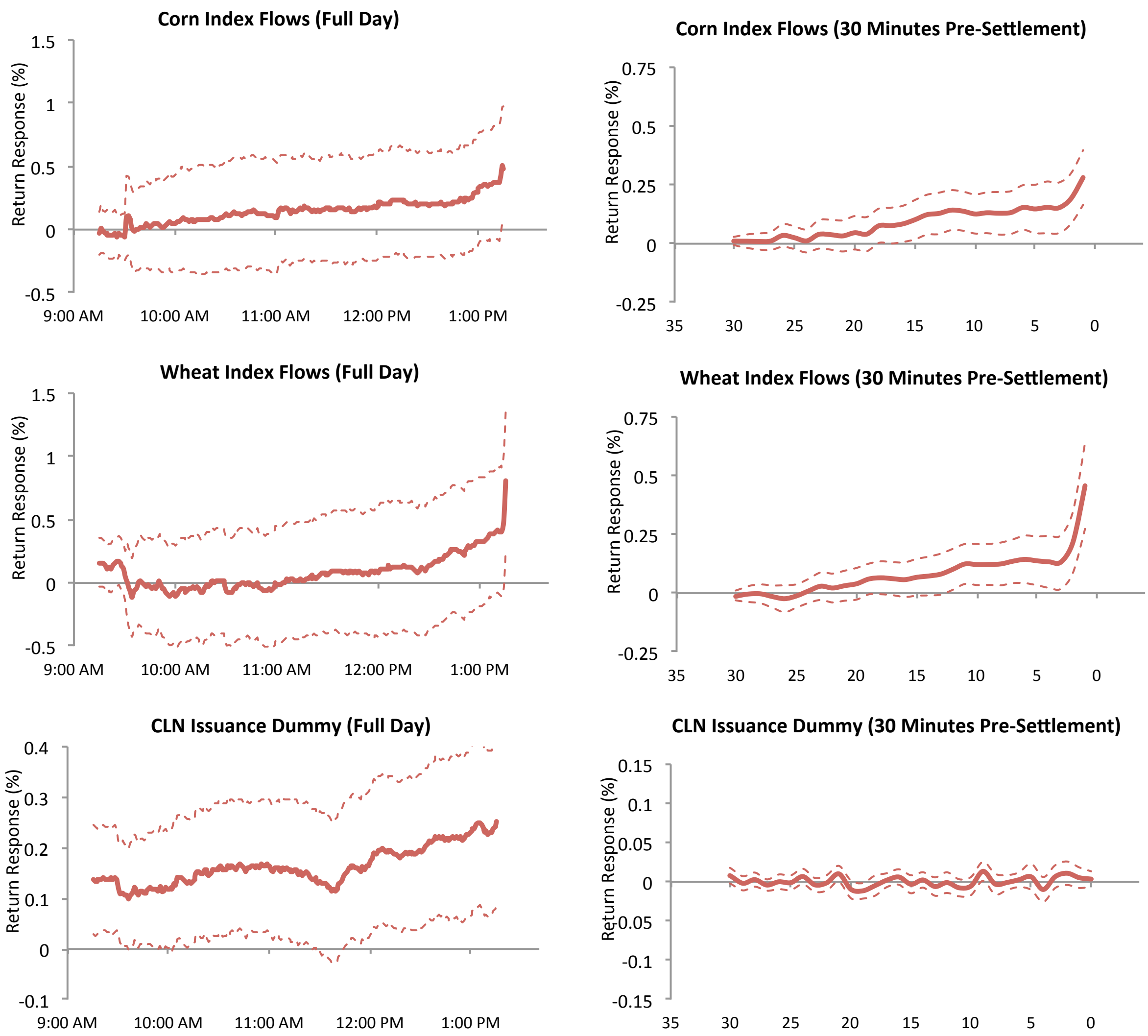
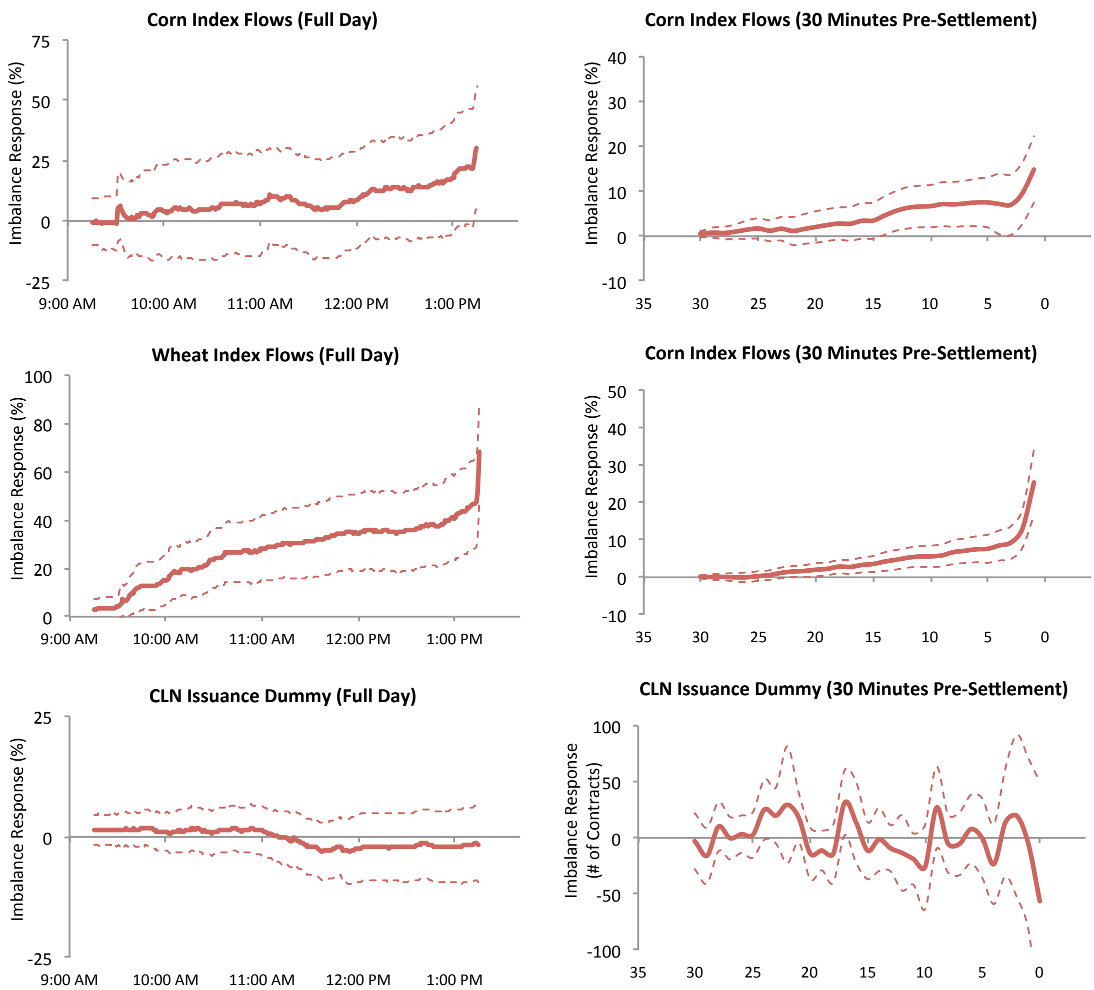


Figure 5: Retail Investment Impacts on Imbalance Across the Trading Day

The figure plots the slope coefficient and 95% confidence interval from regressions where the dependent variables are cumulative futures imbalance for expanding windows across the trading day, and the independent variables are sources of retail investment. The sources of retail investment are weekly changes in the positions of index traders for corn and wheat for the first two rows of plots, and a dummy variable which takes a value of one if a CLN for a given commodity priced on a given trading day for the last row. The left hand plots show the entire trading day, while the right hand plots show the 30 minutes prior to the daily futures settlement.



5 Conclusion

In this paper we construct trade imbalances for six major commodity futures markets. We find that these futures markets play a large role in price discovery. We also document substantial intraday variation in price impacts, with high volumes and low price impact around futures settlements.

We use our findings on trade impacts to examine the potential impacts of retail investors in this market. We find that the positive returns associated with Commodity-Linked Notes are quite large relative to their size, and that these notes are not associated with abnormal trade imbalance. These findings suggest that the positive returns are potentially the result of CLN issuers favouring days with increasing commodity prices, rather than price impacts from associated hedging trades.

We also examine the impact from changing positions in index-fund investment for corn and wheat futures from the CFTC supplemental positions of traders. We find strong evidence for trade imbalances and price impacts associated with these flows, concentrated in the minute prior to the daily futures settlement.

References:

Alquist, Ron, and Olivier Gervais. "The role of financial speculation in driving the price of crude oil." *The Energy Journal* (2013): 35-54.

Basak, Suleyman, and Anna Pavlova. "A model of financialization of commodities." *The Journal of Finance* 71, no. 4 (2016): 1511-1556.

Bessembinder, H., Carrion, A., Tuttle, L., & Venkataraman, K. Predatory or sunshine trading? Evidence from crude oil ETF rolls. *Forthcoming: Journal of Financial Economics* (2016).

Buyuksahin, Bahattin, and Michel Robe. "Does' paper oil'matter? Energy markets' financialization and equity-commodity co-movements." (2011).

Chan, Kalok, Y. Peter Chung, and Wai-Ming Fong. "The informational role of stock and option volume." *Review of Financial Studies* 15.4 (2002): 1049-1075.

Cheng, Ing-Haw, and Wei Xiong. *The financialization of commodity markets*. No. w19642. National Bureau of Economic Research, 2013.

Elder, John, Hong Miao, and Sanjay Ramchander. "Price discovery in crude oil futures." *Energy Economics* 46 (2014): S18-S27.

Fattouh, B., Kilian, L., & Mahadeva, L. (2013). The role of speculation in oil markets: What have we learned so far?. *The Energy Journal*, 7-33.

Hamilton, J. D., & Wu, J. C. (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance*, *42*, 9-37.

Hamilton, James D., and Jing Cynthia Wu. "Effects of Index‐Fund Investing On Commodity Futures Prices." *International economic review* 56.1 (2015): 187-205.

Hasbrouck, Joel. "Measuring the information content of stock trades." *The Journal of Finance* 46.1 (1991): 179-207.

Hasbrouck, Joel. "One security, many markets: Determining the contributions to price discovery." *The journal of Finance* 50.4 (1995): 1175-1199.

Henderson, Brian J., Neil D. Pearson, and Li Wang. "New evidence on the financialization of commodity markets." *Review of Financial Studies* (2014): hhu091.

Irwin, Scott H., and Dwight R. Sanders. "Index funds, financialization, and commodity futures markets." *Applied Economic Perspectives and Policy* 33.1 (2011): 1-31.

Irwin, Scott H., and Dwight R. Sanders. "Testing the Masters Hypothesis in commodity futures markets." *Energy economics* 34.1 (2012): 256-269.

Marshall, Ben R., Nhut H. Nguyen, and Nuttawat Visaltanachoti. "Commodity liquidity measurement and transaction costs." *Review of Financial Studies* 25.2 (2012): 599-638.

Silvennoinen, Annastiina, and Susan Thorp. "Financialization, crisis and commodity correlation dynamics." *Journal of International Financial Markets, Institutions and Money* 24 (2013): 42-65.

Singleton, Kenneth J. "Investor flows and the 2008 boom/bust in oil prices." *Management Science* 60.2 (2013): 300-318.

Sockin, Michael, and Wei Xiong. "Informational frictions and commodity markets." *The Journal of Finance* 70.5 (2015): 2063-2098.

Tang, Ke, and Wei Xiong. "Index investment and the financialization of commodities." *Financial Analysts Journal* 68.5 (2012): 54-74.