Large Traders and Liquidity in Futures Markets

(Job Market Paper)

Stephen Fagan*

Abstract

This paper examines an important and recurring cause of liquidity shocks in futures markets - the accumulation of extreme and opposing positions by hedgers and speculators. Using a combination of market, CFTC, and textual news data, these two classes of traders are found to differ significantly in terms of the impact their trades have on market prices and their response to market-relevant news and past returns. As their positions diverge sufficiently, these differences affect the number and heterogeneity of available counterparties, and can thereby induce periods of increased volatility with high bid-ask spreads and return regularities.

*Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada, V5A 1S6, Email: `sfagan@sfu.ca`.

While market liquidity is a multifaceted concept, it generally refers to the ability to easily buy or sell a security without causing a significant change in the market price. An essential requirement of liquidity is the ample availability of counterparties who are willing to sell when others want to buy and who are willing to buy when others want to sell. Finance theory often assumes perfect market liquidity where market participants can trade any amount of a security without affecting the price. This implicitly requires the unlimited presence of counterparties. While this assumption is clearly false, it is usually a reasonable simplification in large and active markets. There are, however, occasions where, even in the largest and most active markets, liquidity dries up because of the scarcity of counterparties. Essentially, these are times when, for the current market price, everyone who wants to be long is already long, or everyone who wants to be short is already short. When this occurs, someone wanting to trade will be unable to do so unless the market price adjusts to induce someone to trade. While the existence of liquidity shocks has been firmly established, understanding how such occurrences develop is still an important and open question.

This paper presents an empirical microstructural analysis of these liquidity-induced price adjustments in futures markets. The market chosen for this study is that for the New York Mercantile Exchange's light sweet crude oil futures contracts. This is the most liquid market for trading crude oil which is itself the world's most actively traded physical commodity. Because of its liquidity, this market is taken to be very efficient, with the market prices reflecting the value of crude oil. As such, the futures price is used as an international pricing benchmark and an indicator of world energy prices. Despite this overall liquidity, this paper discovers that there are fairly regular occasions where price dynamics, as investigated through returns, realized volatilities and bid-ask spreads, are dominated by the low-liquidity effects resulting from the demographics of available of counterparties. These effects are significant and can persist for up to 15 weeks. This paper identifies the underlying mechanism that causes these effects, and isolates its three main components: (1) significant differences between the dominant classes of traders (i.e. hedgers and speculators), (2) the bounded and mean-reverting nature of trader positions, and (3) the various conditions that cause speculators to enter and exit the market.

Futures markets are designed with many unique features that differentiate them from equity and fixed income markets. One of the most prominent differences is the extensive use of futures markets in risk reduction strategies for those involved in businesses related to the underlying commodity. This feature results in a partition of market participants into two classes of traders: hedgers and speculators. There are large and small traders of both classes, however market activity is dominated large speculators and hedgers. For this study, a trader is considered large if they hold more than 350 futures contracts - the reporting threshold set by the Commodity Futures Trading Commission (CFTC). In the crude oil futures market, large hedgers and speculators hold, on average, 84% of the outstanding futures contracts. While futures markets are anonymous in the sense that information on who is involved in any given trade is not publicly available, the CFTC provides a weekly snapshot of the holdings of these classes of traders.

At any given time, the number of traders in the crude oil market, as well as the aggregated positions held by each class of trader, is bounded. These bounds arise from the number of market participants as well as the constraints faced by these participants. Hedgers participate in futures markets to reduce the price risk that they face in their business activities, and so the number of contracts they hold is determined by the size of their business interests. To use more futures contracts than this bound would be to over-hedge and would increase the price risk they face[1]. Speculators, on the other hand, participate in futures markets

[1]There are many examples of hedgers taking speculative risks. For example, Metallgesellschaft AG, formerly one of Germany's largest industrial conglomerates, lost over 1.4 billion dollars in 1993 through its trading activities in derivatives. Edwards and Canter (1995) and Mello and Parsons (1995) argue that the 1:1 hedge strategy that Metallgesellschaft used was significantly oversized given its underlying oil exposure

to earn profit by accepting additional risk. They too face constraints on their position sizes, but these constraints arise from factors such as available capital, trading strategies, as well as leverage and risk limits. While these differences between hedgers and speculators are well-documented, it is not well understood what, if any, observable market regularities result from these differences. If there are no clear differences between the market impact of these classes of traders, or if there is too much heterogeneity within the classes, then examining opposing positions of hedgers and speculators would not be a useful exercise. This study rejects the hypothesis of no systematic differences between these classes, and establishes reliable differences both in the impact of their trades on market prices and in their response to past market action and market-relevant news.

By examining trade between hedgers and speculators using the net long measure of their positions, large speculators are found to increase prices with their purchases and to engage in trend-following behavior. Large hedgers, on the other hand, decrease prices through their purchases and exhibit contrarian-type behavior. Further, using textual analysis techniques, speculators and hedgers are found to respond differently to market-specific news. These regularities are consistent with the notion that large speculators are informed traders who bring new information to the market through orders that are placed more aggressively than hedgers.

Under liquid market conditions, there are many hedgers and speculators who are not trading near their position limits, and so when a trader wishes to trade, it is highly likely that a counterparty will be available to trade with. Further, the available counterparties on either side of a trade include a mix of both speculators and hedgers. Thus, the systematic differences between the classes of traders effectively gets washed out of the price dynam-

---

and increased its oil price risk instead of reduced it. Despite such examples, the distinction between hedgers and speculators in the oil market is empirically significant as is shown in Section II. This distinction is not always apparent, as is seen in Section IV for the Eurodollar market.

ics. However, when the net long positions of these trader classes significantly diverge, the likelihood of one class of trader dominating the long side of each trade, as the other class dominates the short side, increases. Under such conditions, the differences between trader classes are no longer washed out and begin to significantly impact price movements. Returns, volatilities, and spreads all demonstrate systematic regularities around these times of extreme and opposing positions held by large hedgers and speculators.

Of course, the bounded nature of trader positions prohibits unlimited divergence of trader positions. Consequently, as divergence ends and mean reversion begins, the established concurrent relationship between position accumulation and returns generates sharp changes in price dynamics. The trend following and aggressive order placement of speculators generates persistence in these dynamics for several weeks before and after the divergence of positions reaches its local maximum. For example, when speculators have acquired extremely short net positions and hedgers have acquired extremely long net positions, the price tends to follow a v-shaped reversal pattern, the bid-ask spreads are higher than normal, and the average realized volatility at its 79th percentile. Such liquidity effects are found to be largely determined by the entry and exit of large speculators from the market. In particular, speculators facing increasing losses exit the market using orders that are significantly more aggressive than typical orders, thereby pushing up return volatility. The notion of directional realized volatility is introduced to capture the difference in contributions to volatility from periods of positive and negative returns. Disaggregating the realized volatility in this manner permits the identification of a significant asymmetry in the cause of high volatility before and after these liquidity shocks.

Within the literature, the study of liquidity has focused mainly on equity (Hasbrouck and Sofianos, 1993; Chordia *et al.*, 2001) and fixed income markets (Chordia *et al.*, 2005; Huang *et al.*, 2001; Brandt and Kavajecz, 2002). Trader positions have been related to returns (Sanders *et al.*, 2004) and volatility (Wang, 2003) in futures markets, however, this paper extends the methodology of these studies in three important ways. First, this paper studies futures market microstructure, showing that unique features of these markets can have substantial impacts on market liquidity. Second, the notion that trader positions are bounded is captured by transforming the net long measure into an index that expresses the current position in relation to recent maximum and minimum values. Finally, rather than examining the overall influence of trader positions, this paper focuses on what happens when the position index approaches its bounds.

The paper is organized as follows. Section I discusses the market and CFTC data. Section II examines the empirical regularities that differentiate hedgers from speculators. Section III analyzes ex post market dynamics around extreme holdings of these two classes of traders. Results of robustness checks are described in Section IV and a brief summary concludes the paper in Section V.

I. The Market and CFTC Data

The market under study in this paper is the New York Mercantile Exchange's Light Sweet Crude Oil futures market. This futures contract is the most liquid instrument for trading crude oil, which is the world's most actively traded physical commodity. Approximately $500,000$ contracts are traded per day in this market, with 1.5 million contracts in open interest. This level of daily trading volume represents almost seven times the daily world production of crude oil[2]. The light sweet variety of crude oil is popular among refineries because of its low sulfur content and its high yield of gasoline, diesel, and other petroleum products. This market is chosen because of its high overall liquidity, as well as the fact that

[2] Data on trading activity is from the New York Mercantile Exchange, and data on oil production is from the Energy Information Administration.

the speculator-hedger distinction is not as fuzzy as in other markets such as financial futures markets where contracts can be used as a speculative and hedging instrument by the same fund manager.

Data from three sources are integrated for the analysis: futures market trade data[3], textual news articles on the crude oil market[4], and Commitments of Traders Report (CoT) data from the Commodity Futures Trading Commission (CFTC). The following sections describe these datasets and the construction of the variables of interest.

I.A. Futures Market Data

Futures price data for the NYMEX light sweet crude oil market is collected for the same period for which weekly CoT data is available, namely September 30, 1992 until February 28, 2006 (698 weeks). The data is sampled at the tick level ($5,487,792$ observations) and aggregated to the scales required for variable construction. Since there is trade in several contracts, each with different expiration dates and market prices, a sensible method of constructing a single price series over this period is needed. Since the front contract - the contract with the closest expiration date - usually has the highest trade volume and open interest, the constructed price series is composed almost entirely of prices from the front contract. As the front contract approaches expiration, most traders begin to close out positions in the front contract and enter position in the first-back-month contract. On the first day that the daily number of price ticks, a proxy for the volume of trade, of the first-back-month exceeds that of the front contract, the constructed price series uses the prices from the first-back-month contract from that point onwards. The standard backward-adjustment process corrects the series for price jumps that are only due to the switching of contracts.

[3]The data provider for the crude oil futures market tick data is TickData, www.tickdata.com.
[4]The data provider for the news articles is Dow Jones Factiva, www.factiva.com.

The returns analyzed in this paper are weekly returns, calculated from prices on the close of each Tuesday - the 'as of' date of the CoT report which is released three days later on Friday. In addition to returns, realized volatility (Dacorogna *et al.*, 2001; Andersen and Bollerslev, 1998) is analyzed as a proxy for the true volatility of this return series. Realized volatility is sampled on a weekly basis and constructed from 5-minute intraday return data. Let there be h five-minute intervals during which the market is open in one week. The current day trading hours for NYMEX crude oil futures is from 10:00 until 14:30 EST. Thus, one trading day has 54 five-minute intervals, and one trading week has $h = 270$ five-minute intervals. Then realized volatility is calculated as

$$RV_t = \sum_{i=1}^{h} \left(r^{(h)}_{t-1+i(\frac{1}{h})} \right)^2 \tag{1}$$

where $r_t^{(h)}$ is the five-minute return at time t. Figure 1 presents plots of the price series, the weekly returns, and the realized volatility.

[FIGURE 1 ABOUT HERE]

I.B. Directional Realized Volatility

Realized volatility is a measure of the size of price movements, but it says nothing of the direction of those movements. Yet, the ups and downs of prices might not equally contribute to the return volatility in any given period. During a swiftly rising market, one might expect that the numerous buyers would make selling easy, but buying may be difficult in the sense that it would induce significant slippage. Similarly, during a market crash, selling would be difficult, but buying would be easy. To investigate the notion that liquidity shocks may be asymmetric in this sense, the concept of directional realized volatility is introduced. Directional realized volatility disaggregates the realized volatility into a component that measures the contribution of positive returns to the realized volatility, denoted DRV_t^P, and another component that measures the contribution of negative returns, denoted DRV_t^N. The two components, which are additive variance-preserving transformations of realized volatility, are defined as follows:

$$DRV_t^P = \sum_{i=1}^{h} \left[I\left(r_{t-1+i(1/h)}^{(h)} \geq 0 \right) \cdot r_{t-1+i(1/h)}^{(h)} \right]^2 \tag{2}$$

$$DRV_t^N = \sum_{i=1}^{h} \left[I\left(r_{t-1+i(1/h)}^{(h)} < 0 \right) \cdot r_{t-1+i(1/h)}^{(h)} \right]^2 \tag{3}$$

where $r_{t-1+i(1/h)}^{(h)}$ is the five-minute return as defined above and $I(\cdot)$ is an indicator function taking the value 1 if the condition in the brackets is satisfied, and 0 otherwise. The positive (negative) directional realized volatility sums the squared five-minute returns that are positive (negative).

I.C. Estimating the Bid-Ask Spread

While returns and realized volatility are informative when investigating liquidity dynamics, there is a growing literature that has developed many liquidity measures to quantify specific liquidity related factors. One of the most popular liquidity measures is the bid-ask spread (Roll, 1984) which is often used as a liquidity benchmark for evaluating other liquidity measures (Goyenko *et al.*, 2008). As liquidity dries up, trade can only be induced by offering better prices to potential buyers and sellers resulting in a larger bid-ask spread. Conversely, when there are many willing buyers and sellers, the bid-ask spread will be smaller.

A spread proxy can be constructed from transaction data using the tick test (Lee and Ready, 1991) to partition the transactions into those likely to have originated from market buy orders and those likely to have originated from market sell orders. A transaction is identified as one likely to have originated as a market buy (sell) order if the previous transaction price was below (above) the current transaction price. If the previous transaction price was the same as the current transaction price, then the previous distinct transaction price is used instead. A critical assumption here is that transactions occur only at the bid or ask prices, thus at each transaction price is either an observation of the bid or of the ask price. If the bid (ask) is known at time t, to get a proxy of the spread at that time, we must estimate the ask (bid) price at time t. This estimate is taken to be the most recently identified ask (bid) price constrained to be at least one tick above the observed bid price. Figure 2 shows a sample of the bid and ask estimates given the transaction prices.

[**FIGURE 2 ABOUT HERE**]

This procedure gives an estimated series of bid-ask spreads at the same high frequency as transactions. As pointed out by Lee and Ready (1991), the tick test becomes less precise when the time between transactions increases. Given the high level of activity in the crude oil futures markets, this issue is much less significant then in other less active markets.

With these estimates of the bid-ask spreads, the percentage spread for each transaction is defined to be the ratio of the estimated bid-ask spread to the midpoint between the bid and the ask. These percentage spreads are averaged over each day, t, giving,

$$S_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{Spread_{t,i}}{Midpoint_{t,i}} \tag{4}$$

where N_t is the number of transactions on day t, $Spread_{t,i}$ is the ith estimated bid-ask spread, and $Midpoint_{t,i}$ is the midpoint between the ith estimated bid and ask.

I.D. Textual News Data

To examine potential differences between hedgers and speculators, their response to news articles related to the crude oil markets is analyzed. Since the electronic news database is relatively sparse before 2001, a more recent sample period, between July 25, 2001 and April 28, 2009, is used for this analysis. Approximately $58,000$ news articles related to the crude oil markets were extracted from the Dow Jones Factiva database. These articles were all from the publications used by the Energy Information Administration (EIA) to form their Annual Oil Market Chronology[5] which lists all of the significant world events that have impacted the crude oil markets. The specific uses of this data are detailed in Section II.

I.E. CFTC Trader Data

The Commodity Futures Trading Commission (CFTC) has been releasing weekly Commitments of Traders (CoT) Reports[6] since September 30, 1992, and at various lower frequencies since 1924. The earliest predecessor of the CoT reports were prepared by the U.S. Department of Agriculture's Grain Futures Administration and were released yearly. Currently, these reports provide a breakdown of the open interest in the American futures markets as of each Tuesday. Open interest refers to the total number of futures contracts that have been entered into and not yet exited through a transaction or delivery. One of the main reasons for collecting this data is to detect and deter attempts at market manipulation (such as the cornering of markets) by large traders. More generally, the CoT reporting system allows the CFTC staff to identify large positions that could pose a threat to orderly trading.

In each CoT report, total open interest is broken down according to the type of trader and by the type of positions (long or short) that each type of trader holds. There are three

[5] http://www.eia.doe.gov/cabs/AOMC/Sources.html

[6] The CoT Reports are published on the CFTC website at http://www.cftc.gov/cftc/cftccotreports.htm.

types of traders identified by the CFTC: reporting commercial, reporting non-commercial, and non-reporting. A trader is classified as a reporting trader if they hold positions in excess of the CFTC-determined threshold of 350 contracts. This threshold varies across time and markets in an effort to capture between 70 and 90 percent of the open interest in the market. While the reporting threshold is quite high, the average reporting traders hold much larger positions. For example, in July of 2007, the average reporting trader held over 9000 contracts. A commercial trader is one who self-identifies as being engaged in business activities hedged by use of the futures and options markets. The non-reportable positions are all positions that are held by traders whose total position size is below the CFTC's threshold. In what follows, reporting commercial traders are considered large hedgers, reporting non-commercial traders are considered large speculators, and non-reporting traders are considered small traders who may be hedging or speculating.

Thus, the disaggregation of open interest performed by the CFTC occurs along three dimensions: reporting versus non-reporting, hedging versus speculating, and long versus short. For noncommercials, the CoT report also identifies how much of their position is held in a spread. The spread number measures the extent to which a speculator holds equal long and short positions. For example, if a speculator was long 100 contracts and short 70 contracts of a different expiration, then the spread contribution will be 70 and the long contribution will be 30. The seven components of the total open interest (TOI) that are detailed in the CoT reports are related in the following manner:

$$\overbrace{\overbrace{[S_{Long} + S_{Short} + 2(S_{Spread})]}^{speculators} + \overbrace{[H_{Long} + H_{Short}]}^{hedgers}}^{reporting} + \overbrace{[NR_{Long} + NR_{Short}]}^{nonreporting} = 2(TOI) \quad (5)$$

where S, H and NR refer to positions, measured in the number of contracts, held by large speculators, large hedgers and non-reporting traders, respectively, with subscripts (long, short, and spread) indicating the type of positions. On the left side of equation (5), each contract is counted twice since both long and short positions are counted even though a long and a short position constitute only one contract. Consequently, the total open interest is doubled on the right side of the equation.

As can be seen from the sample CoT report[7] in Figure 3, the report also lists the number of traders in each category. Note that the sum of the numbers of traders from each category exceeds the total number of traders due to the fact that a single trader, such as a spreading speculator, can be counted in more than one category. A trader is counted in each category in which the trader holds a position. For example, a hedger who is long December 2008 crude oil while also being short June 2009 crude oil will be counted only once for the number of total traders, but will be counted in both the 'commercial long' and the 'commercial short' categories. Also of note is that the CoT is available in two versions: a futures-only version, and a futures and options version that converts options positions into equivalent number of futures positions. Since the current study is interested in the microstructure of futures markets, the futures-only version is used.

[**FIGURE 3 ABOUT HERE**]

Various measures of trader positions are constructed to illuminate the relationships between trader classes and market dynamics. These measures are defined as they are introduced in the next two sections.

II. Hedgers and Speculators

The large majority of open interest is held by large hedgers and speculators. In the sample under study, an average of 84% of open interest is held by large traders that must report

[7]See CFTC (2007) for a history of the CoT reports and a full description of CoT variables.

their holdings to the CFTC. Of the reportable holdings, large hedgers dominate the large speculators by holding, on average, 68% of the total open interest compared with 16% for the speculators. Equation (6) demonstrates the calculation of the percent of open interest (POI) held by large speculators. The POI for hedgers is calculated similarly, but without the spread component.

$$POI_t^S = \frac{S_{Long,t} + S_{Short,t} + 2(S_{Spread,t})}{2(TOI_t)} \qquad POI_t^H = \frac{H_{Long,t} + H_{Short,t}}{2(TOI_t)} \tag{6}$$

where S refers to positions held by large speculators with the subscripts (long, short, and spread) indicating the type of positions. From the time series of POI presented in Figure 4, the larger size of the positions of hedgers relative to speculators is clear - hedgers hold between 60% and 80% of the outstanding positions, speculators hold between 5% and 35% of the outstanding positions, and small traders hold the remaining positions.

[**FIGURE 4 ABOUT HERE**]

Another feature of POI evident in Figure 4 is the inverse relationship between movements in positions of hedgers and speculators. Within the sample, the percent of open interest held by large speculators and hedgers, POI_t^S and POI_t^H, have a correlation coefficient of -0.40, indicating that when one trader is entering a position, there is a tendency for a trader of another class to be closing a position. This observation is consistent with the proposition that different trader types provide each other with liquidity.

The main variable of interest in this study is the net long measure of trader positions, denoted NL_t^C for trader class C. The net long measures for speculators and hedgers are calculated by

$$NL_t^S = S_{Long,t} - S_{Short,t} \quad \text{and} \quad NL_t^H = H_{Long,t} - H_{Short,t}. \tag{7}$$

These measures are indicate in the number of contracts held as long positions that are not offset by short positions. A negative net long position indicates that, within the trader class, more contracts are held in short positions than in long positions. A key feature of this measure is that it emphasizes trade between hedgers and speculators since trade within each class cannot change its net long position.

Summary sample statistics for the net long positions by trader types are presented in Table I. The positive mean, median, and skewness of the net long positions of speculators indicates a tendency of speculators to go long, while hedgers tend to go short. This observation reinforces the proposition that multiple trader types increase market liquidity since if speculators were not in the market, hedgers would have a harder time finding counterparties for their short positions. A second observation can be made from the sample kurtosis measures, both of which indicate thinner than Gaussian tails. If trader positions are bounded due to some budgetary or risk constraints, the distribution of their positions would be expected to be finite or, at most, thin.

[TABLE I ABOUT HERE]

As can be seen in Figure 5, the net long position of large hedgers and speculators tend to move in opposing directions. The strong sample correlation of -0.96 is not unexpected since a net change for one trader type requires another trader type to take an opposing position. A measure of -1 is not seen because of a small amount of trade between large and small traders. What is interesting, however, is that the relationship between small traders and large traders is significantly weaker, with correlations of $\rho(NL^S, NL^{small}) = 0.543$ and $\rho(NL^H, NL^{small}) = -0.755$. Since the small trader class is a mix of small speculators and hedgers, the behaviorial regularities associated with trader classes are weaker, though the signs of the correlations indicate that a position taken by small traders is more likely a speculation than a hedge.

[FIGURE 5 ABOUT HERE]

II.A. Trader Positions and Market Prices

It is well-understood that hedgers and speculators face different constraints and participate in futures markets for different purposes. However, it is unclear how these differences impact price dynamics in these markets. This section tests the null hypothesis that, despite their differences, hedgers and speculators are similar in their impact on prices through their trades, as well as in their response to market action. This hypothesis is rejected and several regularities are found that differentiate hedgers and speculators.

To evaluate any potential differences in the relationship between trader positions and market prices for hedgers and speculators, three models are estimated:

$$NL_t^S = \alpha^1 + \beta_1^1 NL_{t-1}^S + \beta_2^1 r_t + \beta_3^1 r_{t-1} + \varepsilon_t^1 \tag{8}$$

$$NL_t^H = \alpha^2 + \beta_1^2 NL_{t-1}^H + \beta_2^2 r_t + \beta_3^2 r_{t-1} + \varepsilon_t^2 \tag{9}$$

$$r_t = \alpha^3 + \beta_1^3 r_{t-1} + \beta_2^3 NL_t^S + \beta_3^3 NL_{t-1}^S + \varepsilon_t^3 \tag{10}$$

where the error terms are assumed to be conditionally heteroskedastic. In order to account for potential endogeneity of returns and positions, continuous updating efficient generalized method of moments (Hansen *et al.*, 1996) was used to estimate the models. Changes in the number of traders holding long and short positions in each trader class were used as instruments for returns while levels of these variables were used as instruments for positions. The results, presented in Table II, indicate several significant differences between hedgers and speculators. First, by examining the concurrent relationship between trader positions and returns, speculators are found to increase prices with their purchases whereas hedgers tend to decrease prices through their purchases. This difference is consistent with the notion that speculators place more aggressive orders than hedgers. Hedgers may employ less aggressive limit orders in order to reduce their costs of hedging.

[TABLE II ABOUT HERE]

Another difference is found in the relationship between trader positions and lagged returns. The significant positive coefficient of lagged returns in the model of speculative net long positions indicates a tendency for speculators to engage in trend-following behavior. Large hedgers, on the other hand, have a tendency towards following contrarian strategies.

II.A. Trader Positions and Market News

A second area in which hedgers and speculators may differ is in their response to news about the crude oil market. In the more traditional equity and bond markets, one might expect that most participants would respond similarly to positive or negative news. However, this may not hold in futures markets where hedgers are not primarily seeking to profit from price changes. To test the relationship between trader positions and market news, this section employs two methods of textual analysis: (i) a reduced dimensionality regression, and (ii) a high dimensionality binary classification.

These quantitative methods are chosen rather than an event-by-event chronological analysis of relevant world events for several reasons. While these methods do not capture the subtleties of complex political and economic news stories, they are objective, consistent, and reproducible methods that have been successfully used in other settings. Additionally, the regularities studied in the remainder of this paper have strong persistence over time, with some lasting more than 15 weeks. These long periods contain many events that are relevant to oil markets, and isolating specific effects would be difficult. Finally, given that many observations are collected for this study, we can abstract from the peculiarities of a single event and identify the underlying microstructural issues.

The first technique is similar to that used by Tetlock *et al.* (2008) to reduce the typically high-dimensional character of textual data. They create a measure of news sentiment by identifying the fraction of positive and negative words in market-specific news stories. Positive and negative words are identified using the Harvard-IV-4 classification dictionary. Using the a traditional regression methodology, they find that the information captured by this simple variable has predictive power for earnings and equity returns. For the current study, approximately $58,000$ crude oil related news stories were used to construct following variables,

$$Neg_t = \frac{\#\text{ of negative words in week } t}{\#\text{ of total words in week } t} \qquad \text{and} \qquad neg_t = g\left(\frac{Neg_t - \mu_t^{Neg}}{\sigma_t^{Neg}}\right)$$

where μ_t^{Neg} and σ_t^{Neg} are the mean and standard deviation of Neg for the six months prior to time t, and where g is an exponential moving average[8] with a smoothing parameter of 0.9. Similarly, Pos_t and pos_t are created for positive words.

Using these proxies for the news content, the following model is estimated.

$$NL_t^S = \alpha + \beta_1 neg_t + \beta_2 pos_t + \beta_3 NL_{t-1}^S + \varepsilon_t \tag{11}$$

where ε_t is assumed to be a white noise process. Results for this estimation are listed in Table III. Tetlock *et al.* (2008) found that for equity markets the negative words contained in news articles contained significant information about firm earnings, beyond what analysts' forecasts and historical accounting data did. Interestingly, for crude oil markets, it is often positive news that will put downward pressure on prices, while bad news drives prices up. After accounting for this difference, similar results are found in crude oil futures markets as speculators significantly decrease their net long positions in the presence of high levels of positive words. Hedgers on the other hand increase their net long positions during such periods.

[8]The results of this section is robust to a wide variety of smoothing operators and parameters.

[TABLE III ABOUT HERE]

Given its large number of words and linguistic relationships, textual data is naturally very highly dimensional. This high dimensionality cannot be maintained in most econometric techniques, however there are some more recent classification and regression frameworks which can accommodate such high dimensionality. The most successful technique in textual studies is the support vector machine (SVM) binary classifier (Joachims, 1998; Burges, 1998). The basic SVM is a binary classifier that finds a hyperplane with the maximum margin between positive and negative training documents. The three key features of SVMs that reduce the likelihood of over-fitting and make them useful for textual classification are

- Not all training documents are used to train the SVM. Instead, only documents near the classification boarder are used.
- Not all features from the training documents are used, so excessive feature reduction is not needed.
- SMVs can construct irregular boarders between positive and negative training documents.

To confirm the systematic response of trader classes to market news, a binary SVM is used to classify whether the net long positions of trader classes will increase (+1) or decrease (-1). Weekly news is aggregated and the frequency of each unique word (there are $59,323$ unique words in the sample) per week is calculated. These frequency vectors are used as input to train the classifier[9]. Using cross-validation techniques, 300 out-of-sample classifications of whether speculative net long positions increased or decreased in response

[9] The SVM used in this study is available at http://svmlight.joachims.org/.

to market news gave the following results[10]:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} = 56\%^{**}$$

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP} = 54\%^{*} \\ &= \text{The proportion of obs. classified as +1 that are truely +1}\end{aligned}$$

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP+FN} = 81\%^{***} \\ &= \text{The proportion of obs. that truely are +1 that are classified as +1}\end{aligned}$$

where TP, FP, TN, and FN denote the number of true positives, false positives, true negatives, and false negatives produced by the classifier, respectively. When this classifier, trained on data of speculative positions, is applied to out-of-sample data on the positions of hedgers, significant misclassifications are observed indicating that hedgers and speculators are responding to news in significantly different ways.

While both of the techniques used in this section capture only a coarse level of linguistic sophistication, they do indicate a clear difference in the response of hedgers and speculators to market news. Results support the notion that speculators respond to news in a similar way to traders in equity markets. That is, with news that puts downward pressure on prices, speculators decrease the net long positions, while hedgers increase their long positions. Combined with the previous results, this section supports the notion that large speculators are informed traders who bring new information to the market through orders that are placed more aggressively than hedgers.

These differences between hedgers and speculators will become particularly pronounced when the net long positions of these two classes of traders diverge and approach their bounds. At such times, the available counterparties for a trade become less heterogeneous,

[10] Note that $*$ indicates $p < 0.10$, $**$ indicates $p < 0.05$, and $***$ indicates $p < 0.01$.

and particular traits of a trader class may temporarily change the price dynamics. The remainder of this paper examines specifically how and why the price dynamics change during these times of extreme and opposing positions.

III. Extreme Positions and Market Dynamics

To study the market dynamics as trader positions approach their bounds, an event study methodology is used. The first step in this methodology is to define the extreme events that are being studied. To do this, the notion of trader positions, as expressed by the net long measure described in the previous section, must be extended to capture the notion that the positions that trader classes can achieve are bounded. Since the true bounds of positions taken by large hedgers and speculators would be difficult to determine, locally realized limits are used as a proxy for this true limit.

The market under study has been growing over the sample period, so an absolute value of the position bounds for all times is inappropriate. Instead, a local measure, $P_t^C(\tau)$ is employed which uses the maximum and minimum NL_t^C values achieved by trader class C within a moving window of τ periods. In this study, a centered window is used, thereby using information from the recent past and future. Consequently, in this ex post study, we can study return regularities, but no claim of predictability can be made from the results. The advantage of an ex post analysis is that a better proxy for position bounds is permitted. That is, extreme positions are easier to identify when future information is used, whereas in an ex ante setting, a large position may get labeled as extreme and then be dwarfed by an even larger position a few periods later. This trader position measure is expressed as

$$P_t^C(\tau) = \frac{NL_t^C - \min\{NL_{t'}^C\}}{\max\{NL_{t'}^C\} - \min\{NL_{t'}^C\}} \tag{12}$$

where $t' \in \{t-(\tau-1)/2, \cdots, t-1, t, t+1, \cdots t+(\tau-1)/2\}$ for an odd-valued τ. In an ex ante form, using only past information, this variable is similar to the market sentiment indicator of Briese (1990). However, rather than relying on notions of sentiment, the current study argues that the resulting dynamics are the consequence of structural constraints imposed by the bounded size of the market which induces mean-reverting behavior when these bounds are approached.

For the current study, τ is taken to be around nine months (39 weeks), but the results are robust across many window sizes as is shown in Section IV. Clearly, values of $P_t^C(\tau)$ take values between 0 and 1, with values achieving 1 when net long positions exceed all values within the nine month centered moving window. Figure 6 plots this position index for large speculators and hedgers.

[FIGURE 6 ABOUT HERE]

The events of interest occur when one trader class is extremely long and the other class is extremely short. To identify the times of these events, two indicator variables are constructed. SL_t identifies times when speculators (S) are extremely long (L) and hedgers are extremely short, and HL_t identifies times when hedgers (H) are extremely long (L) and speculators are extremely short. These variables are defined to be

$$SL_t = \begin{cases} 1 & \text{if} \quad P_t^S(\tau) - P_t^H(\tau) = \max[P_{t'}^S(\tau) - P_{t'}^H(\tau) : t' \in [t - \frac{(\tau-1)}{2}, t + \frac{(\tau-1)}{2}]] \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

$$HL_t = \begin{cases} 1 & \text{if} \quad P_t^S(\tau) - P_t^H(\tau) = \min[P_{t'}^S(\tau) - P_{t'}^H(\tau) : t' \in [t - \frac{(\tau-1)}{2}, t + \frac{(\tau-1)}{2}]] \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

An SL-type event occurs when the difference between the position indices of speculators and hedgers is larger than at any other time within a centered window of length τ. For example, if $P_t^S(\tau) = 1$ and $P_t^H(\tau) = 0$, then speculators are extremely long and hedgers are extremely short, and so $SL_t = 1$ captures the extreme opposition of the positions of these two classes of traders. Within the sample period there are 20 events of each type.

Such extreme positions do not develop without reason and are clearly dependent on world events related to the crude oil markets. However, it is not within the scope of this paper to identify the specific causes of these extreme positions. That said, Section II.A identified divergent position accumulation by hedgers and speculators in response to market related news. Therefore, if news variables have temporal persistence, then extreme positions would regularly result from the chance occurrence of many news innovations of the same type (i.e. either a 'good' or 'bad' cluster of news events). Such persistence in news variables is in fact observed, with significant first order autocorrelations of $\rho_1(Neg_t) = 0.54$, $\rho_1(neg_t) = 0.97$, $\rho_1(Pos_t) = 0.45$, and $\rho_1(pos_t) = 0.96$.

III.A. Extreme Positions and the Number of Traders

The argument that bounded trader positions can induced liquidity shocks is based on the limited heterogeneity of counterparties when extreme positions are taken. Of course, the number of traders of each class in the market is another determinant of the heterogeneity of available counterparties. As is shown below, the number of traders in the market and the acquisition of extreme positions is not independent. In the case of HL-type events, both factors move in conjunction and result in strong liquidity induced dynamics. In the case of SL-type events, the two factors more in opposition to each other thereby decreasing the strength of the liquidity-induced dynamics.

Up until this point in the analysis, no consideration has been given to the demographics of the groups of large hedgers and speculators. The Commitments of Traders Reports list the number of large traders in each of these groups, thereby allowing an analysis of their composition around the times that they acquire extreme positions. For the class of SL-type events, where speculators are extremely long and hedgers extremely short, the left panel of Figure 7 shows that, on average, traders are entering the market before the event date and then leave the market afterwards. Conversely, the right panel of the same figure shows that for the class of HL-type events, where hedgers are extremely long and speculators extremely short, traders tend to leave the market before the event date and then return afterwards.

[**FIGURE 7 ABOUT HERE**]

Two questions arise from the observation that traders leave or enter the market before and after extreme positions are taken: Who is it that is leaving the market, and why are they leaving? A partial answer to the first question is seen in Figure 8 which shows that the majority of the traders that systematically leave and enter the the market around the event dates are speculators, and not hedgers. The fact that hedgers do not enter or exit the market in a regular fashion around these extreme position events is due to the fact that their positions are not dependent on the profitability of their trades, but are rather determined by their business interests.

[**FIGURE 8 ABOUT HERE**]

The number of speculators can be further disaggregated into the number of speculators who are long and the number that are short. Figure 9 shows these numbers, thereby giving a fuller answer to the question of who is leaving the market and why they are leaving. The right panel of the figure indicates that virtually all of the traders that are leaving the market before the HL-type events are speculators that were long. Similarly, the majority of speculators who enter the market leading up to an SL-type event are also those speculators that were long. In the following sections the set of speculators that enter and exit long positions are shown to be significant determinants of the market dynamics around these extreme events.

[**FIGURE 9 ABOUT HERE**]

III.B. Extreme Positions and Returns

It was shown in Section II that a strong relationship exists between trader positions and market returns. Purchases by speculators, and sales by hedgers, tend to push prices higher. Sales by speculators, and purchases by hedgers, tend to push prices lower. This relationship implies that extreme events are often preceded by periods of trending prices. Once trader positions are near their upper or lower bounds, they inevitably move back towards their average since they cannot become much more extreme than they currently are. When trader positions revert away from their extremes, the price trend often reverses.

In the case of SL-type events, where speculators are extremely long and hedgers extremely short, there is a significant increase in prices that begins around seven weeks before the extreme event resulting, on average, in an 8.8% cumulative abnormal return by the event date, where the abnormal return is defined to be the return minus the sample mean weekly return of $\overline{r} = 0.36\%$. More formally, the cumulative abnormal return between times $t - \tau$ and t is

$$CAR(t-\tau, t) = \prod_{i=0}^{\tau}(1 + r_{t-i} - \overline{r}) - 1. \tag{15}$$

After the event, the upward trend is reversed and prices tend to fall for the next eight weeks. Thus, SL-type events tend to coincide with local price maximums. In the case of HL-type events, where speculators are extremely short and hedgers extremely long, there is a significant decrease in prices that begins around six weeks before the extreme event resulting, on average, in a -9.3% cumulative abnormal return by the event date. The price trend reverses after the event date and tends to rise for the next seven weeks. Thus, HL-type events tend to coincide with local price minimums. Both of these cases are seen in Figure 10.

[FIGURE 10 ABOUT HERE]

To evaluate the effect of trader positions on returns for the week before and after those positions become extreme, the following model is estimated

$$r_t = \alpha + \beta_1 SL_t + \beta_2 SL_{t-1} + \gamma_1 HL_t + \gamma_2 HL_{t-1} + \varepsilon_t \tag{16}$$

where ε_t is assumed to be a white noise process, SL_t is an indicator variable for times when large speculators are extremely long and large hedgers are extremely short, and HL_t is an indicator variable for times when large speculators are extremely short and large hedgers are extremely long. The result of this estimation is found in Table IV. All of the estimated variable coefficients except for γ_1 are significant. Since the return, r_t, measures the price change between times t and $t-1$, β_1 is the change in price leading up to an SL-type event occurring at time t. Similarly, β_2 is the change in price that occurs when the SL-event occurred at time $t-1$, that is, the change in price after such an event. Estimates of β_1 and β_2 are positive and negative, respectively, indicating that prices tend to rise during the week before an SL-type event and fall the following week. Conversely, estimates of γ_1 and γ_2 are negative and positive, respectively, indicating that prices tend to fall during the week before an HL-type event and rise the following week. The low R^2 of this and other regressions in the paper are low, as expected, since there are only 20 events of each type. Clearly, there is a lot more going on in the return dynamics than just those resulting from extreme trader positions. However, when such events occur, the effects are significant, both economically and statistically. These results are consistent with the price trends and reversals around extreme events seen in Figure 10.

[TABLE IV ABOUT HERE]

This relationship between positions and returns can be explained through the bounded and mean reverting nature of extreme positions, but there is also an influence from the entry and exit of speculators around these event dates. Recall from the right panel of Figure 9 that virtually all of the traders that are leaving the market before the HL-type event are speculators who are long. Given the returns during this time, these traders were long in a falling market, so it is likely that they were leaving the market to avoid further losses. At the same time leading up to the event date, the number of speculators that are short is actually increasing. These traders who are selling short in a falling market are realizing significant profits, and immediately after the event date they begin to take their profits by closing their short positions. This is seen in Figure 9 where the number of speculators who are short after the HL-type event dates steadily falls. Also after the event date, the traders who were long and had left the market begin to re-enter the market establishing new long positions. Thus, there are regularities in the entry and exit of traders around the event dates that are consistent with speculative profit-taking and with capital constraints forcing traders to exit losing trades. A similar argument can be made to explain the regularities around SL-type events in the left panel of Figure 9.

To check whether these regularities in the entry and exit of speculators plays a role in the microstructure of returns the following model is estimated,

$$r_t = \alpha + \beta_1 \triangle T^{spec}_{long,t} + \beta_2 \triangle T^{spec}_{short,t} + \varepsilon_t \tag{17}$$

where $\triangle T^{spec}_{long,t}$ is the change since the previous period in the number of speculators who are long, $\triangle T^{spec}_{short,t}$ is the change in the number of speculators who are short, and ε_t is a white noise process. The estimation results in Table V indicate that an increase in the number of long speculators is associated with a significantly higher return.

[TABLE V ABOUT HERE]

An interesting observation arises when market demographics are examined around the price reversals associated with SL and HL-type events. If a speculative trader were aware of the date of a significant reversal of price movements, then they would take positions that would profit from this price action. An informed trader[11] would therefore want to be long on an HL-event date and short on an SL-event date. However, these events are essentially defined to be times when large speculators are on the wrong side of the market. On average, at SL-event dates, only 36% of large speculators are short, and only 38% are long on HL-event dates. Thus, an essential feature of these liquidity shocks is that a significant number of traders be uninformed in their timing.

Figure 9 shows that the majority of new speculative bets are made in the 'correct' direction with respect to the price trends both before and after the event dates. The longer the trend has been in effect, the more speculators trade in that direction. However, some traders enter the trend too late as is indicated by the spike at time zero. Since only aggregate trader numbers are reported in the CoT reports, it is impossible to tell when a specific trader both enters and exits the market. Consequently, the distribution of profits within a class of traders is unknown.

With the majority of large speculators being on the wrong side of the market at SL and HL-event dates, some conclusions about uninformed traders can be drawn. First, it would be inappropriate to model these uninformed traders as being equally likely to take long or

[11]Since a hedger's trading behavior is determined by business interests and risk management strategies, their actions would be independent of knowledge of these event dates. In this sense, hedgers act as if they are uninformed traders.

short positions, since that would imply that on average the majority of speculators would be on the right side of the market[12]. Second, uninformed trades appear to be dependent on past trades since they appear to want to follow a trend that is about to end. Finally, the number of speculators trading in the direction of a trend tends to increase with the duration of the trend. This seems to indicate that there are varying degrees to which traders are informed, with well informed traders entering a trend early, other informed traders entering as the trend develops, and the uninformed traders entering too late as the trend ends.

III.C. Extreme Positions and Realized Volatility

So far in the analysis, significant differences in the behavior of hedgers and speculators, as well as the bounded nature of trader positions, have been found to drive return regularities around dates of extreme and opposing trader positions. These dates also coincide with systematic entries and exits of large speculators from the market, primarily by those holding long positions. In both SL and HL-type events, the changes in the number of speculators who are short moves in opposition to the changes in the number of speculators who are long, however the magnitude of the latter changes are more than twice as large. Since speculators facing losses may leave a market through orders placed more aggressively than other orders, there is reason to suspect that return volatility may differ between SL-type and HL-type events.

As was shown earlier in Figure 10, HL-type events are characterized by first falling and then rising prices. As prices initially fall, a significant number of speculators holding long positions leave the market (see the right panel of Figure 9). These speculators were long in a

[12]If uninformed traders are equally likely to take long or short positions, then in expectation half of the uninformed traders are on the right side of the market. Since informed traders would be on the right side of the market as well, then overall there are more traders on the right side of the market than on the wrong side.

falling market, and hence exited the market while they were facing increasing losses. Under such circumstances, orders to exit a losing trade may well be placed more aggressively than would ordinarily be observed. Such aggressive orders would be observed as an increase in the return volatility. After the price trend reverses at the event date, and speculators return to the market, volatility is expected to return to normal levels.

Around SL-type events, a similar argument would suggest that speculators holding long positions who leave the market after the event date would also drive up the return volatility. To examine the conjecture that speculators facing losses leave the market using aggressive orders thereby causing high return volatilities, this section examines the behavior of the realized volatility around both types of extreme position events. Realized volatilities are sampled weekly and constructed with five-minute returns, as described in Section I. Figure 11 plots the average realized volatility around the event dates. There is a clear difference between the two classes of extreme events. The average realized volatility before and immediately after type-SL extreme positions, where speculators are long and hedgers are short, is significantly lower than the overall sample average[13]. In contrast, it is significantly higher before and just after HL-type extreme positions, where speculators are short and hedgers are long. In this case, the average realized volatility on the event date is at the 79th percentile of the entire sample and is even higher for several weeks before the event date. Both panels of Figure 11 are consistent with the notion that traders facing losses can significantly increase return volatility as they exit the market.

[FIGURE 11 ABOUT HERE]

To quantify the increase in volatility around HL-type events, the following model is

[13]For Figure 11 and the following estimations, a single outlier realized volatility value was removed from the series. The outlier was on 24 March 1998 and was over 19 standard deviations away from the sample mean. The value was replaced with an average of the preceding and following values. After examining the high-frequency price data, the outlier appears to have been caused by an incorrect price entry.

estimated:

$$\ln(RV_t) = \alpha + \beta \ln(RV_{t-1}) + \gamma HL_t^* + \varepsilon_t \qquad (18)$$

where RV_t is the realized volatility at time t, and ε_t is assumed to be a white noise process. HL_t^* is an indicator variable for whether a HL-type extreme event occurs at time t or during the next five weeks following t. The five week window captures the fact that volatility builds, and in fact peaks, during the several weeks leading up the the event date. This new indicator is defined more formally as

$$HL_t^* = \begin{cases} 1 & \text{if} \quad HL_t \vee HL_{t+1} \vee HL_{t+2} \vee HL_{t+3} \vee HL_{t+4} \vee HL_{t+5} \\ 0 & \text{otherwise} \end{cases} \qquad (19)$$

where $\vee$ is the logical 'OR' operator. The results of this estimation are found in Table VI. The estimated coefficient of HL_t^* indicates realized volatility is expected to rise by 6.6% five weeks prior to an HL-type event and remain higher until the event date. Since HL_t^* dates occur in groups of six consecutive weeks, and since the model (19) accounts for previous weeks, the realized volatility at the event date will tend to be 17% higher than it was six weeks prior to the event date, indicating that HL-type extreme events can be an economically important cause of high-volatility periods.

[TABLE VI ABOUT HERE]

The different volatility effects that are observed before and after HL and SL-type events, as well as the trend reversals seen at these times, it is likely that realized volatilities are not symmetrically composed of positive and negative returns. Figure 12 plots the average positive and negative directional realized volatilities for the weeks around HL-type events, when hedgers are extremely long and speculators extremely short. It is clear that the realized volatilities before and after the event date are composed of two nonsymmetric components. Before the event, the majority of the volatility comes from negative returns as is seen from the increase in DRV_t^N before the event date in Figure 12(a). After the event date, DRV_t^N falls sharply, but DRV_t^P, in Figure 12(b), peaks thereby keeping the overall realized volatility higher than average. This asymmetric composition of return volatility indicates that the potential ill-effects of liquidity shocks do not affect all market participants equally. High negative directional volatility would be bad for sellers, while high positive directional volatility would be bad for buyers. Further, a significant difference between the negative and positive volatilities indicates trending prices.

[FIGURE 12 ABOUT HERE]

III.D. Extreme Positions and Bid-Ask Spreads

The significance of these dynamic regularities around extreme position dates can be further verified using the prominent percentage bid-ask spread measure of liquidity. Figure 13 plots the average daily percentage bid-ask spread around both types of event dates. The pattern of larger than average spreads around HL-type events and lower than average spreads around SL-type events are similar to the patterns of realized volatility around these dates. With fewer speculators in the market at HL-type event dates, trade can only be induced by offering better prices to potential buyers and sellers resulting in a larger bid-ask spreads. On SL-type event dates, there are more speculators in the market and finding willing counterparties is easier, resulting in smaller bid-ask spreads. These findings indicate that, despite the fact that speculators can raise return volatility as they close losing positions, their increase presence in the market during SL-type events improves liquidity. While this study is concerned with the movement of large speculators into and out of a single market, if these movements actually constitute movements from or into other futures markets, then in addition to impacting liquidity within the crude oil futures market, these movements of traders may also account for some of the time-varying cross-market liquidity variation (Spiegel, 2008) in futures markets.

[FIGURE 13 ABOUT HERE]

To quantify the effects of these events on the bid-ask spread, the following model is estimated:

$$S_t = \alpha + \beta(HL_{t-1} + HL_t + HL_{t+1}) + \gamma(SL_{t-1} + SL_t + SL_{t+1}) + \varepsilon_t \tag{20}$$

where S_t is the estimated daily percentage bid-ask spread described in Section I, and ε_t is assumed to be a white noise process. Since extreme events cannot occur in consecutive weeks, $(HL_{t-1} + HL_t + HL_{t+1})$ and $(SL_{t-1} + SL_t + SL_{t+1})$ are indicator functions taking the value 1 when the time t is within one week of an extreme event and taking the value 0 otherwise. Table VII contains the estimation results where we see that both β and γ are significant. The coefficient β has the expected positive sign since the spreads are higher on HL-type event dates when there are fewer speculators in the markets, and γ has the expected negative sign since there are more speculators in the market during SL-type events resulting in smaller spreads.

[TABLE VII ABOUT HERE]

IV. Robustness Checks

The results of this study depend largely on the identification of dates during which hedgers and speculators are holding extreme and opposing positions. The extremity of trader positions is defined by their size relative to positions held within a centered window of width

τ weeks. Thus, τ is a critical parameter in the study and the results should be robust to changes in its value.

Return and realized volatility regularities are central results of this study, and models (17) and (19) were used to quantify these regularities. The estimation of these models used a centered moving window of $\tau = 39$ weeks. The results were also checked for window sizes $\tau = 31, 33, 35, 37, 41, 43, 45$, and 47 weeks. The coefficient estimates and their p-values for models (16) and (18) for each of these window lengths are listed in Table VIII. For robustness to hold, the sign, magnitude, and significance of these estimates should be reasonably close to each other and to the values reported in this paper for $\tau = 39$. Of the 45 coefficient estimates for the crude oil market in Table VIII, all have the same sign as those listed in the paper, all are of the same order of magnitude as those listed in the paper, and all but four are significant or not, as listed in the paper.

[**TABLE VIII ABOUT HERE**]

Similar results should be found in other futures markets where the distinction between hedgers and speculators is clear. The lower portion of Table VIII lists the estimates of models (16) and (18) for three additional markets: Soybeans from the Chicago Board of Trade, Live Cattle and 3-Month Eurodollars, both from the Chicago Mercantile Exchange. All estimated coefficients from the model of returns have the same sign as those of crude oil, though several are not significant. The model for Eurodollars returns has no significant estimates, and this is likely because of the fuzzy distinction between hedgers and speculators in financial futures markets. When a trader uses a market for both hedging and speculating, as is common in interest rate and equity markets, they are identified as hedgers in the Commitments of Traders data even if their dominant activity is speculative. This reporting standard causes the identifiable systematic differences between hedgers and speculators to disappear. In agricultural markets, on the other hand, the classes of hedgers and speculators are much more distinct, and consequently more estimates are significant.

Cross-sectional differences across markets are more pronounced in the volatility regularities around event dates. For example, as seen in the top panels of Figure 14, there is a clear rising of volatility around SL-type events and a clear falling of volatility around HL-type events in the Soybean market. This is the opposite of the effect found in crude oil futures, and is due to the confounding effects of the differing systematic movements of speculators into and out of the market around these dates.

[Figure 14 ABOUT HERE]

There is a large literature on the market equilibrium approach to futures pricing (Gibson and Schwartz, 1990; Fama and French, 1987) which use fundamental market factors to model prices. For crude oil, prominant examples of such factors include convenience yields, crude oil inventories, and crude oil production measures. Data for crude oil inventories, production and spot prices are available from the Energy Information Administration[14] (E.I.A.), and 1-month Treasury bill rates are available from Kenneth French's website[15]. While several event studies of these factors did not reveal any systematic regularities around extreme position accumulations, a full incorporation of these variables is difficult due to different release dates and frequency of observations. Given the current focus on the microstructural mechanism of liquidity shocks, investigating the relationship between extreme positions and market fundamentals is left for future study.

In summary, the central results of this study are robust within the crude oil futures market across several neighboring parameter values, and supporting evidence is found in agricultural futures markets where the hedger-speculator distinction is clear. There appear to be some cross-sectional differences between markets that induce differing regularities

[14] http://www.eia.doe.gov/

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

around event dates. Since these results depend on reliable differences in the behavior of hedgers and speculators, supporting results are not found in financial futures markets where the distinction between hedgers and speculators is not always clear. Finally, there may be fundamental factors at play which drive position accumulation, however the relationship between trader positions and these factors is not simple and will be pursued in a future study.

V. Conclusion

Liquidity is a complex notion that captures the impact and costs of trading in a market. In liquid markets, trades are executed with low costs and result in little impact on the market price. Yet even in large and active markets, there are periods of illiquidity characterized by high volatility and bid-ask spreads, and even directional return regularities. This paper presents an empirical microstructural study of such liquidity-induced dynamics in the New York Mercantile Exchange's crude oil futures market, a very large and active market which is usually considered a very liquid market.

The paper's results suggest that during times when large hedgers and speculators acquire large and opposing positions, with one group going extremely long and the other going extremely short, liquidity-induced dynamics temporarily dominate price action. These dynamics can include price trend reversals, high volatility, and high bid-ask spreads. The underlying mechanism that cause these effects has three components: (1) significant differences between hedgers and speculators, (2) the bounded and mean-reverting nature of trader positions, and (3) the various conditions that cause speculators to enter and exit the market.

Large hedgers and speculators are the dominant participants in futures markets, and each group have different constraints as well as different reasons for participating in the market. Their impact on prices through their trades are also found to be significantly different. Large speculators are found to increase prices with their purchases, to engage in trend-following behavior, and to react to market-related news. Large hedgers, on the other hand, decrease prices through their purchases and exhibit contrarian-type behavior.

Speculators and hedgers are also found to differ in their response to news about the crude oil market. In two distinct analyses using over 58,000 news stories relating to crude oil markets, we found that news variables have temporal persistence and that traders systematically accumulate positions (with positions of hedgers and speculators diverging) in response to news. Combined with the previous results, these results support the notion that large speculators are informed traders who bring new information to the market through orders that are placed more aggressively than hedgers.

While positions of hedgers and speculators can diverge in response to news, there are bounds on how far they may diverge. When the positions held by hedgers and speculators diverge and approach their limits, one class of traders begins to dominate the set of available counterparties for purchases, and the other class dominates that for sales. When this happens, the differences between trader classes, which are usually 'washed out' when the set of counterparties is well mixed, begin to have a pronounced impact on price changes. Prices trend strongly until position bounds are approached, and then reverse as positions revert towards normal levels.

The impact on liquidity during these episodes is found to be largely determined by the entry and exit of large speculators from the market. In particular, speculators facing increasing losses exit the market using orders that are significantly more aggressive than typical orders, thereby pushing up return volatility. On the other hand, speculators entering the market tend to improve market liquidity.

The asymmetry of liquidity shocks is also investigated. Both before and after speculators are extremely short and hedgers extremely long, realized volatility is significantly higher than its average. However, while these positions are being accumulated the volatility is dominated by negative returns. As these positions are being unwound, the volatility is dominated by positive returns. These effects are captured by a directional realized volatility measure and indicate that liquidity shocks can affect different market participants in opposing ways. That is, an illiquid market for a seller can be a liquid market for a buyer, and liquidity measures should be able to capture such asymmetries.

References

Andersen, T. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**(4), 885–905.

Brandt, B. and Kavajecz, K. (2002). Price discovery in the U.S. treasury market: The impact of orderflow and liquidity on the yield curve. University of Pennsylvania Working Paper.

Briese, S. (1990). Commitments of traders as a sentiment indicator. *Technical Analysis of Stocks and Commodities*, **8**(5), 199–204.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, 121–168.

CFTC (2007). Backgrounder. *Commodity Futures Trading Commission.*

Chordia, T., Roll, R., and Subrahmanyam, A. (2001). Market liquidity and trading activity. *Journal of Finance*, **56**(2), 501–530.

Chordia, T., Sarkar, A., and Subrahmanyam, A. (2005). An empirical analysis of stock and bond market liquidity. *The Review of Financial Studies*, **18**(1), 85–129.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. (2001). *An Introduction to High Frequency Finance.* Academic Press, San Diego.

Edwards, F. and Canter, M. (1995). The collapse of metallgesellschaft: Unhedgeable risks, poor hedging strategy or just bad luck? *Journal of Futures Markets*, **15**.

Fama, E. and French, K. (1987). Commodity future prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, **60**(1), 55–73.

Gibson, R. and Schwartz, E. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, **45**(3), 959–976.

Goyenko, R., Holden, C., and Trzcinka, C. (2008). Do liquidity measures measure liquidity? *Journal of Financial Economics*, (forthcoming).

Hansen, L., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, **14**, 262–280.

Hasbrouck, J. and Sofianos, G. (1993). The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance*, **48**(5), 1565–1593.

Huang, R., Cai, J., and Wang, X. (2001). Inventory risk-sharing and public information-based trading in the treasury note interdealer broker market. University of Notre Dame Working Paper.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Proceedings of the Tenth European Conference on Machine Learning*, pages 137–142.

Lee, C. and Ready, M. (1991). Inferring trade direction from intraday data. *Journal of Finance*, **46**, 733–746.

Mello, A. and Parsons, J. (1995). Maturity structure of a hedge matters: Lessons from the metallgesellschaft debacle. *Journal of Applied Corporate Finance*, **7**.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, **39**, 1127–1139.

Sanders, D., Boris, K., and Manfredo, M. (2004). Hedgers, funds, and small speculators in the energy futures markets: An analysis of the cftc's commitments of traders reports. *Finance Research Letters*, **26**(3), 425–445.

Spiegel, M. (2008). Patterns in cross market liquidity. *Finance Research Letters*, **5**, 2–10.

Tetlock, P., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, **LXIII**(3), 1437–1467.

Wang, C. (2003). Investor sentiment, market timing, and futures returns. *Applied Financial Economics*, **13**(12), 891–898.

Trader Class	Min	Max	Median	Mean	Std Dev	Skewness	Kurtosis
Speculators	$-71,928$	$88,712$	$8,361$	$8,471$	$30,544$	0.06207	2.777
Hedgers	$-103,854$	$94,868$	$-8,103$	$-7,804$	$39,144$	-0.05284	2.617

Table I.: Summary Sample Statistics for the Net Long Measure of Position Holdings for Large Hedgers and Speculators. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006, as reported in the CFTC Commitments of Traders Reports.

$$NL_t^S = \alpha^1 + \beta_1^1 NL_{t-1}^S + \beta_2^1 r_t + \beta_3^1 r_{t-1} + \varepsilon_t^1$$

	Intercept	NL_{t-1}^S	r_t	r_{t-1}
Coefficient	-1388.58	0.89	$417,665$	$90,670$
t-stat	-2.20	44.20	14.75	5.94
p-value	0.03	0.00	0.00	0.00
R^2	0.91			

$$NL_t^H = \alpha^2 + \beta_1^2 NL_{t-1}^H + \beta_2^2 r_t + \beta_3^2 r_{t-1} + \varepsilon_t^2$$

	Intercept	NL_{t-1}^H	r_t	r_{t-1}
Coefficient	1991.80	0.88	$-549,829$	$-124,626$
t-stat	2.55	46.25	-14.60	6.55
p-value	0.01	0.00	0.00	0.00
R^2	0.91			

$$r_t = \alpha^3 + \beta_1^3 r_{t-1} + \beta_2^3 NL_t^S + \beta_3^3 NL_{t-1}^S + \varepsilon_t^3$$

	Intercept	r_{t-1}	NL_t^S	NL_{t-1}^S
Coefficient	0.0020	-0.23	0.0000026	-0.0000023
t-stat	1.39	-5.89	12.57	-11.45
p-value	0.17	0.00	0.00	0.00
R^2	0.35			

Table II.: Trader Positions and Returns. Estimation is through continuous updating GMM. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

$$NL_t^S = \alpha + \beta_1 neg_t + \beta_2 pos_t + \beta_3 NL_{t-1}^S + \varepsilon_t$$

	Intercept	neg_t	pos_t	NL_{t-1}^S
Coefficient	$27,524$	$312,887$	$-1,263,992$	0.90
Std Error	$20,187$	$410,409$	$551,890$	0.02
t-stat	1.36	0.76	-2.29	44.08
p-value	0.17	0.45	0.02	0.00
R^2	0.84			

Table III.: Impact of Positive and Negative Words on Trader Positions. Statistics are calculated from 405 weekly observations from July 25, 2001 until April 28, 2009. The *pos* and *neg* variables represent the frequency of positive and negative words in news articles from each week. Trader positions are reported in the CFTC Commitments of Traders Reports.

$$r_t = \alpha + \beta_1 SL_t + \beta_2 SL_{t-1} + \gamma_1 HL_t + \gamma_2 HL_{t-1} + \varepsilon_t$$

	Intercept	SL_t	SL_{t-1}	HL_t	HL_{t-1}
Coefficient	0.000	0.019	-0.031	-0.018	0.024
Std Error	0.002	0.007	0.007	0.012	0.011
t-stat	0.095	2.530	-4.331	-1.465	2.130
p-value	0.924	0.012	0.000	0.143	0.034
R^2	0.025				

Table IV.: Regression Model of the Effect of Current and Past Extreme Trader Positions on Weekly Returns. The variable SL_t is an indicator variable for times when large speculators are extremely long and large hedgers are extremely short, and similarly HL_t is an indicator variable for times when large speculators are extremely short and large hedgers are extremely long. Heteroskedasticity robust standard errors and p-values are reported. The error term is assumed to be from a white noise process. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

$$r_t = \alpha + \beta_1 \triangle T_{long,t}^{spec} + \beta_2 \triangle T_{short,t}^{spec} + \varepsilon_t$$

	Intercept	$\triangle T_{long,t}^{spec}$	$\triangle T_{short,t}^{spec}$
Coefficient	0.000	0.003	-0.001
Std Error	0.002	0.000	0.000
t-stat	-0.080	10.771	-1.524
p-value	0.937	0.000	0.128
R^2	0.173		

Table V.: Regression Model of the Effect of Changes in the Number of Speculators on Weekly Returns. The variable $\triangle T_{short,t}^{spec}$ is the change in the number of speculators who are short, and $\triangle T_{long,t}^{spec}$ is the change in the number of speculators who are long. Heteroskedasticity robust standard errors and p-values are reported. The error term is assumed to be from a white noise process. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

$$\ln(RV_t) = \alpha + \beta \ln(RV_{t-1}) + \gamma HL_t^* + \varepsilon_t$$

	Intercept	$\ln(RV_{t-1})$	HL_t^*
Coefficient	-2.254	0.629	0.066
Std Error	0.712	0.116	0.027
t-stat	-3.167	5.434	2.464
p-value	0.002	0.000	0.014
R^2	0.465		

Table VI.: Regression Model of the Effect of Extreme Trader Positions on Weekly Realized Volatility. RV_t is the realized volatility at time t, HL_t^* is an indicator variable for whether a HL-type event occurs within the next five weeks, and ε_t is assumed to be a white noise process. HL_t-type events occur when large speculators are extremely short and large hedgers are extremely long. Heteroskedasticity robust standard errors and p-values are reported. The error term is assumed to be from a white noise process. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

$$S_t = \alpha + \beta(HL_{t-1} + HL_t + HL_{t+1}) + \gamma(SL_{t-1} + SL_t + SL_{t+1}) + \varepsilon_t$$

	Intercept	β	γ
Coefficient	0.112	0.012	-0.009
Std Error	0.001	0.004	0.004
t-stat	96.05	3.258	-2.351
p-value	0.000	0.001	0.025
R^2	0.025		

Table VII.: Regression Model of the Effect of Extreme Trader Positions on Average Daily Percentage Bid-Ask Spreads. The variable SL_t is an indicator variable for times when large speculators are extremely long and large hedgers are extremely short, and similarly HL_t is an indicator variable for times when large speculators are extremely short and large hedgers are extremely long. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

$$r_t = \alpha + \beta_1 SL_t + \beta_2 SL_{t-1} + \gamma_1 HL_t + \gamma_2 HL_{t-1} + \varepsilon_t$$

$$\ln(RV_t) = \alpha + \beta_3 \ln(RV_{t-1}) + \delta HL_t^* + \varepsilon_t$$

τ	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\delta}$
31	0.025	−0.034	−0.015	0.020	0.117
	(0.000)	(0.002)	(0.155)	(0.038)	(0.010)
33	0.026	−0.024	−0.016	0.029	0.118
	(0.000)	(0.002)	(0.099)	(0.001)	(0.011)
35	0.026	−0.024	−0.017	0.023	0.093
	(0.000)	(0.002)	(0.547)	(0.036)	(0.057)
37	0.021	−0.028	−0.012	0.021	0.122
	(0.007)	(0.000)	(0.370)	(0.064)	(0.029)
39	0.019	−0.031	−0.018	0.024	0.132
	(0.012)	(0.000)	(0.143)	(0.034)	(0.014)
41	0.021	−0.027	−0.016	0.024	0.126
	(0.007)	(0.001)	(0.246)	(0.043)	(0.020)
43	0.026	−0.031	−0.015	0.023	0.141
	(0.000)	(0.000)	(0.296)	(0.050)	(0.012)
45	0.026	−0.034	−0.008	0.018	0.140
	(0.000)	(0.000)	(0.627)	(0.132)	(0.020)
47	0.026	−0.034	−0.011	0.017	0.146
	(0.001)	(0.000)	(0.486)	(0.182)	(0.018)
Soybeans	0.004	−0.013	−0.012	0.011	−0.102
$\tau = 39$	(0.658)	(0.076)	(0.017)	(0.050)	(0.033)
Live Cattle	0.008	−0.013	−0.002	0.007	0.137
$\tau = 39$	(0.013)	(0.000)	(0.795)	(0.260)	(0.002)
Eurodollar	0.000	−0.000	−0.000	0.000	0.106
$\tau = 39$	(0.119)	(0.600)	(0.768)	(0.680)	(0.109)

Table VIII.: Robustness Check on the Size of the Centered Window Size and on Other Markets. A potentially critical parameter in the determination of extreme event dates is the size of the moving window τ. Throughout the paper $\tau = 39$ weeks is used. This table shows that the sign, magnitude, and significance of the critical regression coefficients are reasonably close for 8 other window lengths. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

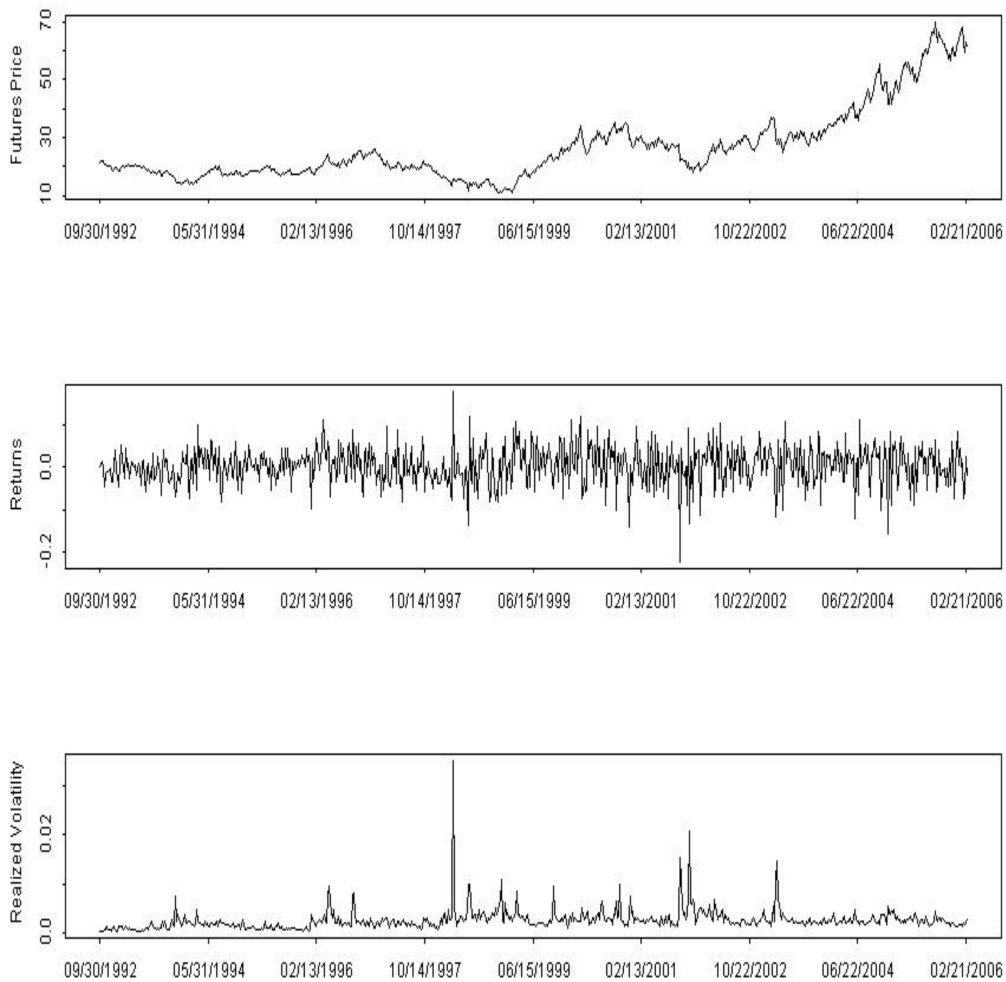


Figure 1: Weekly Prices (top frame), Returns (middle frame), and Realized Volatility (bottom frame) for NYMEX Light Sweet Crude Oil Futures. Prices are the closing price on each Tuesday. Returns are based on an unleveraged investment of the value of the underlying commodity at the futures price of the previous Tuesday. Realized volatility is also sampled each Tuesday, and is constructed by summing the squared five-minute returns over the previous week. The sample period is from September 30, 1992 to February 28, 2006 with 698 observations. Source: TickData, www.tickdata.com.

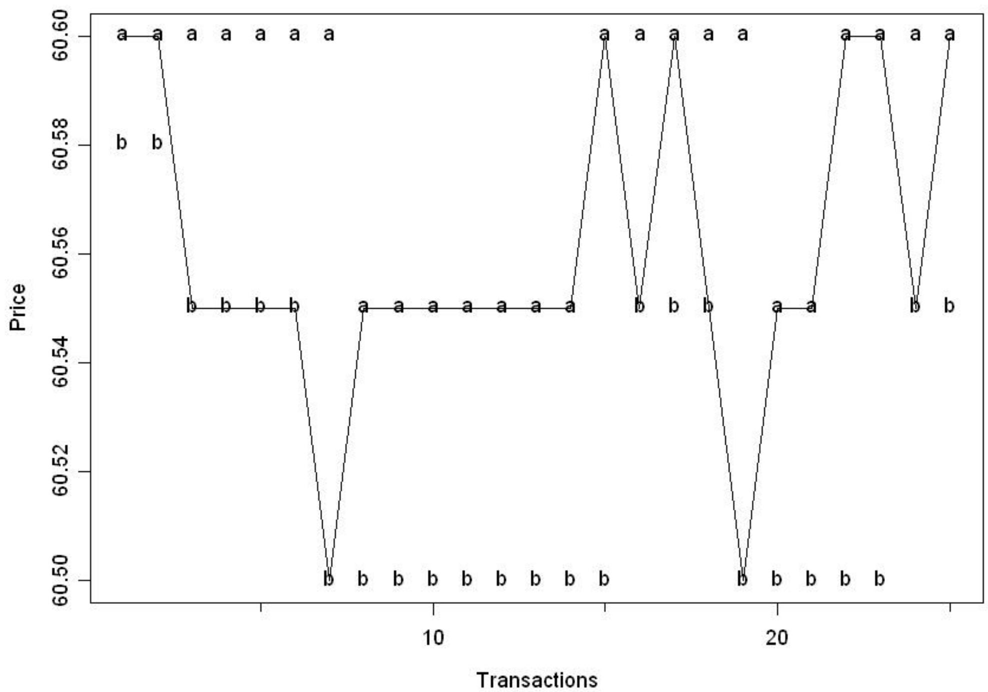


Figure 2: Transaction Prices with Estimated Bids and Asks. Transaction prices are shown by the solid line, with the estimated bids for each transaction at the 'b' markers and the estimated asks at the 'a' markers. The estimated bid-ask spread is $a_t - b_t$. Source: TickData, www.tickdata.com.

CRUDE OIL, LIGHT SWEET - NEW YORK MERCANTILE EXCHANGE Code-067651
FUTURES ONLY POSITIONS AS OF 07/17/07

NON-COMMERCIAL			COMMERCIAL		TOTAL		NONREPORTABLE POSITIONS	
LONG	SHORT	SPREADS	LONG	SHORT	LONG	SHORT	LONG	SHORT
(CONTRACTS OF 1,000 BARRELS)					OPEN INTEREST:			1,549,425
COMMITMENTS								
246,844	137,421	310,801	911,229	1015492	1468874	1463714	80,551	85,711
CHANGES FROM 07/10/07 (CHANGE IN OPEN INTEREST: 3,053)								
-2,201	663	12,566	-10,804	-18,584	-439	-5,355	3,492	8,408
PERCENT OF OPEN INTEREST FOR EACH CATEGORY OF TRADERS								
15.9	8.9	20.1	58.8	65.5	94.8	94.5	5.2	5.5
NUMBER OF TRADERS IN EACH CATEGORY (TOTAL TRADERS: 324)								
107	94	127	91	102	268	269		

Figure 3: Example of a CFTC Commitments of Traders (CoT) Report. Released every Friday, the CoT report summarizes the positions of various classes of traders as of the previous Tuesday. The commercial class refers to large hedgers, and the non-commercial class refers to large speculators. Non-reportable positions are those held by small traders. The Commitments of Traders Reports are published every Friday on the CFTC website at http://www.cftc.gov/cftc/cftccotreports.htm.

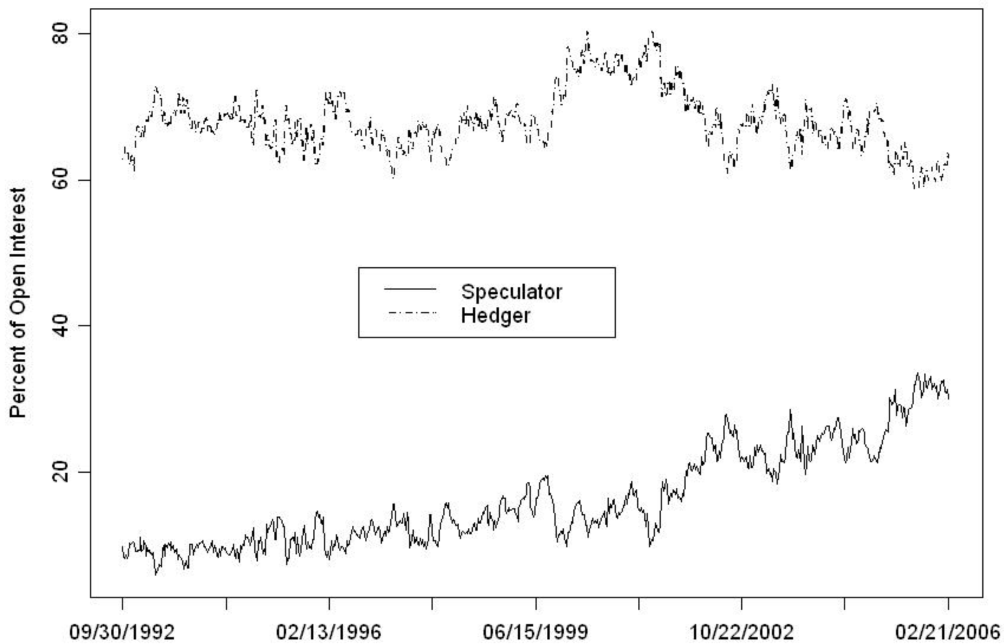


Figure 4: Percent of Open Interest for Large Hedgers and Speculators. Information is from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports.

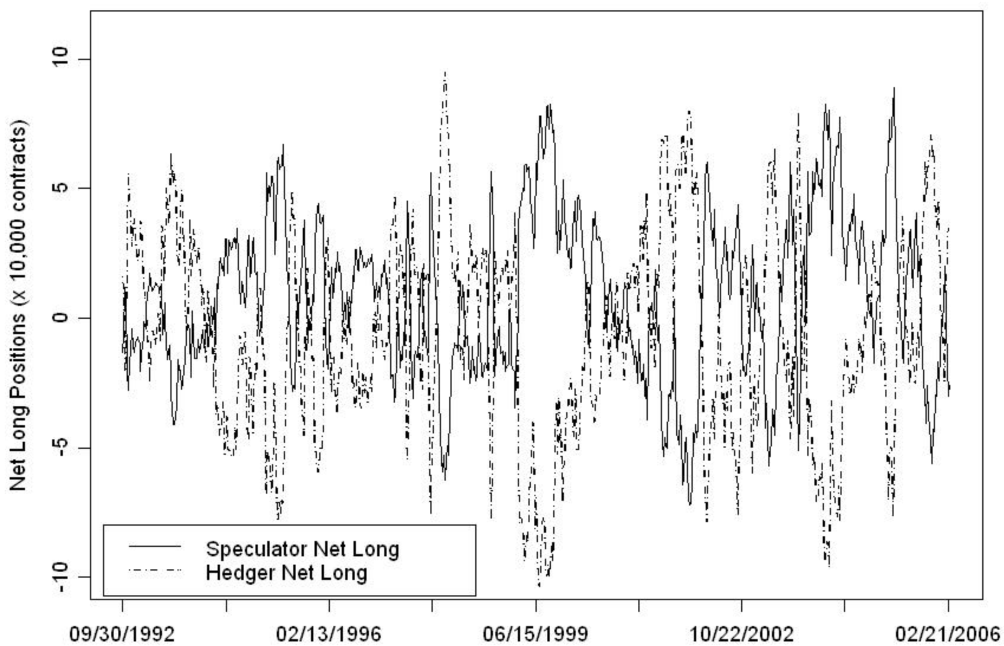


Figure 5: Net Long Positions of Large Hedgers and Speculators. The net long positions are the number of long positions minus the number of short positions within the trader class. Values are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006, as reported in the CFTC Commitments of Traders Reports.

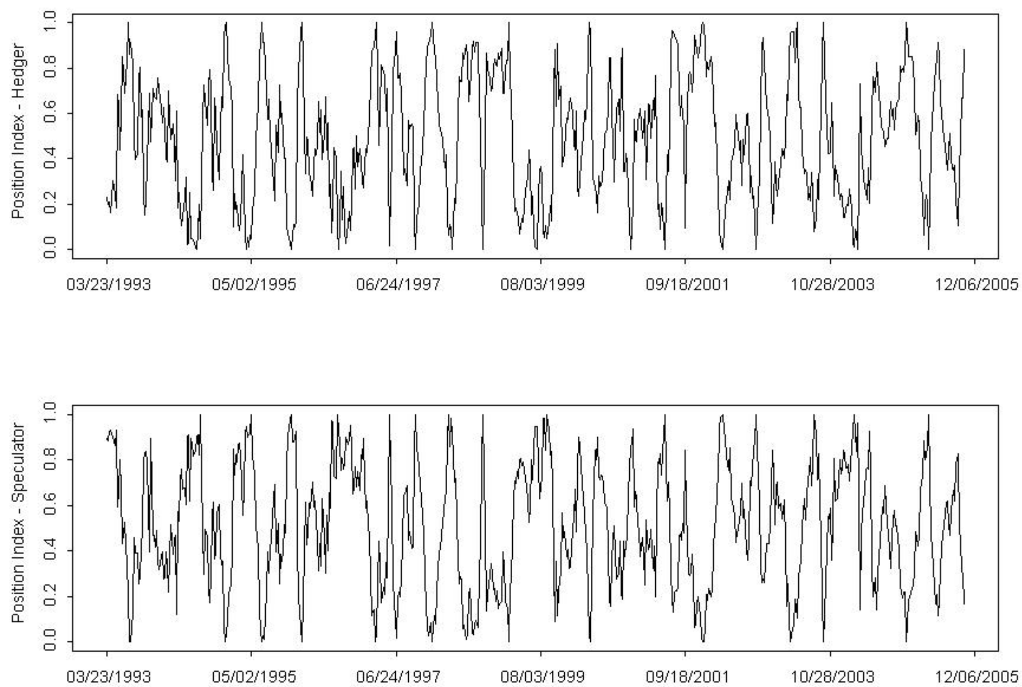


Figure 6: Trader Position Index, $P_t^C(\tau)$, for Hedgers and Speculators with $\tau = 39$ Weeks. The position index is a measure of how long a trader class relative to the maximum and minimum net long positions of that class over the centered window of length τ weeks. A value of 1 indicates that a new τ-period net long maximum has been reached. A value of 0 indicates that a new τ-period net long minimum (a net short maximum) has been reached. The upper panel is the position index for large hedgers, and the lower panel is the position index for large speculators. Values are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006, as reported in the CFTC Commitments of Traders Reports.

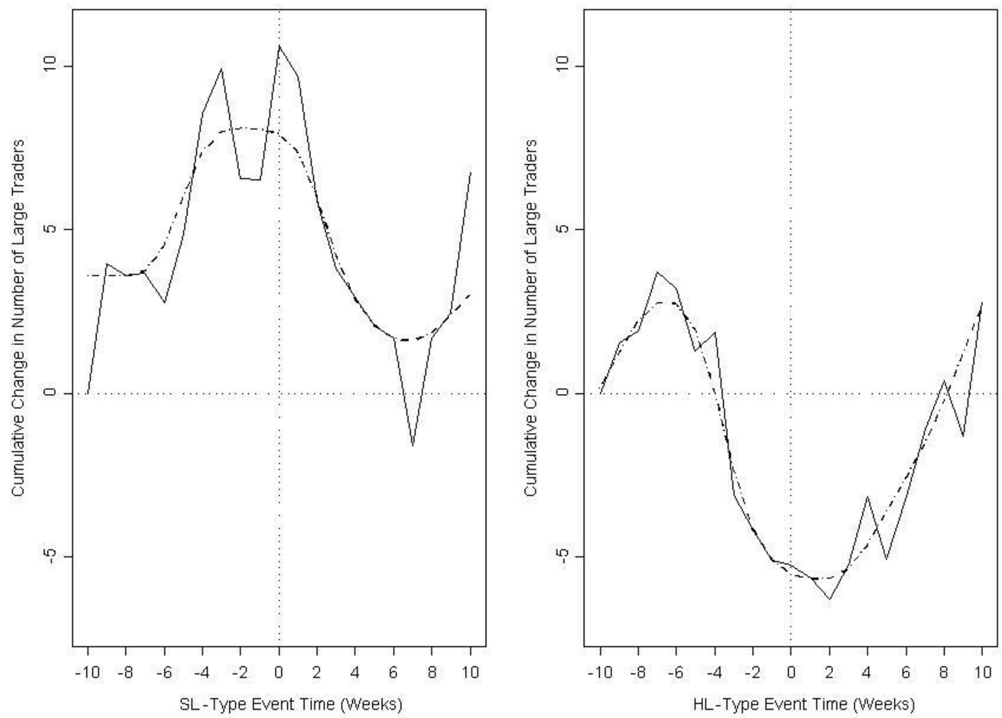


Figure 7: Averaged Cumulative Change in the Number of Large Traders Around Dates when Speculators and Hedgers Have Extreme and Opposing Positions. *SL*-type events are when speculators are extremely long and hedgers extremely short. *HL*-type events are when speculators are extremely short and hedgers extremely long. The dashed line is a smoothed version of the series, computed by using a method of running medians known as 4(3RSR)2H with twicing. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. The left panel shows that traders are entering the market before an *SL*-type event, and then leave the market afterwards. The right panel shows that traders are leaving the market before and slightly after an *HL*-type event, and then begin entering again. The statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. The number of traders in the market is reported in the CFTC Commitments of Traders Reports.

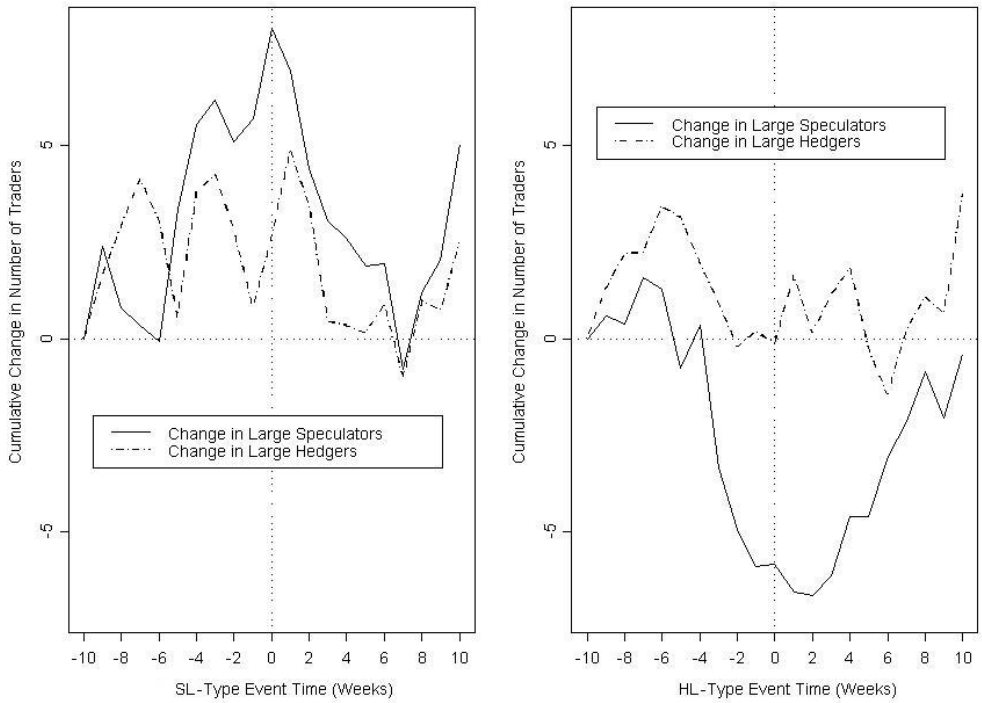


Figure 8: Averaged Cumulative Change in the Number of Speculators and Hedgers Around Dates when Speculators and Hedgers have Extreme and Opposing Positions. SL-type events are when speculators are extremely long and hedgers extremely short. HL-type events are when speculators are extremely short and hedgers extremely long. The solid and dashed lines represent the cumulative change in the number of speculators and hedgers, respectively. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. The left panel shows that traders are entering the market before an SL-type event, and then leave the market afterwards. The right panel shows that traders are leaving the market before and slightly after an HL-type event, and then begin entering again. The statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. The number of traders in the market is reported in the CFTC Commitments of Traders Reports.

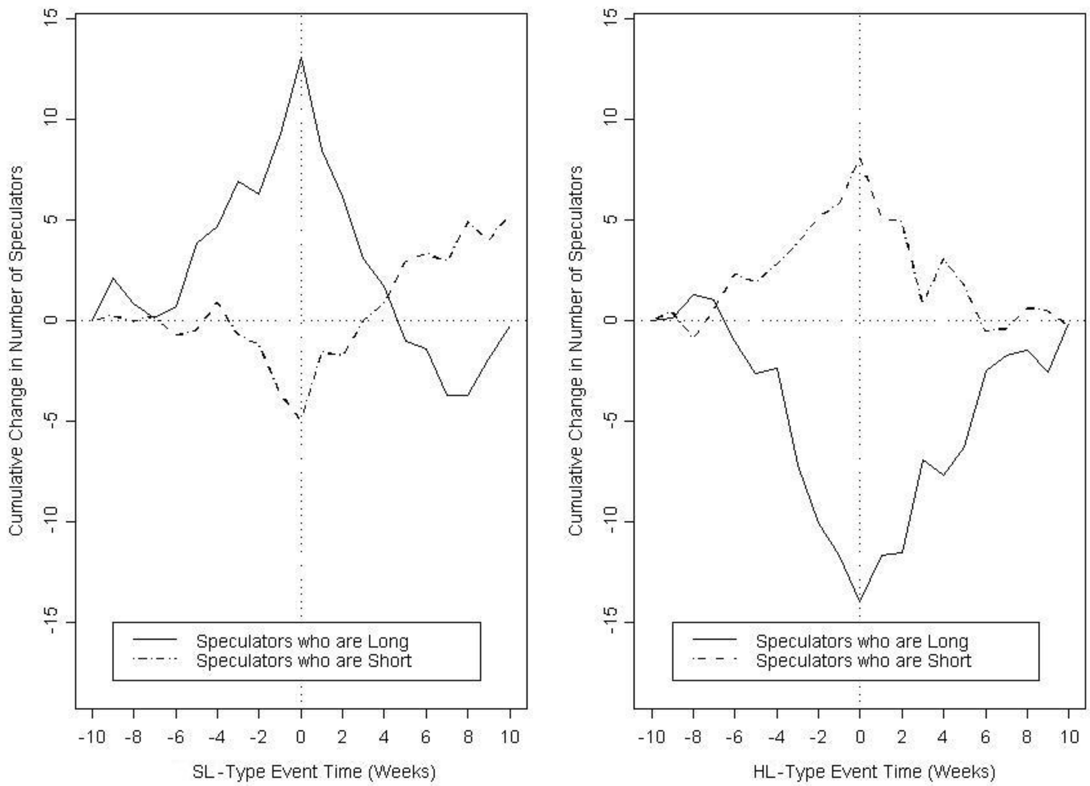


Figure 9: Averaged Cumulative Change in the Number of Speculators who are Long Versus who are Short Around Dates when Speculators and Hedgers have Extreme and Opposing Positions. SL-type events are when speculators are extremely long and hedgers extremely short. HL-type events are when speculators are extremely short and hedgers extremely long. The solid and dashed lines represent the cumulative change in the number of speculators long and short, respectively. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. The dotted vertical line indicates the event time. The averaged cumulative change in the total number of large hedgers. The statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. The number of traders in the market is reported in the CFTC Commitments of Traders Reports.

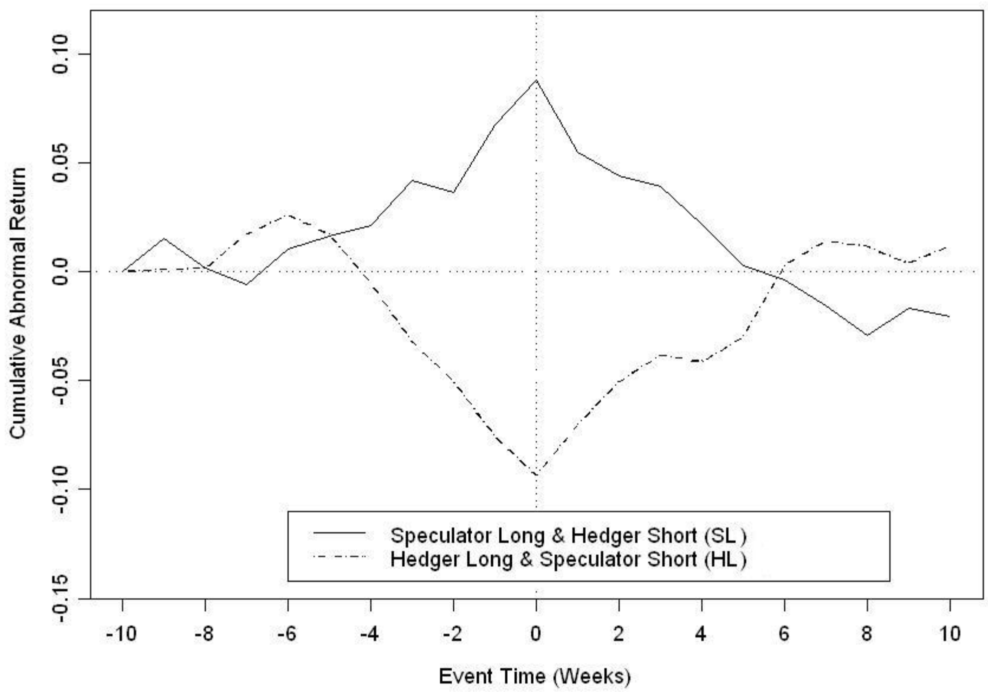


Figure 10: Averaged Cumulative Abnormal Returns. Returns are assumed to have a constant mean. Time relative to the event date, measured in weeks, are on the horizontal axis. Events at time zero (indicated by a vertical dotted line) are extreme opposing positions by large hedgers and speculators. When speculators are long and hedgers are short (solid line), the event date coincides with a local price top. When speculators are short and hedgers long, the event date coincides with a local price bottom. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Within the sample period, there are 20 extreme events of each type for a total of 40 extreme events. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

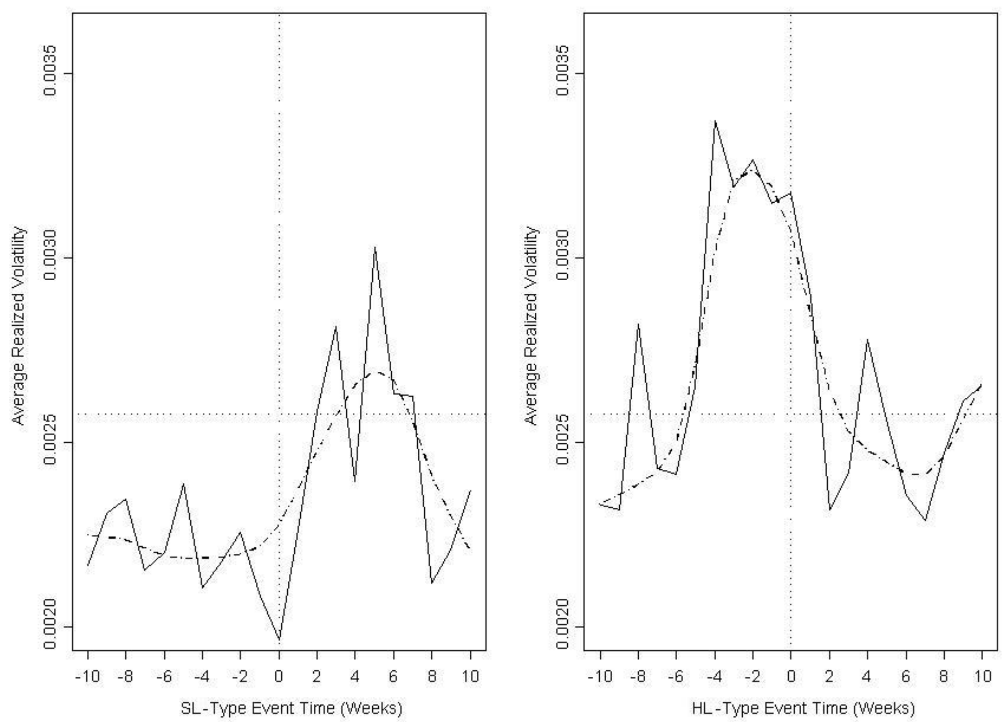
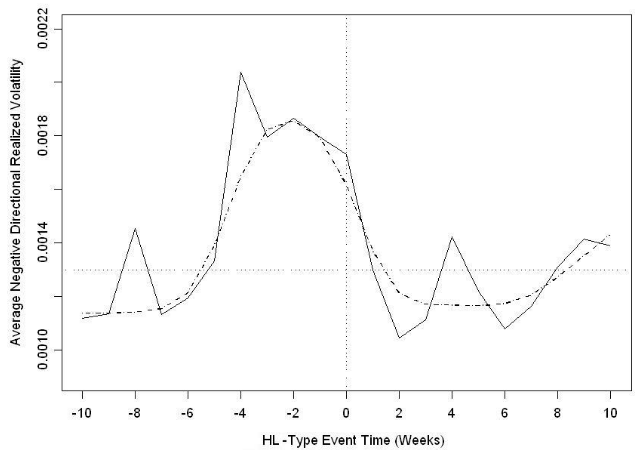
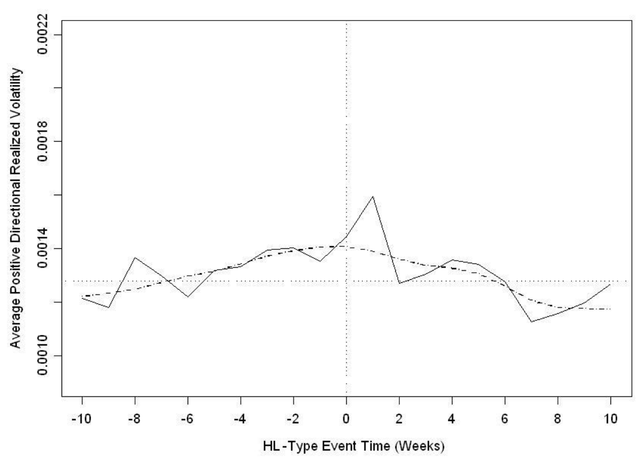


Figure 11: Average Realized Volatility Around Type SL-Type Event Dates and Type HL-Type Event Dates. SL-type events are when speculators are extremely long and hedgers extremely short. HL-type events are when speculators are extremely short and hedgers extremely long. The dashed lines are a smoothed versions of the average realized volatility computed by using a method of running medians known as 4(3RSR)2H with twicing. The dotted horizontal line is the average realized volatility over the entire sample. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.



(a)



(b)

Figure 12: Average Directional Realized Volatility Around Type HL-Type Event Dates. HL-type events are when speculators are extremely short and hedgers extremely long. Panel (a) plots the average negative directional realized volatility, and panel (b) plots the positive directional realized volatility. The dashed lines are a smoothed versions of the average realized volatility computed by using a method of running medians known as 4(3RSR)2H with twicing. The dotted horizontal lines in the respective panels are the average negative and positive directional realized volatility over the entire sample. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. Note that the spike in panel (a) at time -1 week is spurious and is caused by a single large value. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

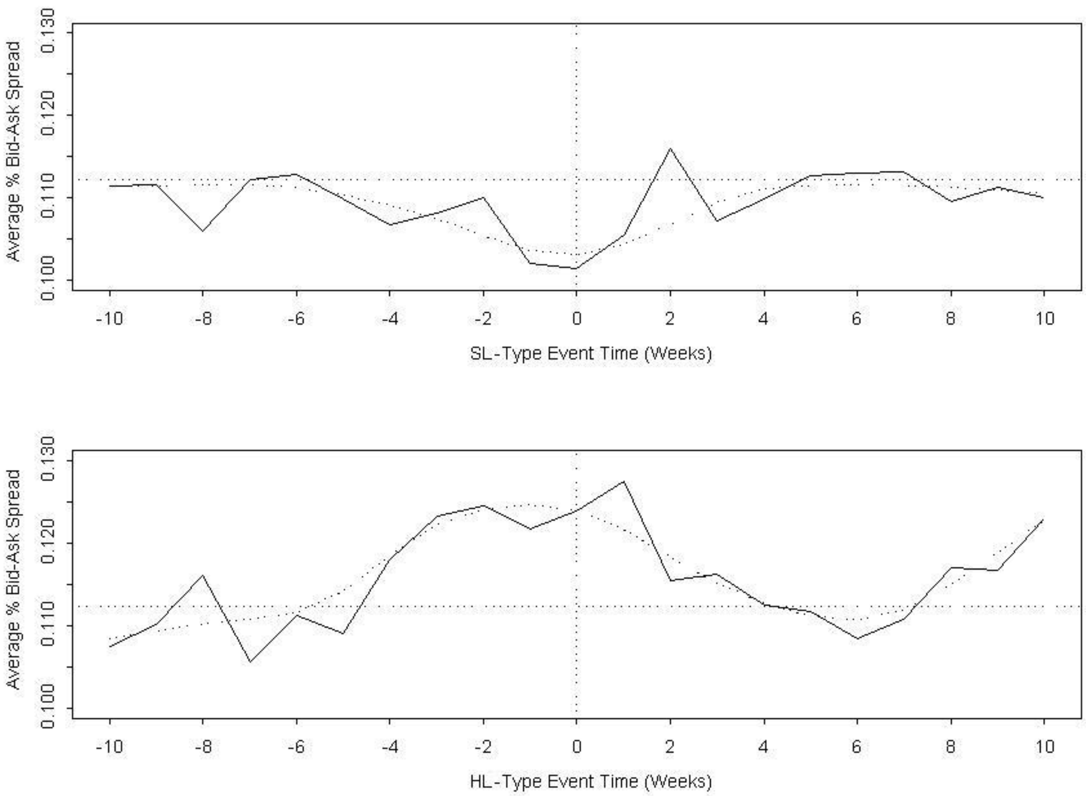


Figure 13: Averaged Daily Percentage Bid-Ask Spread Around Event Dates. Spreads were estimated from transaction price data. Time relative to the event date, measured in weeks, are on the horizontal axis. Events at time zero (indicated by a vertical dotted line) are extreme opposing positions by large hedgers and speculators. The dotted horizontal line indicates the average daily percentage bid-ask spread from all observations. The dashed lines are a smoothed versions of the average daily percentage bid-ask spread computed by using a method of running medians known as 4(3RSR)2H with twicing. Statistics are calculated from 698 weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.

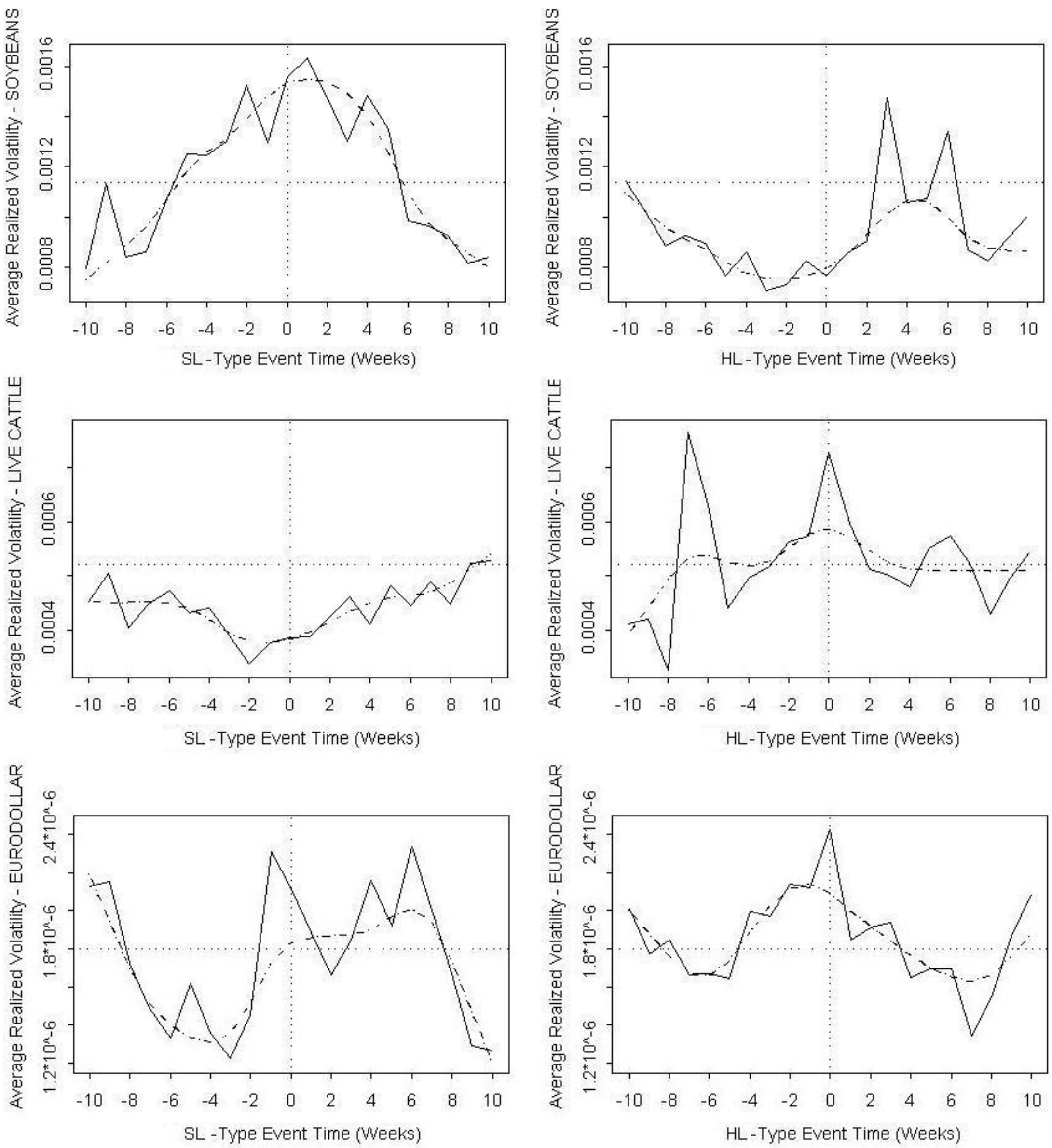


Figure 14: Average Realized Volatility Around Type SL-Type Event Dates and Type HL-Type Event Dates for Three Markets. SL-type events are when speculators are extremely long and hedgers extremely short. HL-type events are when speculators are extremely short and hedgers extremely long. The dashed lines are a smoothed versions of the average realized volatility computed by using a method of running medians known as 4(3RSR)2H with twicing. The dotted horizontal line is the average realized volatility over the entire sample. Times relative to the event date, measured in weeks, are on the horizontal axis from 10 weeks before traders take an extreme position until 10 weeks after. Statistics are calculated from weekly observations from September 30, 1992 until February 28, 2006. Trader positions are reported in the CFTC Commitments of Traders Reports and the weekly returns are calculated from tick-level price data from TickData, www.tickdata.com.