Received: 20 November 2019 | Accepted: 20 November 2019

DOI: 10.1002/fut.22079

Check for updates

RESEARCH ARTICLE

FUTURES MARKETS WILEY

Trading and information in futures markets

Guillermo Llorente[1] | **Jiang Wang**[2]

[1]Facultad de C. Económicas, Universidad Autónoma de Madrid, Madrid, Spain

[2]Sloan School of Management, CAFR and NBER, Massachusetts Institute of Technology, Cambridge, Massachusetts

Correspondence
Guillermo Llorente, Facultad de C. Económicas, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
Email: guiller@uam.es

Funding information
Ministerio de Economía y Competitividad, Grant/Award Numbers: ECO02012-32554, ECO2017-85356-P; Fundación BBVA

Abstract

This paper studies the trading behavior of different types of traders (customer type indicators [CTI's]) in corn futures. Nonmembers (CTI4) consume most of the intraday liquidity while local traders (CTI1) as market makers are its main provider. Both groups combine most of the intraday trading volume. Interday trading comes mainly from proprietary accounts (CTI2) and other local traders' trades (CTI3), reflecting their longer-term needs for hedging and speculation. Changes in the overnight positions of the general public (CTI4) and clearing members (CTI2) contribute mostly to daily price discovery, while the positions of CTI3 group reflect possible information advantage about future price movements.

KEYWORDS

CTI, futures, liquidity, price discovery

JEL CLASSIFICATION

G10; G12; G13; G14; G18

1 | INTRODUCTION

A financial market allows different market participants to meet their trading needs. In doing so, it serves the purpose of liquidity provision and price discovery. Despite a body of theoretical analysis, our understanding of how the market performs these functions remains limited, especially empirically.[1] This is mainly due to the lack of comprehensive and detailed information on who trade in the market, why they trade, and how they trade. This type of information is often proprietary, even at relatively aggregated levels. In this paper, we utilize a unique data from the futures market, which categorizes trades according to trader types, to analyze the trading behavior of each type. In particular, we are able to characterize their trading behavior, their gains and losses, their role in liquidity provision and price discovery, and their potential informational advantage. These results allow us to gain more insight into how the futures market functions and serves its participants as a financial market, different from the usual "commodity" perspective.

The Liquidity Data Bank (LDB) compiled by the Chicago Board of Trade (CBOT)/Chicago Mercantile Exchange (CME) categorize each trade by its Customer Type Indicator (CTI).[2] It identifies each trade according to four customer types: for a member's own account (CTI1), a commercial clearing member's proprietary accounts (CTI2), another member's own account (CTI3), or a customer (CTI4). Roughly speaking, CTI1 represents market makers, CTI2

[1]Theoretical analysis often relies on somewhat abstract and simplified formulation of the heterogeneity in investors' trading needs and information. Grossman (1976), Grossman and Stiglitz (1980), Kyle (1985) provide the basic analytical framework in analyzing the price discovery process in financial markets. Grossman and Miller (1988), among others, consider the underlying mechanism for liquidity consumption and provision. Wang (1994) describes a dynamic model incorporating both the needs to trade for risk-sharing/liquidity and for information motivated speculation. See also Vayanos and Wang (2012) for a survey on the related theoretical literature.

[2]During the period of our sample, the data was compiled and provided by CBOT, which merged with CME to form the CME Group in July 2007.

 © 2019 Wiley Periodicals LLC

represents the proprietary accounts of exchange clearing members, CTI3 represents other member traders, and CTI4 represents the general public (i.e., nonmembers).[3] We only consider corn futures, for the reason that private information may be more important for commodity futures, especially agricultural. From LDB, we can construct the transactions by each CTI group during a day and by cumulating these transactions we can construct their closing positions for each trading day.

We find that for corn futures, the four types of traders exhibit very different trading behavior and play different roles in the market. First, CTI group 1 and 4 constitute most of the intraday trading and maintain little overnight positions. In particular, group CTI1 contributes around 50–60% of the intraday volume and group CTI4 contributes around 30–40%. In comparison, group CTI2 and CTI3 contribute about 5% each. To the contrary, CTI group 2 and 3 contribute the most to interday trading, carrying most of the overnight positions, for about 30–40% each. The contribution to the interday trading volume is minuscule for group CTI1 and less than 10% for group CTI4. We also find that the relative shares of the intraday trading is quite stable for groups CTI1 and CTI4, the two dominant groups, but the relative shares of interday trading is highly variable for the two main contributing groups, CTI2 and CTI3. This pattern of trading behavior suggests that group CTI4 conducts mostly intraday trading, likely to speculate on very short-term price trends. Groups CTI2 and CTI3 are mostly trading for longer-term motivations, such as hedging, market making or speculation. And these trading needs can change substantially over time.

The trading patterns of different CTI groups suggest that group CTI4 is the main consumer of intraday liquidity while group CTI1 is the main provider. As a result, we find that group CTI1 consistently earns profits from its intraday trading while group CTI4 consistently loses in intraday trading. In addition, both the profits of groups CTI1 and CTI4 exhibit strong time consistency, with low variability. The Sharpe ratio for CTI1 group's intraday trading exceeds 0.79, while it is −0.58 for CTI4 traders. Group CTI2 generally breaks even in intraday trading while group CTI3 also loses money on intraday trading, but at a lower level than group CTI4 as it trades significantly less. We further show that the profits and losses of CTI1 and CTI4 groups are positively related to unexpected changes in daily turnover and price volatility.

Using the day to day changes in each CTI group's end of day positions, we also examine its relative importance in price discovery. We start by looking at the contemporaneous correlation between changes in market prices and market-wide variables. We find that predicted change in turnover and imbalance between buy and sell volume exhibit a significant correlation with daily price changes over the same day. After controlling for these variables, we further show that only changes in group CTI4's daily closing positions exhibit additional explanatory power for contemporaneous price changes.

Given the different trading needs of the four CTI groups, we further examine their potential information advantage. We find that changes in the overnight positions of group CTI3 can forecast the price change of the following day. In contrast, changes in positions of other CTI groups have no predictive power for future price movements. Using nonparametric analysis, we further show that changes in CTI3's overnight positions can also predict higher moments of the price change in the following day. In particular, an increase in their overnight position predicts a positive skewness and higher kurtosis, while a decrease predicts negative skewness and lower kurtosis. In addition, while other CTI groups' overnight positions tend to have mixed correlations across contracts with different maturities, group CTI3's positions are significantly positively correlated across maturities. These results suggest that group CTI3 collectively possess private information that is not fully reflected in market prices.

The results above paint an informative picture about the overall trading needs of different CTI groups, how they trade in the market, and their role in liquidity provision/consumption and in price discovery. Clearly, group CTI1 serves as market makers, mainly providing intraday liquidity to other traders, mostly to the general public. Group CTI2 trades for longer-term (more than a day), with time-varying needs. Their trades have limited information content, indicating that they trade mostly for risk-sharing or hedging reasons. Group CTI3 also trades for the longer term, but its trades clearly contain information beyond what is reflected in market prices. Group CTI4, the general public, consists of mostly short term traders. It consumes most of the short-term liquidity, incurring nontrivial loses. However, it is also through group CTI4's trades (at least some of them), market prices move to reflect more fundamental information.

Compared with many other financial markets, more disaggregated data is available for the futures market, partially because of the exchanges' desire to promote transparency and the unique regulatory environment it faces. Many researchers have obtained various types of disaggregated data on trading and analyzed the behavior of different market

[3]The actual description of the CTI is given later in the paper, with more details in Appendix A.2. As of July 2, 2015, the CME closed most of its trading floors, including the corn pit. Nevertheless, the CTI classification continues to apply for electronic trades. See CME Rule Book.

participants. Earlier work primarily focus on intraday trading and liquidity provision. Kuserk and Locke (1993) are among the earliest to use CTI related data to analyze their trading behavior. They have a short sample period (only 3 months), but more detailed information on the trading records of a set of traders within each CTI group for a set of futures.[4] They mainly focus on the short-term trading profit of "scalpers," those in group CTI1 identified as market makers, and find similar results as we do for the whole CTI1 group.[5]

Using a similar data set, Manaster and Mann (1996) further examine the market-making behavior of a sample of scalpers in their intraday trading. In particular, they look at the relationship between market makers' inventory, customer spread, market depth and price variability over short (1-, 5- or 15-min) intervals. They also find that at 1-min intervals, the net flow of customers (CTI4) can predict price changes over the next minute. They did not examine the behavior of CTI2 and CTI3 groups.[6] Wiley and Daigler (1998) use data similar to ours on several financial futures to analyze the daily trading volume for all four CTI groups, in particular, their dynamics and cross-correlation. Daigler and Wiley (1999) further examine the relationship between price volatility and unexpected CTI group's trading volume. They do not separate intraday versus interday trading and their connection with price discovery and information asymmetry.

Brandt, Kavajecz, and Underwood (2007) use the CTI data on Treasury futures to examine the role of different CTI groups in the price discovery process for both the futures and cash market at a daily frequency. Their methodology is similar to ours, as in Evans and Lyons (2002), among others. They find that changes in group CTI4's overnight positions are significantly positively correlated with contemporaneous price changes for Treasury futures, which we also find for the corn futures. Since they bundle together groups CTI2 and CTI3, they find a significant negative correlation between changes in their total overnight positions and price changes. We consider the two groups separately and show that such a relationship is only weakly significant for group CTI2. Our results on the predicative power of group CTI3's position changes for future price changes indicate its distinctive information advantage.

More recently, several authors have conducted fruitful analysis based on new data sets with detailed information on position records. Dewally, Ederington, and Fernando (2013) utilize the data from Commodity Futures Trading Commission's (CFTC's) Large Trader Reporting System (LTRS) to study the interday trading profits of large traders in the futures on energy-related products. The LTRS contains information of the daily closing positions of all large traders whose open positions exceed a certain threshold. The advantage of this data is it contains more information on trade types for these large traders.[7] This information allows a finer grouping of traders according to their "physical" characteristics, such as refiners versus hedge funds and market makers.[8] In fact, Dewally et al. (2013) formed 11 groups from the set of large traders and analyze their interday trading patterns, profits, and possible determinants. Grouping by "physical" characteristics, however, may not best reflect the "financial" characteristics of these traders such as liquidity demanders/providers versus information motivated speculators. The market makers using LTRS characteristics contain mostly CTI1, but also CTI3. The other groups refine but also mix CTI2 and CTI4 groups in the LDB data set. Instead of using the physical characteristics, Fishe and Smith (2012) rely on the performance of different traders to help identifying informed traders and then examine the nature of their potential information, its relationship with trader characteristics and trading profits. This empirical approach can potentially reveal additional information about a trader, but it is also subject to the limitations of statistical inference, including the accuracy of the underlying hypotheses.[9] Based on the complete transaction record on S&P 500 futures, Locke and Onayev (2007) examine the relationship between customer order flow, price change, and dealer predictability at 5-min intervals. They find significant connections among these variables only over a short time horizon, contemporaneously or with 1 or 2 5-min lags.

Several papers, such as Sanders, Irwin, and Merrin (2009), Brunetti, Buyuksahin, and Harris (2016), Buyuksahin and Harris (2011), utilize CFTC's reported daily closing positions from the Commitments of Traders (COT) data to test the Granger causality and led–lag relationships between changes in futures prices and trades. The evidence mostly finds

[4]The data Kuserk and Locke (1993) used is based on the audit transaction trail data from the exchanges, which records all the trades occurred on the exchange floor with the trader identifications. It covers a set of futures, including eight financials and four agricultural. Trader ID can be further grouped into CTI categories. The data also gives the time stamp of each trade up to a minute, with some errors. This data used by researchers typically consists of a random sample from the whole population.

[5]See also Silber (1984) for an earlier study on the trading records of two scalpers in stock index futures, showing their role in liquidity provision at a profit.

[6]Based on the LDB data on 30-year Treasury futures, Menkveld, Sarkar, and van der Wel (2012) show that net customer volume (CTI4) contributes to price discovery over 15-min intervals. They do not consider the order flow of other CTI groups.

[7]See Haigh, Hranaiova, and Overdahl (2005) for a more detailed discussion of the LTRS data set.

[8]The CFTC trader classification is commonly referred to as done by the "business line of activities." Ederington and Lee (2002) discuss the accuracy of this classification.

[9]For example, to make inferences about the intraday performance and its source, Fishe and Smith (2012) have to rely on several hypotheses in measuring performance and its determinants.

that traders' end of day positions do not help to forecast futures price changes. These papers do not control for market variables, which make the results, if any, also a bit hard to interpret.

It is worth noting that CFTC's data provides position data at a given sample frequency (e.g., daily). The position data do not provide information on the prices the positions were established, which limits how much we can learn about the trading behavior of different market participants. In terms of categorization, the CFTC data relies in the market participants' business lines of activities. In contrast, the LDB data we use provides information on actual trades, which CTI group is trading, how much, and at what prices, which allows a more detailed characterization of different group's trading behavior, and their study as agents trading a financial asset. It is difficult to match the CTI classification with descriptions or business lines in the CFTC data.

There is also a rich empirical literature that studies the underlying determinants of futures prices based on the theory of normal backwardation (e.g., Keynes, 1923 and Hicks, 1939), the theory of storage (e.g., Kaldor, 1939 and Working, 1949), theories that relate futures returns to market risks (e.g., Dusak, 1973 and Black, 1976), or combinations of them (e.g., Stoll, 1979, Hirshleifer, 1988, 1989, and de Roon, Nijman, and Veld, 2000). See, for example, Hartzmark (1987), Kucher and Kurov (2014), Erb and Harvey (2006), and Bessembinder (1992), which mostly rely on CFTC's COT data.[10] While it is highly desirable to connect futures prices to economic fundamentals, the results from these empirical studies are mixed and inconclusive, depending upon the sample period, sample frequency, the particular futures markets considered, and measures of fundamentals and systematic risks. Our approach in this paper is mainly empirical, documenting market behavior empirically rather than testing specific theories.

Our analysis complements and expands on the existing literature to study the futures markets as any other financial market. Instead of focusing on a specific aspect of the futures market, such as intraday behavior versus interday, liquidity versus price formation or a particular type of traders, we aim at providing a more broad and comprehensive picture. The four CTI groups, although a bit coarse, are complete and capture meaningful heterogeneity among market participants. We provide an overall characterization of their trading behavior and market impact in terms of trading, liquidity supply, and demand, profits and loses (P&L), price discovery, and information advantage. Although we also confirm some of the existing results in a slightly different context, we deliver several new findings, such as the informational role of CTI3 group and factors driving the cost of liquidity and the price formation process. The potential significance of group CTI3 in the market is particularly intriguing. It was often overlooked in previous analysis and typically bundled with either group CTI1 as floor brokers/traders (FBTs) or with group CTI2 as hedgers.

In sum, the uniqueness of the database allows us to study the futures market as any other financial market, different from the usual "commodity" perspective. The broad picture we present about the market adds important unknown features worth considering in these markets.

The rest of the paper is organized as follows. Section 2 describes the data sets we use and presents some basic summary and characterization of the data. Section 3 analyzes the P&L of each CTI group. Section 4 examines the role of each CTI group in the price discovery process. Section 5 studies the potential information advantage each CTI group may have. We conclude in Section 6.

2 | DATA

In this paper, we focus on corn futures. This choice is motivated by at least two considerations. First, many practitioners believe that there is potentially more private information in the trading and pricing of agricultural futures than other futures. Second, corn is by far the most liquid single grain futures traded on the CME exchanges (previously only the CBOT).[11] Being one of the commodities with the oldest exchange-traded futures (started in January 2, 1877), corn has also become more important as a commodity due to its many uses.[12] Corn's many and diverse usages makes the corn futures market very important not only for agricultural but also for industrial output.

[10]See also Houthakker (1957), Chang (1985), Hartzmark (1991), Bessembinder and Chan (1992), Leuthold, Garcia, and Lu (1994), Wang (2003), Wang (2004), Hong and Yogo (2012), Rouwenhorst and Tang (2012), Gorton, Hayashi, and Rouwenhorst (2013), Kang, Rouwenhorst, and Tang (2014), Singleton (2014), Hamilton and Wu (2015).

[11]For example, for December 2014, the trading volume (in thousand contracts) on the CME was 1,237 for corn futures, 599 for soybean futures, and 503 for wheat futures. The "soybean futures complex" (which includes soybean, soybean meal, and soybean oil) had a volume of 1,295 thousand contracts, slightly higher than corn futures alone.

[12]In addition being used as food, especially as feed for livestock and poultry, either as the main source or as a substitute of wheat and/or soybean, source for corn oil, as well as soft drinks, corn is also used for ethanol, alcohol, antibiotics, aspirin, lubricants agents, metal plating, adhesives, construction materials, cardboard, laminated building products, and so forth.

2.1 | Data sources

Our data comes from two complementary sources both from the CBOT. The basic source of information about the market is from the End Of the Day (EOD) files. For each trading day and for each contract open for trade, this data set provides the information on the open, high, low, close, and settlement prices, open interest at the end of each day as well as daily trading volume (in number of contracts).

The primary source for our analysis is the LDB, which describes the trading activities of different types of traders in the exchange. In particular, LDB assigns each trade according to its Customer Trade Indicator (CTI) into four categories: CTI1, CTI2, CTI3, and CTI4. According to CBOT, each category is defined as follows:[13]

- CTI1: Orders entered/trades executed by an individual member for his/her own account, for an account he/she controls, or for an account in which he/she has an ownership or financial interest. However, transactions initiated and executed by a member for the proprietary account of a member firm must be designated as CTI2 transactions.
- CTI2: Orders entered/trades executed for the proprietary accounts of a member firm or Qualified Affiliate.
- CTI3: Orders entered/trades executed by a member on behalf of another individual member, or for an account such other member controls or in which such other member has an ownership or financial interest.
- CTI4: All orders/transactions not included in CTI categories 1, 2, or 3. These typically are orders entered by or on behalf of nonmember entities.

Each day, for every contract available for trade and for every traded price, LDB provides information on the number of contracts each CTI category buys, sells and its net trade (change in position or buy minus sell), the half-hour time brackets where that price has been negotiated, and the total volume (measured by the number of contracts) traded at that price during the floor trading hours. Only nonspread trades executed in the pit through open outcry or through the electronic trading platform (Globex) are included in our LDB data, where spread traders refer to paired trades in two different maturities.[14] It is worth noting that the LDB data does not follow off-hours trading. This can potentially affect our calculation of overnight positions.[15] Transactions outside the pit, also referred to as "ex-pit" transactions, such as block trades, cleared-only contracts, exchange for physicals (EFP), exchange for risk (EFR), and so forth are not available in LDB (see LDB Guide and CME Group).[16]

Our sample period is from January 3, 1995 to May 8, 2006, with 2,856 days. CBOT did not provide corn futures LDB data before 1995. In the second half of 2006, CBOT went through a structural change in the trading of agricultural futures. On August 1, 2006, CBOT launched electronic trading side by side with the traditional open outcry trading in the pit. Orders could be submitted and executed through the electronic trading system or in the pit on the trading floor. After a period of side by side trading, most of the trading migrated to electronic trading.[17] Given that this change in the trading process may alter the trading behavior of different participants, especially as categorized by the CTI groups, we stay with the data before the change. Combining the two sources of data lead to a total of 2,856 daily observations on prices, trading volume, open interest, and different CTI groups' trading activities.

[13]Detailed CTI definitions by CBOT and their changes over time are given in Appendix A.2. See, for example, Wiley and Daigler (1998) for a discussion of the different CTI categories and the likely type of traders included in each group.

[14]Electronic trading (Globex) became available for daytime trade in September 1998. Data on electronic and spread trades were available to us in LDB form since January 2003. Table A.2 in Appendix A.1 compares the mean daily contracts traded in each venue during this period. Electronic and spread (pit and electronic) trading represents about 7% of the total LDB volume, with LDB pit trading representing approximately the 93% of the total LDB reported traded volume.

[15]CBOT launched the "Project A," an electronic trading platform to trade in after hours, in 1995. It seems that trading in this overnight e-market was quite thin (see, e.g., Leuthold & Kim, 2000).

[16]Ex-pit transactions (also called by CME as privately negotiated transactions, PNT) refer to any transaction that is executed on a noncompetitive basis and outside of a traditional open outcry or electronic trading environment. These transactions are known by a variety of names, depending on the details of the different nature of the transaction. Block trading is an outright trade or spreads trade in futures positions privately negotiated. It is mainly used in the context of very large transactions. Block trades are subject to unique rules. Cleared only contracts is offered on a "cleared only" basis. It is intended to allow transactions executed in an OTC market to be submitted or novated to the Clearing Corporation for purposes of processing and clearance. Exchange for Physicals (EFP) refers to privately negotiated transactions between two counterparties and involves the simultaneous exchange of a futures position for an economically offsetting position. When offsetting is by an over-the-counter derivatives position, it is termed Exchange for Risk (EFR). Other transactions might involve an Exchange for Swaps (EFS). All of these are generally known as Exchange of Futures for Related Positions (EFRPs). See Labuszewski, Nyhoff, Co, and Peterson (2010) for more details about market structure and products/transactions. Ex-pit trading represents a small fraction of the total trading volume. For example, on October 10, 2014, for corn futures traded on the CME, ex-pit transactions represented about 2% of the total number of contracts traded on different venues of CME (in number of contracts, Globex 334,376, Pit 8,065, Ex-Pit 7,068, total traded volume 349,869, Open Interest 1,297,789). On December 31, 2014, ex-pit trading represented 5% of the total number of contracts traded.

[17]In April 2007, 58% of the total CBOT agricultural futures trading volume was conducted electronically and 42% in the pit; in November 2006, corn trading volume on the electronic venue was already higher (above 50%) than in the pit by open outcry; by the end of 2008 electronic trade represented more than 90% of the trading volume in corn futures.

It would be interesting to conduct our analysis using an actual sample and extend it to other futures markets. Electronic trading is already mature. Nowadays most of the trading is electronic. Unfortunately, CBOT stopped providing the LDB data, though the CTI rules are still on place (see Footnote 3).[18]

2.2 | Variable definitions

From the EOD files, we obtain daily observations for each contract open to trade. Since multiple contracts with different maturities are traded on a given day, we use m to denote a contract with a particular maturity. We will set $m = 0$ for the contract in its maturing month (when it exists), $m = 1$ for the contract with the shortest maturity but before its maturing month, $m = 2$ for the contract with the next shortest maturity, and so forth. For each contract m, $m = 0, 1, 2, \ldots$, we then have

$$P_t^{(m)} = \text{ Settlement price for date } t, \tag{1a}$$

$$\mathrm{OI}_t^{(m)} = \text{ Open interest by the close of date } t, \tag{1b}$$

$$\mathrm{HML}_t^{(m)} = \text{ Difference between the highest and the lowest price during date } t. \tag{1c}$$

Market participants use $\mathrm{HML}_t^{(m)}$ as a gauge of daily price volatility. Given the potential nonstationarity in prices, we will use

$$X_t^{(m)} = \frac{\mathrm{HML}_t^{(m)}}{P_t^{(m)}}, \tag{2}$$

as a measure of the relative daily price volatility for contract m. This is in the spirit of range-based measure of volatility suggested by Graman and Klass (1980).[19] It is well known that a contract becomes illiquid as it approaches its maturity.

Figure 1 shows the "life-cycle" of contracts with staggering maturities in their open interests, which are correlated with their liquidity. To avoid potential liquidity issues, sometimes we only include the most liquid contracts in our analysis. As Figure 1 shows, the open interest of a contract drops substantially when it enters its maturing month (and so is its liquidity). Also, for contracts with the longest maturities, their open interests are also very low. Thus, we define $m = 1, 2, 3, 4$ as the most liquid contracts.[20] As the data shows, they represent over 91% of the total open interest and 95% of the total volume in the market.

To simplify our analysis and presentation, sometimes we use the market average for certain variables instead of looking at each individual contract. Factoring in liquidity considerations, we only include the four most liquid contracts. In particular, we have

$$P_t = \frac{1}{4}\sum_{m=1}^{4} P_t^{(m)}, \tag{3a}$$

$$X_t = \frac{1}{4}\sum_{m=1}^{4} X_t^{(m)} = \frac{1}{4}\sum_{m=1}^{4} \mathrm{HML}_t^{(m)}/P_t^{(m)}. \tag{3b}$$

[18]Boyd and Kurov (2012) examined the consequences of the side-by-side pit and electronic trading for group CTI1 from January 2005 through February 2009 in energy futures traded on NYMEX. They find a decrease in the number of CTI1 traders, a decrease of P&L of the surviving CTI1 traders, a decrease in their transaction size, but an increase in the number of trades. Considering the 30-year Treasury futures with the nearest maturity and for CTI1 and CTI4 groups, Menkveld et al. (2012) show a decline in their trading volume after the switch from open out-cry to electronic trading. In addition, the CTI4's trading volume decreases less than that of CTI1. They do not look at CTI2 or CTI3 groups.

[19]Daigler and Wiley (1999) used the Graman and Klass (1980) volatility measure as well as the difference between the high minus low prices to study the influence of CTI's traded volume on market volatility for the silver, munis, MMI index, T-notes and T-bonds futures markets, using daily data from June 1986 to June 1998. The results of their regressions are similar for both measures.

[20]The time series for the considered variables are formed by rolling the contracts the first day of the delivery month to contracts with the next nearest delivery month. Changes in prices are calculated always using prices on the same contract.

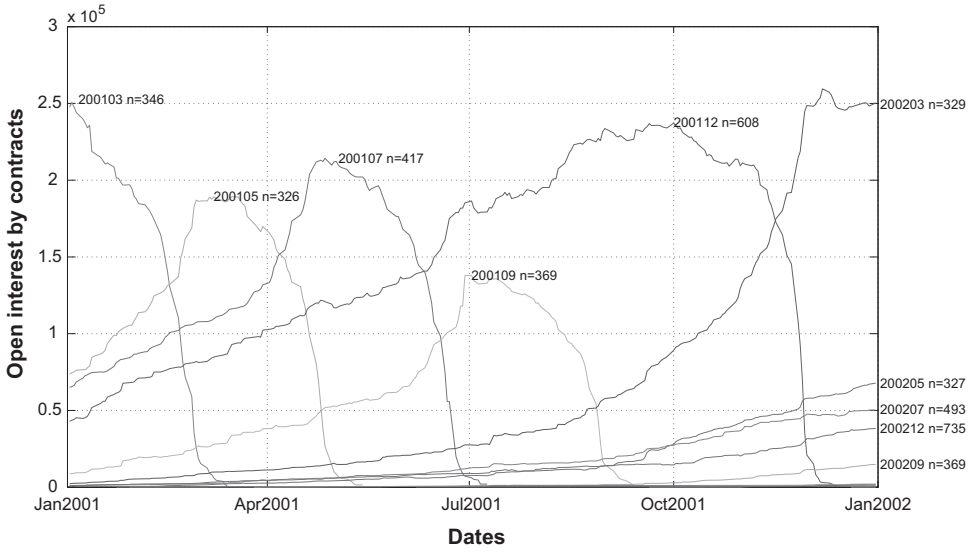

FIGURE 1 Open interest for different maturities. The figure plots the open interest of traded contracts during the year of 2001. The maturities of these contracts are March, May, July, September, December of 2001 and March, May, July, September and December of 2002, respectively

From the LDB data, for each contract we can obtain the buy and sell transactions of each CTI group for each transaction price during the day and the time intervals they belong. For date t and contact m, let $F_{tk}^{(m)}$ denote the kth transaction price, where $k = 1, ..., K_t$. We then have, for each CTI group i and contract m,

$$F_{tk}^{(m)} = \text{The } k\text{th transaction price during date } t, k = 1, ..., K_t, \tag{4a}$$

$$B_{i,tk}^{(m)} = \text{Number of contracts bought at transaction price } F_{tk}^{(m)} \text{ by CTI group } i, \tag{4b}$$

$$S_{i,tk}^{(m)} = \text{Number of contracts sold at transaction price } F_{tk}^{(m)} \text{ by CTI group } i. \tag{4c}$$

It is important to recognize that LDB does not provide the precise information on the time of a transaction. It only gives the time intervals, in half-hour brackets, during which a trade at a given transaction price occurred.[21] Thus, the trade index here does not really describe the sequencing of trades within a bracket. Since the time interval of our analysis is over a day, we further ignore the time bracket of a trade.

Following the convention established above, we will add subscript i for CTI group i, $i = 1, 2, 3, 4$, and superscript (m) for contract m, $m = 0, 1, 2, 3, 4, ...$, to these and other variables when needed.

From the transaction data, we construct several additional variables on each CTI group's trading activity. In particular, we can define the change in CTI group i's position in a given contract m in date t by

$$\Delta N_{i,t}^{(m)} = \sum_k \left(B_{i,tk}^{(m)} - S_{i,tk}^{(m)}\right), \quad i = 1, 2, 3, 4. \tag{5}$$

If we start from the beginning of trading for a contract, we can further construct the cumulative position in a contract by a CTI group at the close of each trading day by

$$N_{i,t}^{(m)} = \sum_{s=0}^{t} \Delta N_{i,s}^{(m)}, \quad i = 1, 2, 3, 4. \tag{6}$$

[21]There may be multiple transactions at a given price in a day, falling into different half-hour time intervals. The LDB data does not specify which one of these intervals a particular transaction belongs.

We also define the sum of changes in positions over all contracts, that is, $m = 1, 2, 3, 4, \ldots$, as a measure of change in each CTI group's net positions

$$\Delta N_{i,t} = \sum_m \Delta N_{i,t}^{(m)}. \tag{7}$$

Adding up the changes in a CTI group's net positions will yield the cumulative net position of the group at the close of each trading day:

$$N_{i,t} = \sum_m N_{i,t}^{(m)} = \sum_m \sum_{s=0}^{t} \Delta N_{i,s}^{(m)}, \quad i = 1, 2, 3, 4. \tag{8}$$

We can also compute the trading volume of each CTI group from LDB transactions for each contract, by the number of contracts and dollar amount, respectively, as follows:

$$V_{i,t}^{(m)} = \sum_k \left(|B_{i,tk}^{(m)}| + |S_{i,tk}^{(m)}| \right), \quad V_{i,t}^{\$(m)} = \sum_k \left(|B_{i,tk}^{(m)}| + |S_{i,tk}^{(m)}| \right) F_{tk}^{(m)}. \tag{9}$$

Summing over the four groups gives the trading volume for a given contract:

$$V_t^{(m)} = \frac{1}{2} \sum_{i=1}^{4} V_{i,t}^{(m)}, \quad V_t^{\$(m)} = \frac{1}{2} \sum_{i=1}^{4} V_{i,t}^{\$(m)} \tag{10}$$

where the coefficient 1/2 corrects for double counting.

It is worth noting that even though EOD files also provide data on daily trading volume, we do not use it for our volume measure. This is because the volume number in EOD files also contains electronic transactions (outright and spread trades), pit spread trades and ex-pit transactions (block trades, EFP, EFR, etc.), which are not included in our entire LDB sample. To be consistent, we use the LDB data to construct a better-matched volume measure. Moreover, in our LDB volume data, all trades are competitively determined and netted trades, for each buy-side, there is one sell-side. The difference between the volume measure from LDB data and EOD data is relatively small. The two measures are highly correlated (the correlation coefficient is 0.983). Appendix A.1 provides a more detailed comparison between these two measures. More important, our results do not depend on the choice between the two measures. Following this procedure and staying as close as possible to one consistent source of data, we built the $\mathrm{HML}_t^{(m)}$ data from the LDB files. The correlation between this series and the one from the OED file is 0.98. Note also that the prices used in the $\mathrm{HML}_t^{(m)}$ variable are transaction prices with nontrivial volume attached—they are not quoted prices.

Summing over all contracts give us a measure of volume for the overall market

$$V_t = \sum_m V_t^{(m)}, \quad V_t^{\$} = \sum_m V_t^{\$(m)}. \tag{11}$$

Using the data of daily open interest from the EOD files, we can also define a measure of turnover for the whole market by

$$\tau_t = \frac{V_t}{\mathrm{OI}_t}, \quad \mathrm{OI}_t = \sum_m \mathrm{OI}_t^{(m)}. \tag{12}$$

In a similar way to Fishman and Longstaff (1992), we also compute each CTI group's profit and loss (P&L) from its intraday transactions in a given contract m by:

$$\mathrm{PNL}_{i,t}^{0\,(m)} = -\sum_k \left(B_{i,tk}^{(m)} - S_{i,tk}^{(m)} \right) F_{tk}^{(m)} + \Delta N_{i,t}^{(m)} P_t^{(m)}, \tag{13}$$

TABLE 1 Descriptive statistics for the market variables

	Mean	*SD*	Skew	Kurt	ρ_1	ρ_2	ρ_5	ρ_{10}
P_t	2.529	0.425	1.499	5.091				
$OI_t(\times 1,000)$	460.559	162.867	1.917	7.567				
$X_t(\%)$	1.526	0.793	1.551	6.636	0.373 (.000)	0.381 (.000)	0.326 (.000)	0.292 (.000)
$V_t(\times 1,000)$	51.112	23.553	1.575	7.320	0.510 (.000)	0.443 (.000)	0.350 (.000)	0.302 (.000)
$\tau_t(\%)$	11.576	5.124	1.506	6.532	0.471 (.000)	0.396 (.000)	0.292 (.000)	0.248 (.000)
Z_t	−0.007	0.054	0.059	3.601	0.086 (.000)	0.060 (.001)	0.002 (.903)	0.011 (.556)
$\Delta \ln P_t$	−0.000	0.013	0.185	4.800	0.053 (.004)	0.019 (.305)	−0.051 (.006)	0.005 (.808)
ΔX_t	−0.000	0.009	0.243	5.716	−0.506 (.000)	0.029 (.121)	0.014 (.464)	−0.007 (.726)
$\Delta\tau_t(\%)$	0.000	5.265	0.519	6.663	−0.428 (.000)	−0.015 (.421)	0.018 (.339)	0.003 (.878)

Note: The descriptive statistics for the daily average of the market variables are shown. ρ_i denotes the autocorrelation coefficient for lag i. The sample period is 1995–2006. The p values are provided inside the parenthesis.

where $P_t^{(m)}$ is the settlement price of the m contract for date t. This definition takes into account the mark to market for futures positions taken during the date. It, however, does not include the P&L from the mark to market on remaining positions from the previous day, which is $N_{i,t-1}^{(m)}(P_t^{(m)} - P_{t-1}^{(m)})$. We call this part the interday P&L:[22]

$$\mathrm{PNL}_{i,t}^{1\,(m)} = N_{i,t-1}^{(m)}\left(P_t^{(m)} - P_{t-1}^{(m)}\right). \tag{14}$$

Thus, the total P&L (for a given CTI group and contract) is given by

$$\mathrm{PNL}_{i,t}^{(m)} = \mathrm{PNL}_{i,t}^{0\,(m)} + \mathrm{PNL}_{i,t}^{1\,(m)}. \tag{15}$$

Summing over all contracts, we obtain the total P&L for each CTI group

$$\mathrm{PNL}_{i,t} = \sum_m \mathrm{PNL}_{i,t}^{(m)}. \tag{16}$$

In addition, we introduce a variable that measures the daily "imbalance" in the trading of a contract between the buy- and sell-side. In particular, on each trading day for each contract, we first divide the range of transaction prices into three intervals with an equal number of distinctive prices. We then compute the dollar volume for prices in each of the intervals. Taking the difference between the dollar volume in the top and the bottom interval reflects the imbalance in trading on the buy- and sell-side. We sum this imbalance in dollar volume over the four most liquid contracts, that is, $m = 1, 2, 3, 4$, and normalize it by the total dollar volume to arrive at our measure of market-wide imbalance in trading:

$$Z_t = \frac{1}{4}\sum_{m=1}^{4} \frac{V_t^{\$(m)}\,(\text{top price interval}) \; - V_t^{\$(m)}\,(\text{bottom price interval})}{V_t^{\$}}. \tag{17}$$

[22]Hartzmark (1987) noted that this is the procedure used by the central clearing house to mark each trader's account to the market price at the end of each trading session: multiply the end-of-day position by the change in the settlement price.

TABLE 2 Correlations among the changes in the market variables

	$\Delta \ln P_t$	$\Delta\tau_t$	ΔX_t	Z_t
$\Delta \ln P_t$	1.000 (.000)	0.006 (.742)	0.046 (.016)	0.425 (.000)
$\Delta\tau_t$		1.000 (.000)	0.667 (.000)	0.004 (.828)
ΔX_t			1.000 (.000)	0.043 (.023)
Z_t				1.000 (.000)

Note: The correlations among the changes in the market variables are shown. The sample period is 1995–2006. The *p* values are provided inside the parenthesis.

2.3 | Data summary

In this section, we present a brief summary of the data. In Table 1, we provide some summary statistics for several of the variables defined above. The *p* value is provided in the parenthesis as in all tables in this paper. Clearly, the daily change in the average log settlement price has a sample mean of zero and daily volatility of 1.3%. It also exhibits a mild positive skewness and a substantial kurtosis, which reflects the fat-tails in its distribution. In addition, the change in log price has a weak but statistically significant serial correlation at 1-day lag and a seasonal 5-day lags, but the correlation dies off with more lags. This weak serial correlation will be taken into account in our future analysis of the price formation process.

The normalized difference between the high and low prices during a day, $X_t = \frac{1}{4}\sum_{m=1}^{4} \mathrm{HML}_t^{(m)}/P_t^{(m)}$, roughly reflects the daily price variation (with a sample mean of 1.5%). It exhibits similar positive skewness and kurtosis as the change in log prices. It also exhibits a strong serial correlation, which persists even after many lags. For example, even at 10-day lags, the serial correlation of X_t remains at 0.292 and highly significant. This is consistent with the general property of the financial market that price volatility is highly persistent over time. However, when we look at the changes in X_t, that is ΔX_t, the serial correlation is significantly negative at one lag, but becomes insignificant afterwards.

The average level of daily turnover in the sample is 11.576%, which means that daily trading volume is on average over 11% of the end-of-day open interest. This clearly suggests that in this market there is a high level of intraday trading. As is typically the case for measures of trading activities, τ_t exhibits positive skewness and kurtosis. Moreover, our turnover measure has a persistent serial correlation. However, when we look at changes in the turnover, it has a strong serial correlation at only one lag, which is negative, reflecting a strong mean reversion tendency in the level of turnover. The serial correlation becomes insignificant even at the 2-day lag.

The measure of trading imbalance, Z_t, has a small and negative sample mean. It exhibits no skewness but a nontrivial kurtosis. It is positively serially correlated but only at 1- and 2-day lags. The correlation, even though statistically significant, is quite small.

In Table 2, we report the correlation between the main market variables we will use in future analysis. It is not surprising that there is a strong correlation (0.667) between turnover τ_t and normalized high minus low prices. It is well known that turnover and price volatility are related in the futures market (as well as in other financial markets).[23] We also see a significant correlation between changes in log settlement price and trading imbalance, which is 0.425.

As Table 1 shows, the market variables such as turnover and normalized high minus low prices exhibit significant persistence. This persistence still exists even for first-order differences. In our future analysis, we will use the first differences to describe the dynamics of these market variables. Moreover, we use an MA(2) model to characterize the dynamics of the first differences of these market variables and to decompose them into their expected and unexpected parts.[24] See Appendix A.3 for the details of these decompositions.

[23]Bessembinder and Seguin (1993) examined eight futures, including two currencies (Deutsche mark and Japanese yen), two metals (gold and silver), two agricultural products (cotton and wheat) and two financials (T-bonds and T-bills), using daily data for the period of May 1982 through March 1990, and found that their price volatility is positively related to volume, both its expected and unexpected parts. Daigler and Wiley (1999) found similar results with CTI4 trading volume for silver, MMI index, Munis, T-Notes, and T-Bonds futures, using daily data from June 1986 to June 1998.

[24]Bessembinder and Seguin (1993) were among the first to show the strong relationship between volume and volatility when considering their unexpected parts, hence the importance of doing so. Latter, Daigler and Wiley (1999) used a similar approach when using the CTI trading volume as an explanatory variable for market volatility.

TABLE 3 Share of open interest and share of volume by the four nearest contracts

Contract (%)	Mean	*SD*	Skew	Kurt
Share of open interest by contracts				
M1	38.651	15.986	−0.224	2.244
M2	29.368	14.894	0.737	2.465
M3	15.249	11.406	1.553	4.726
M4	8.447	7.965	1.706	5.315
Share of volume by contracts				
M1	51.136	21.682	−0.419	1.736
M2	27.870	21.908	1.002	2.539
M3	11.156	14.015	2.386	8.216
M4	5.142	7.614	2.822	12.794

Note: The descriptive statistics for the share of open interest ($OI_t^{(m)}/OI_t$) and share of volume ($V_t^{(m)}/V_t$), by the four nearest contracts (m = M1, M2, M3, and M4) are shown. The sample period is 1995–2006.

The discussion above focuses mostly on market-wide variables. In particular, we aggregate across contracts with different maturities and over different CTI groups. In the following sections, we will examine in detail the behavior of different CTI groups and their connection with prices. Here, we provide some information on the distribution of trading over contracts with different maturities.

Table 3 reports some summary statistics for the share of total open interest and trading volume of the four most liquid contracts, $m = 1, 2, 3, 4$, denoted by M1, M2, M3, and M4, respectively in the table. Not surprisingly, the most liquid contract, which is also the contact with the nearest maturity (excluding the contract soon to mature), has over 38% of the total open interest. The next maturity has about 29% and it tapers off quite quickly. The fourth contract ($m = 4$) has about 8%. The distribution of trading volume across contracts with different maturities exhibits the same pattern. The nearest contract contributes 51% to the total trading volume. The next contract contributes around 27%. The volume share drops to around 11% and 5%, respectively, for the next maturities. Thus, as it was already mentioned, the first four maturities represent over 91% of the total open interest and over 95% of the total trading volume in the corn market futures. The above distribution of trade among maturities is the reason why in our analysis of price discovery process in this market, we will primarily focus on either the nearest contract or the four nearest contracts.

3 | TRADING AND PROFITS OF CTI GROUPS

3.1 | Trading behavior

We now examine the trading behavior of the four CTI groups. From the LDB, we can construct the trades of each CTI group in each contract. We can then compute the daily trading volume and the net change in positions. The former gives us a picture of the relative share of each group in trading while the latter describes its net overnight position-taking.

In Figure 2, we plot the relative share of the total trading volume (including all contracts) of each CTI group. It is obvious that group CTI1 has the largest share of the total volume. Over the sample period, it typically has more than 50% of the volume. Group CTI4 is the next most active in trading, contributing to between 30% and 40% of the total volume. Both groups CTI2 and CTI3 have typically less than 10% of the total volume each. Group CTI2 has less than 5% share of the volume until 2004 when the volume share jumped close to 10%.[25]

[25]For the five futures and the 2-year period (June 1986 through June 1988) they cover, Daigler and Wiley (1999) report (Table I) the following volume breakdowns by CTI groups (means calculated for this footnote): 50% for CTI1, 16% for CTI2, 7% for CTI3, and 27% for CTI4. Fishman and Longstaff (1992) report that the CTI1 represent nearly 60% of the total volume for the 15 random days on the last quarter of 1998 in the soybean market. Manaster and Mann (1996) look at all the futures traded on CME during the first half of 1992 and find that CTI1 represents 47% of the total volume, CTI2 7%, CTI3 6%, and CTI4 40%.

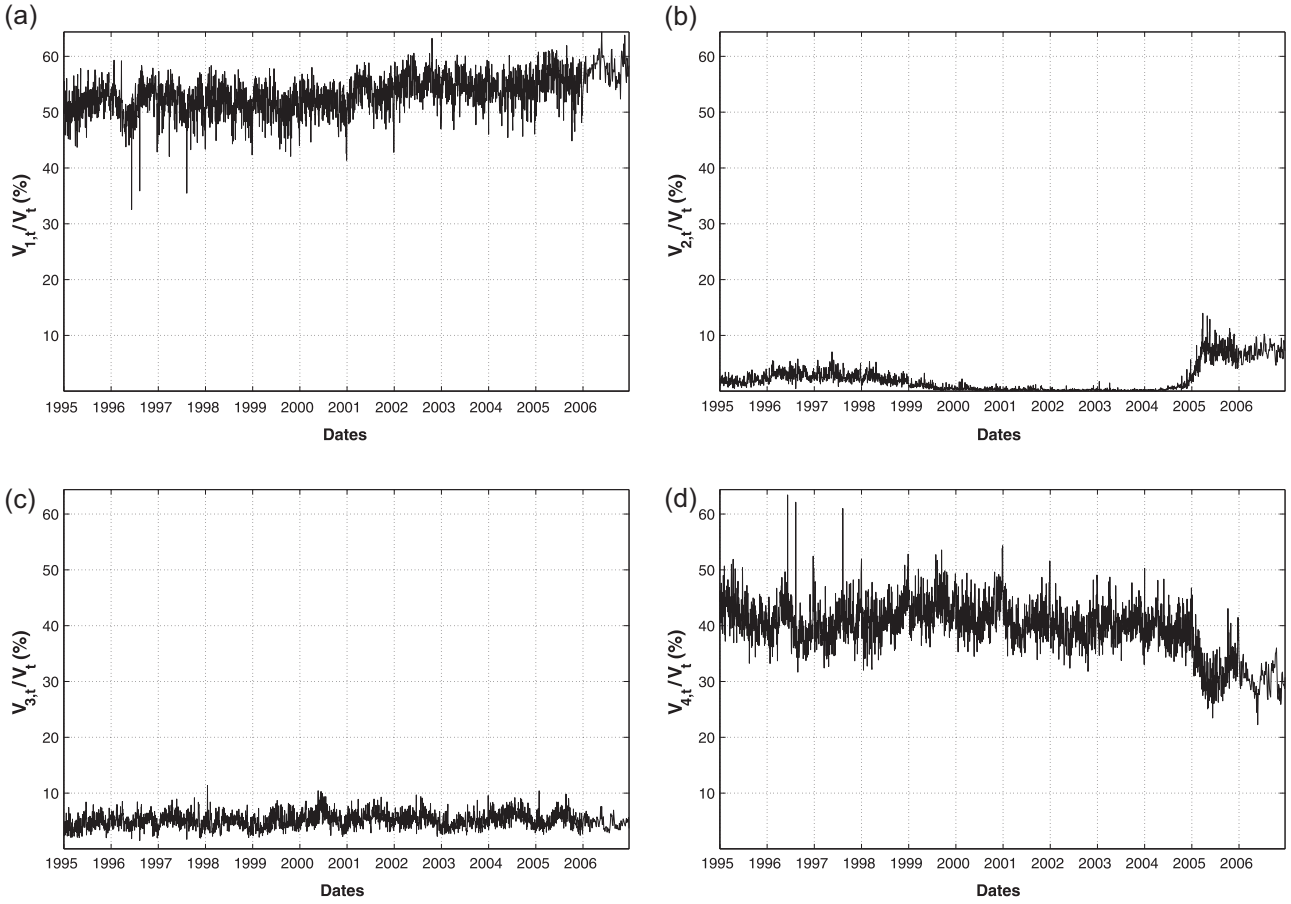

FIGURE 2 Share of trading volume by CTI groups. This figure shows the total trading volume of each CTI group ($V_{i,t}$) normalized by the market volume (V_t) in percentage terms, over the sample period. (a) CTI1, (b) CTI2, (c) CTI3, and (d) CTI4. CTI, customer type indicator

It is worth noting that over the same period, the trading of group CTI2 has increased and that of CTI4 has decreased accordingly. This change is in part due to some of the reclassification of trades between the groups during 2004.[26]

Among the trades during a day by a CTI group, some are offsetting with each other. This part of the trading can be labeled as "intraday" trading (some authors name them as "round trip trading"). The change in the net position from the previous day, that is, $\Delta N_{i,t}^{(m)}$, can then be labeled as "interday" trading. We refer to its absolute value as the interday trading volume. The breakdown of a CTI group's into these two parts is indicative of its trading behavior and potential trading motive. Figure 3 plots the interday trading volume by each CTI group, normalized by the group's own daily volume.

The difference across different CTI groups is stark. For group CTI1, most of its trading is intraday. Its interday volume is only a tiny fraction of its total volume.[27] The same is true for group CTI4, although the share of interday trading is slightly higher. For groups CTI2 and CTI3, however, interday trading constitutes a large fraction of their total volume. For group CTI2, often more than 50% of its trading is interday and sometimes it reaches 80% or higher. For group CTI3, the share of interday trading typically ranges between 20% and 50%.[28]

[26]See the Appendix A.2 for the reclassification, or "Harmonization," on the CTI codes on CME and CBOT.

[27]When analyzing the inventory of the market makers (CTI1) in all the futures traded on CME using transaction data for the first half of 1992, Manaster and Mann (1996) point to three possible sources for the "nonzero daily inventory changes." One may be due to CTI3 trades and the other two are potential data errors and the overnight unwinding of positions in Singapore International Monetary Exchange (SIMEX).

[28]Studying the futures energy markets in the NYMEX and using daily open interest positions of individual traders, Dewally et al. (2013) find that the group of "market makers/floor traders" (corresponding to CTI1 or CTI1 plus CTI3 in our grouping) is the one with the highest interday turnover (daily change in positions of the group over its own open interest position). Our findings are consistent but cover to other market participants.

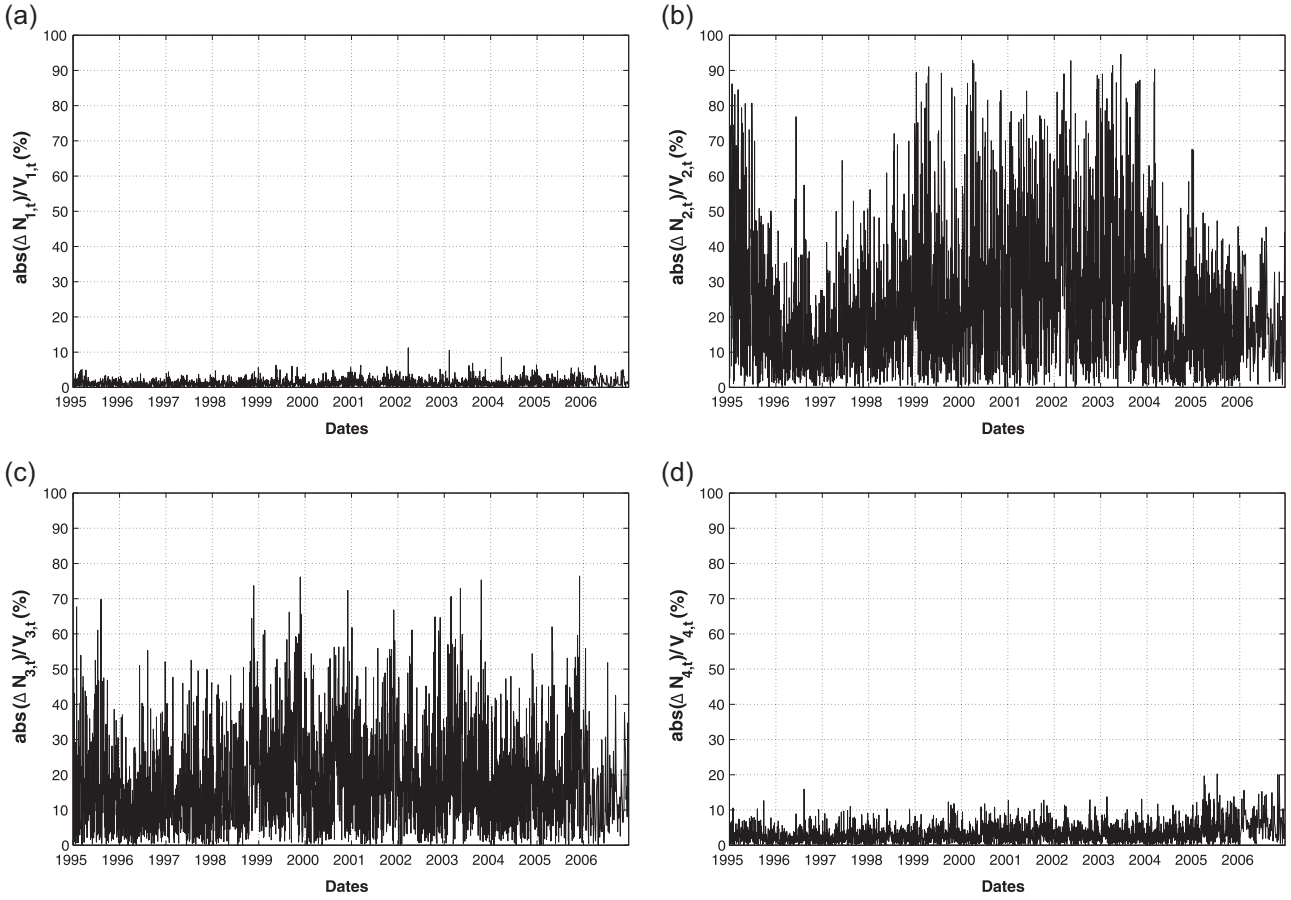

FIGURE 3 Interday trading volume by each CTI group relative to its total volume. This figure shows the interday volume of each CTI group ($|\Delta N_{i,t}|$) normalized by its total volume ($V_{i,t}$), in percentage terms. (a), (b), (c), and (d) are for groups CTI1, CTI2, CTI3, and CTI4, respectively. CTI, customer type indicator

The two figures above give the following picture: Groups CTI1 and CTI4 produce most of the trading volume. But most of their trading is intraday. Groups CTI2 and CTI3, however, do not contribute much to the intraday volume. Nonetheless, their trading is mostly interday, which constitutes most of the interday volume. As we will see later, this distinction in different CTI groups' trading behavior is closely related to their trading motive, the role they play in liquidity provision and consumption and in the price formation process.

Next, we examine in more detail different CTI groups' interday trading, that is, overnight position-taking. We first note that theoretically, all groups' net interday trading has to sum up to zero. This implies that the net daily changes in the CTI groups' interday positions are not dependent. In Table 4, we report the correlations between net changes in the

TABLE 4 Correlations among changes in overnight positions by different CTI groups

	CTI1	CTI2	CTI3	CTI4
CTI1	1.000 (.000)	−0.026 (.174)	0.031 (.104)	−0.526 (.000)
CTI2		1.000 (.000)	−0.034 (.069)	−0.491 (.000)
CTI3			1.000 (.000)	−0.683 (.000)
CTI4				1.000 (.000)

Note: The correlations among changes in overnight positions for the CTI groups ($\Delta N_{i,t}$, $i = 1, 2, 3, 4$). The sample period is 1995–2006. The p values are provided inside the parenthesis.

Abbreviation: CTI, customer type indicator.

TABLE 5 Correlations between changes in interday contract positions for each CTI group

	$\Delta N_{1,t}$				$\Delta N_{2,t}$			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
M1	1.000 (.000)	−0.545 (.000)	−0.418 (.000)	−0.217 (.000)	1.000 (.000)	−0.085 (.000)	−0.143 (.000)	0.076 (.000)
M2		1.000 (.000)	−0.129 (.000)	−0.168 (.000)		1.000 (.000)	0.081 (.000)	−0.062 (.001)
M3			1.000 (.000)	0.008 (.657)			1.000 (.000)	−0.007 (.714)
M4				1.000 (.000)				1.000 (.000)
	$\Delta N_{3,t}$				$\Delta N_{4,t}$			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
M1	1.000 (.000)	0.100 (.000)	0.079 (.000)	0.058 (.002)	1.000 (.000)	−0.357 (.000)	−0.277 (.000)	−0.159 (.000)
M2		1.000 (.000)	0.117 (.000)	0.060 (.001)		1.000 (.000)	−0.109 (.000)	−0.112 (.000)
M3			1.000 (.000)	0.109 (.000)			1.000 (.000)	−0.007 (.720)
M4				1.000 (.000)				1.000 (.000)

Note: The correlations between changes in interday positions ($\Delta N_{i,t}$) in different contracts (M1, M2, M3, and M4) for each CTI group ($i = 1, 2, 3, 4$). The sample period is 1995–2006. The p values are provided inside the parenthesis.

total positions of each CTI groups. It is interesting to note that the daily position changes of group CTI4 is significantly negatively correlated with that of groups CIT1, CTI2, and CTI3. The other correlations are all relatively small. This suggests that to a large extent, for interday trades, group CTI4 is usually on one side and the other three groups are on the other side.

Several previous studies, such as Manaster and Mann (1996), Daigler and Wiley (1999), and Ferguson and Mann (1999), find the same result that trading between locals (CTI1) and customers (CTI4) accounts for a large part of the total volume.[29] Based on a longer sample, we reconfirm this result and also extend the analysis to provide a more complete picture in the distribution of volume across different CTI groups for both intraday trading and interday trading. As we will show later, these trading are driven by different trading needs of these groups and impact the market in different dimensions.

In the discussions above, we only consider the total overnight position of each CTI group in the four most liquid contracts. As one might expect, for various reasons such as hedging or market making, each group (or investor) may hold offsetting positions across contracts. In Table 5, we report the correlations between the changes of each CTI group's overnight positions in contracts with different maturities. Indeed, for groups CTI1, CTI2, and CTI4, changes in interday positions in different contracts often exhibit a mixture of positive and negative signs. For example, changes in their interday positions in the nearest contract are often negatively correlated with changes in interday positions in other contracts (with the exception of CTI2 and contract 4). However, it is interesting to note that for group CTI3 all correlations are positive, implying that for this group changes in positions behave in the same way for all maturities. Contrary to other CTI groups, which seem to hedge across different maturities, CTI3 group "hedges or bets" the market as a whole (i.e., taking "directional trade" on the whole market). As we will see in Section 5, this behavior of group CTI3 is no accident.

[29]Based on the transaction data for all futures trades on the CME during 6 month of 1992, Manaster and Mann (1999) conclude the following breakdown in terms of share of total volume from trading between: locals (CTI1) and customers (CTI4) 41%, customers and other customers 11%, commercials (CTI2) and customers 9%, floor hedgers (CTI3) and customers 4%, locals and locals 9%, locals and commercials 15%, and locals and floor hedgers 6%. For the five futures markets during the 2 years (June 1986 through June 1998) considered by Daigler and Wiley (1999), they find that groups CTI1 and CTI4 have the highest cross-correlation in trading volume among the four CTI groups. Ferguson and Mann (1999) also find that customer trading (CTI4) against any other CTI groups takes between 51% and 75% of the total volume, with the minimum (50%) in Eurodollar and the maximum (75%) in lumber futures.

TABLE 6 Descriptive statistics for changes in positions of each CTI group

	Mean	Median	*SD*	Skew	Kurt	ρ_1	ρ_2	ρ_5	ρ_{10}
$\Delta N_{1,t}$	**−126.998** [−7.021]	**−106.000** [−8.800]	947.973	−0.708	22.777	−0.105 (.000)	−0.003 (.866)	0.011 (.555)	0.006 (.735)
$\Delta N_{2,t}$	**56.821** [2.462]	**7.000** [3.086]	964.583	0.610	44.136	0.080 (.000)	0.118 (.000)	0.025 (.181)	0.027 (.146)
$\Delta N_{3,t}$	**276.417** [7.866]	**257.000** [9.858]	1251.306	−0.088	6.143	0.179 (.000)	0.167 (.000)	0.093 (.000)	0.038 (.045)
$\Delta N_{4,t}$	**−206.239** [−4.635]	**−249.000** [−5.809]	1827.064	0.375	11.608	0.081 (.000)	0.107 (.000)	0.043 (.021)	0.009 (.643)

Note: The descriptive statistics for the changes in positions ($\Delta N_{i,t}$, $i = 1, 2, 3, 4$) of each CTI group. ρ_i denotes the autocorrelation for lag i. The sample period is 1995–2006. The p values are provided inside the parenthesis, and the t statistics are provided in square brackets. Boldface means significative at the 5% level.

Table 6 presents the descriptive statistics for the changes in positions of each CTI group in all contracts. As in all tables in this paper, the t statistics, corrected for heteroscedasticity and autocorrelation using the Newey–West procedure, are provided in square-parenthesis and numbers in bold face indicate significance at the 5% level. Summing the average position changes across groups the total is zero, meaning we are not losing any trades across groups. Yet, different groups on average hold different positions by the end of each day. In particular, group CTI2 and CTI3 finish each day with long positions on average, while group CTI1 and CTI4 finish each day with short positions. It is worth noting that changes in the overnight positions of CTI3 group changes are the most substantial and exhibit significant term persistence. Changes in the overnight positions of CTI4 exhibit similar behavior but with smaller magnitudes.

3.2 | Profit and loss

We now consider the profitability of different CTI groups in their trading. As discussed in Section 2.2, we decompose the profit of each CTI group into two components: intraday profits defined in Equation (13), which arises from their intraday trading, and interday profits defined in Equation (14), which arises from their overnight positions.

In Figure 4, we plot the daily intraday P&L for each CTI group, normalized by the group's own daily volume, and the cumulative normalized intraday P&L over the sample period. Since groups CTI1 and CTI4 do most of the intraday trading, as shown by panels (a) and (d) of Figure 2, their bear a substantial amount of intraday P&L. Furthermore, from panels (b) and (h) of Figure 4, we clearly see that group CTI1 consistently makes money on intraday trading while group CTI4 consistently loses money. This demonstrates that group CTI1 as market makers are compensated from their liquidity provision. This result is consistent with Manaster and Mann (1996), who observes that market makers are not only "passive order fillers," they also manage their inventory as "active profit-seeking individuals." They make profits from both execution and timing, taking advantage of the information they have over the next price movements.

In contrast, group CTI4 is paying for the liquidity in their intraday trading. Group CTI2 does only a small amount of intraday trading, as Figure 2b shows. Thus, it incurs little intraday P&L, as shown in Figure 4d. Group CTI3, on the other hand, does some intraday trading, for which it pays a nontrivial amount. As Figure 4f shows, over time, group CTI3 consistently loses money in their intraday trading.

Table 7 reports the basic summary statistics for the normalized intraday P&L of each CTI group. Clearly, group CTI1 has a positive and significant average daily P&L from intraday trading, while groups CTI3 and CTI4 have significantly negative average P&L from intraday trading. Since the P&L numbers are normalized by each group's own daily volume (by the number of contracts), multiplied by 100, they give the average P&L per contract traded, in cents. Thus, for group CTI1, it earns an average profit of 0.034 cents per contract traded, while for groups CTI3 and CTI4, they on average incur a loss of 0.028 and 0.041 cents, respectively. Group CTI2, however, incurs no profit or loss for its intraday trading. Another striking feature of the daily P&L from intraday trading is that it is highly persistent over time. For example, for group CTI1, its daily intraday P&L has a serial correlation of 0.150 even after 10 days.

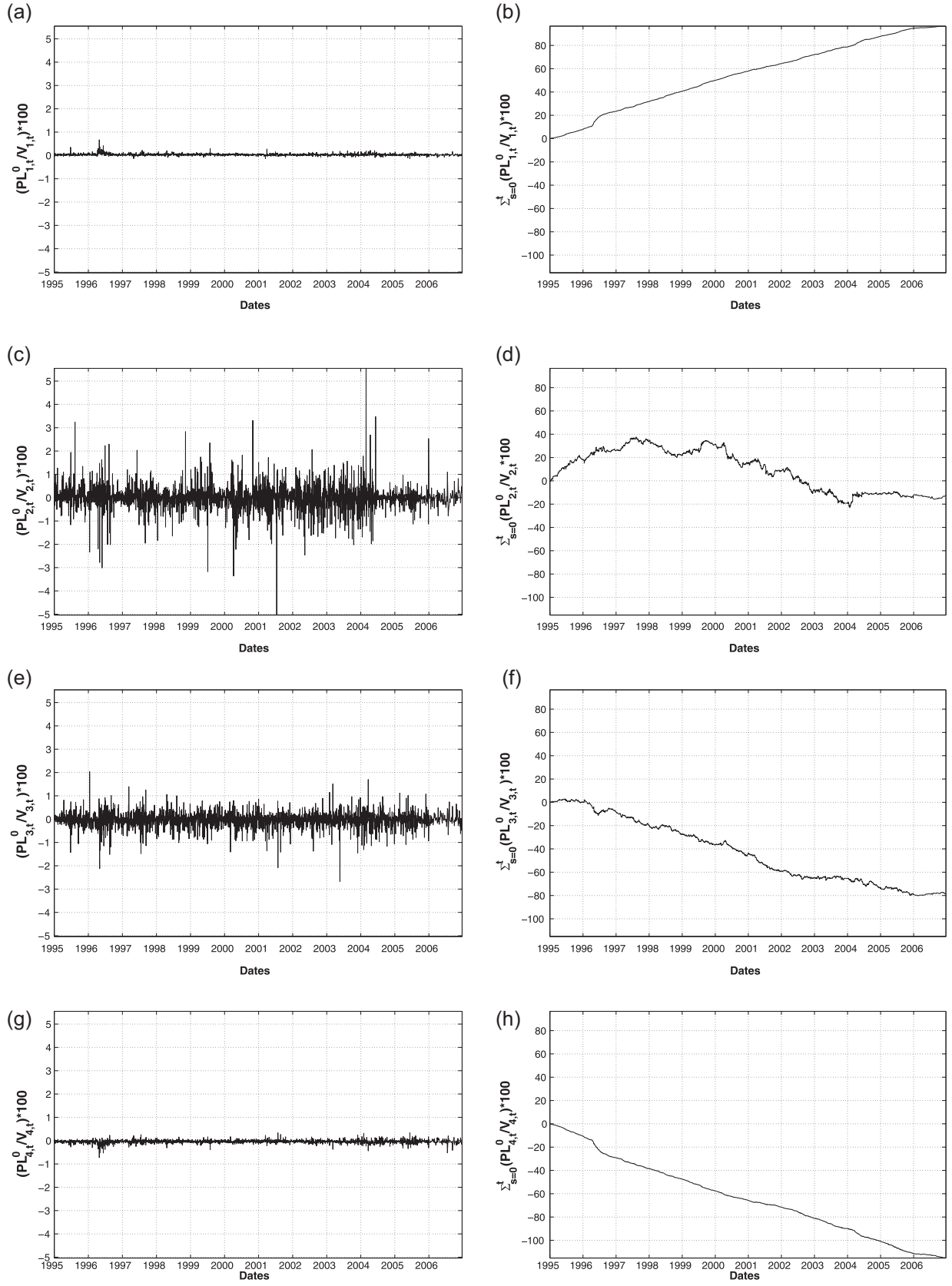

FIGURE 4 Intraday P&L of each CTI group over its daily volume ($\mathrm{PNL}^0_{i,t}/V_{i,t}$) and their corresponding cumulative ($\sum_{s=0}^{t}\mathrm{PNL}^0_{i,s}/V_{i,s}$), multiplied by 100. The left panels (a), (c), (e), and (g) are daily intraday P&L for groups CTI1, CTI2, CTI3, and CTI4, respectively, and the right panels are the corresponding cumulative intraday P&L. P&L, profit and loss

TABLE 7 Descriptive statistics for the intraday P&L of each CTI group normalized by its daily trading volume

	Mean	*SD*	Skew	Kurt	SR	ρ_1	ρ_2	ρ_5	ρ_{10}
CTI1(%)	**0.034** [24.495]	0.043	2.681	28.356	0.791	0.237 (.000)	0.166 (.000)	0.140 (.000)	0.150 (.000)
CTI2(%)	−0.004 [−0.457]	0.514	0.096	17.420	−0.008	−0.019 (.315)	−0.014 (.451)	−0.002 (.913)	−0.012 (.525)
CTI3(%)	**−0.028** [−4.883]	0.315	−0.686	9.635	−0.088	−0.036 (.058)	−0.024 (.201)	−0.009 (.628)	0.006 (.762)
CTI4(%)	**−0.041** [−23.190]	0.070	−0.952	12.311	−0.585	0.067 (.000)	0.080 (.000)	0.071 (.000)	0.064 (.001)

Note: The descriptive statistics for the intraday P&L of each CTI group, normalized by its daily trading volume ($PNL^0_{i,t}/V_{i,t}$, $i = 1, 2, 3, 4$) are shown. SR denotes the Sharpe ratio for each CTI group's P&L, and ρ_i the autocorrelation for lag i. The sample period is 1995–2006. The p values are provided inside the parenthesis, and the t statistics are provided in square brackets. Boldface means significative at the 5% level.

In Table 7, we also report the Sharpe ratio for the intraday P&L for each CTI group, which is defined as the ratio between the mean of their intraday P&L and its standard deviation. For group CTI1, the Sharpe ratio of its intraday P&L is 0.791, which is both statistically and economically significant. As a comparison, the Sharpe ratio on S&P 500 stock index is around 0.4.[30] For CTI2 and CTI3, the Sharpe ratio on their intraday P&L is −0.008 and −0.088, respectively. For CTI4, the Sharpe ratio is −0.585, which is substantial.

Several previous studies have found that group CTI1 typically makes money while group CTI4 loses money, for example, Fishman and Longstaff (1992) and Manaster and Mann (1999). Using a longer sample, we reconfirm these results.[31] Moreover, we show that these profits and loses are mainly from intraday trading, not from interday trading
(at least for corn futures). We further show that group CTI2 roughly break even on their intraday trading while group CTI3 loses money. This result is different from that in Manaster and Mann (1999), which find group CTI2 also loses money. Given that we have much longer sample, while they only have 6 months, our results should be more robust.

Next, we examine the P&L from different CTI groups' interday trading, that is, overnight position-taking. In Table 8, we report the basic summary statistics for their interday P&L. Clearly, group CTI2 on average earns positive profits on its interday trades while group CIT3 bears a loss on its interday trading. The other two groups, CTI1 and CTI4, have no significant profit or loss. The other interesting observation is that unlike the P&L from intraday trading, which exhibits persistence after many lags, P&L from interday trading exhibits some persistence only over the first a few lags.[32]

Dewally et al. (2013) find that the market makers as a group have significant trading loses on their overnight positions. Fishe and Smith (2012) find that market makers are "over-represented" among traders who have positive performance from their overnight positions.

Our results on the performance of different CTI groups' overnight positions are less informative, in part because of the slippage from missing spread trades and ex-pit trades.

3.3 | Intraday P&L and market variables

In this section, we study how the intraday profits for each CTI group ($PNL^0_{i,t}$) is related to the market variables, including turnover and normalized high minus low prices. Given the persistence in P&L, we consider changes in P&L,

[30]Gorton and Rouwenhorst (2006) compare the Sharpe ratio for the Ibbotson corporate bond total return index for the U.S. bonds, the S&P 500 total return index for the U.S. stocks, and a equally weighted commodity futures index, for the period July 1959 through December 2004. They find a Sharpe ratio of 0.43 for commodities futures, 0.38 for stocks, and 0.26 for bonds. Hong and Yogo (2012) find similar numbers for a portfolio of commodities, and for the 10-year U.S. Treasury bond, for the CRSP value-weighted stock portfolio they find a Sharpe ratio of 0.36, monthly data from January 1965, through December 2008.

[31]See, also, Dewally et al. (2013), Fishe and Smith (2012), Hartzmark (1987), Leuthold et al. (1994), and among others. Since their data are based on traders' "end of day positions" from CFTC's Commitments of Trades (COT) data set, their calculations of profits and loses of intraday trades are not reliable because of the lack of precise information on actual transactions prices.

[32]Remember that our LDB data misses about 30% of the EOD trading volume from non-pit transactions, off-hours trades, and potential recording errors.

TABLE 8 Descriptive statistics for the interday P&L of each CTI group, normalized by its daily trading volume

	Mean	*SD*	Skew	Kurt	SR	ρ_1	ρ_2	ρ_5	ρ_{10}
CTI1 (%)	0.045 [1.100]	2.358	−0.259	5.460	0.019	0.015 (.434)	−0.085 (.000)	−0.025 (.195)	0.021 (.265)
CTI2 (%)	**0.609** [2.422]	13.491	−0.213	8.254	0.045	0.025 (.182)	−0.062 (.001)	−0.018 (.341)	0.007 (.696)
CTI3 (%)	**−0.765** [−2.267]	19.590	0.002	4.042	−0.039	−0.001 (.960)	−0.046 (.016)	−0.027 (.157)	−0.001 (.974)
CTI4 (%)	−0.003 [−0.072]	2.715	−0.008	6.866	−0.001	0.038 (.042)	−0.032 (.092)	−0.080 (.000)	0.001 (.946)

Note: The descriptive statistics for the interday P&L of each CTI group, normalized by its daily trading volume ($\mathrm{PNL}^1_{i,t}/V_{i,t}$, $i = 1, 2, 3, 4$) are shown. SR denotes the Sharpe ratio for each CTI group's P&L, and ρ_i the autocorrelation for lag i. The sample period is 1995–2006. The p values are provided inside the parenthesis, and the t statistics are provided in square brackets. Bold face means significative at the 5% level.

which still exhibits nontrivial autocorrelation within two lags. We thus decompose the changes in P&L into a predictable part using an MA(2) model and a residual as follows:

$$\Delta\mathrm{PNL}^0_{i,t} = \widehat{\Delta\mathrm{PNL}}^{\,0}_{i,t} + \hat{u}^0_{i,\Delta\mathrm{PNL},t}, \tag{18}$$

where $\widehat{\Delta\mathrm{PNL}}^{\,0}_{i,t}$ is the forecast of $\Delta\mathrm{PNL}^0_{i,t}$ from the estimated MA(2) model.[33] We then consider how $\hat{u}^0_{i,\Delta\mathrm{PNL},t}$ can be explained by changes in τ_t and X_t, that is $\Delta\tau_t$ and ΔX_t, as well as the measure imbalance Z_t. As control variables, we also include lagged changes in log settlement prices. All the market variables are averaged over the four most liquid contracts.

Following Bessembinder and Seguin (1993), we run the regression by decomposing the highly persistent explanatory variables (changes in volume and volatility) by its forecasted value based on the MA(2) model and the residual, if the variable shows up significant (see the appendix for more details). Formally, the relationship we estimate is given as follows:

$$\begin{aligned}\hat{u}^0_{i,\Delta\mathrm{PNL},t} = a &+ b_1\widehat{\Delta\tau_t} + b_1'\hat{u}_{\Delta\tau,t} + b_2\widehat{\Delta X_t} + b_2'\hat{u}_{\Delta X,t} + b_3 Z_t \\ &+ \text{ lagged changes in log settlement prices as controls}.\end{aligned} \tag{19}$$

Since only groups CTI1 and CTI4 incur significant persistent intraday P&L, we report the regression results only for them. Table 9 gives the results for groups CTI1 and CTI4.

From the left panel of Table 9, we see that lagged changes in log prices show certain statistical significance in explaining changes in intraday P&L for group CTI1. However, the power is quite weak, with a R^2 of less than 1.5%. Trade imbalance measure Z_t has no explanatory power. Both innovations in turnover and price volatility, as measured by normalized high minus low prices, have significant power in explaining changes in group CTI1's intraday P&L. The former yields a R^2 of around 15% alone while the latter gives a R^2 of 13% by itself. Together, they explain over 17% of the daily changes in CTI1 group's intraday P&L. These results are not surprising. As market makers, one would expect CTI1 group to make more profits when volume or volatility rises (they demand higher spread when volatility is high).[34]

It is worth noting that changes in τ_t and X_t are highly correlated. We have also examined if each one contributes additional explanatory power by looking at their orthogonal parts and the answer is positive. For brevity, we have omitted these results here.

The right panel of Table 9 reports the results for CTI4, which are similar to those for CTI1 except that the signs are reversed. This is expected as group CTI4 consumes the liquidity provided by group CTI1. It should be noted that the

[33]Given the standard nature of this procedure, we omit the details here, which are given in Appendix A.3.

[34]Dewally et al. (2013) find a positive relationship between interday profits for the group of speculators, which includes market makers and floor traders, and its trading volume. Chang (1985) analyzes monthly data on net positions of large hedging, speculative, and nonreporting traders and concludes that the positive performance for the large speculators is mainly due to the risk they bear for corn and soybean while for wheat it is also due to forecasting ability for price movements.

TABLE 9 Regressions of intraday P&L of groups CTI1 and CTI4 on market variables

	CTI1				CTI4					
c	0.000 [0.287]	0.000 [0.208]	0.000 [0.242]	0.000 [0.371]	0.000 [0.210]	−0.000 [−0.198]	−0.000 [−0.133]	−0.000 [−0.161]	−0.000 [−0.288]	−0.000 [−0.135]
$\widehat{\Delta\tau_t}$		−0.002 [−0.371]					0.003 [0.485]			
$\hat{u}_{t,\Delta\tau}$		**0.059** [8.359]			**0.042** [4.993]		**−0.051** [−6.237]			**−0.043** [−4.533]
$\widehat{\Delta X_t}$			−0.011 [−0.367]					0.008 [0.224]		
$\hat{u}_{t,\Delta X_t}$			**0.349** [8.356]		**0.163** [3.109]			**−0.273** [−5.340]	−0.085	[−1.411]
Z_t				0.002 [0.562]					−0.002 [−0.620]	
$\Delta \ln P_{t-1}$	**0.041** [3.088]	0.022 [1.922]	**0.029** [2.282]	**0.041** [3.012]	0.022 [1.935]	−0.020 [−1.030]	−0.003 [−0.181]	−0.011 [−0.555]	−0.019 [−0.979]	−0.004 [−0.196]
$\Delta \ln P_{t-2}$	**0.039** [3.118]	**0.033** [2.832]	**0.033** [2.812]	**0.039** [3.065]	**0.032** [2.854]	**−0.036** [−2.016]	−0.030 [−1.691]	−0.032 [−1.757]	**−0.036** [−1.989]	−0.031 [−1.743]
$\Delta \ln P_{t-3}$	**0.024** [2.074]	0.010 [0.951]	0.015 [1.417]	**0.024** [2.070]	0.010 [0.977]	**−0.034** [−2.364]	−0.021 [−1.541]	−0.027 [−1.910]	**−0.034** [−2.361]	−0.022 [−1.569]
R^2	1.362	15.852	13.142	1.343	17.266	0.604	7.641	5.169	0.586	7.874

Note: The results for the regressions of intraday P&L of groups CTI1 and CTI4 on selected market variables (decomposed into expected and unexpected parts) are shown. The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets. Boldface means significative at the 5% level.

main significant variable when put together, that can explain the daily variation in CTI4's intraday P&L is innovations in turnover, which leads to a R^2 around 8%.

4 | PRICE DISCOVERY AND CTI TRADING

Other than public information, futures price changes are driven by the private information contained in the trading of market participants (see, e.g., Grossman, 1976 and Kyle, 1985). To the extent that different market participants may have different information regarding the underlying factors, their trading would have a different impact on the evolution of futures prices. Given the information on the trading of different CTI groups, we can explore the role each CTI group plays in the price discovery process.

To do so, we examine in this section the relationships between the change in the log of the settlement price ($\Delta \ln P_t$) and the contemporaneous change in the interday (i.e., overnight) positions of different CTI groups ($\Delta N_{i,t}$). First, we select the relevant market variables that will be included as control variables in explaining price changes. Following the procedure in previous section, we will also consider forecasts and residuals of volume and volatility as separate explanatory variables. Next, we consider the contemporaneous changes in the positions of different CTI groups. In addition, we also include the lagged dependent variables as controls.

The most general specification of the regression is as follows:

$$\begin{aligned}\Delta \ln P_t = a &+ b_1 \Delta N_{1,t} + b_2 \Delta N_{2,t} + b_3 \Delta N_{3,t} + b_4 \Delta N_{4,t} \\ &+ \text{ contemporaneous changes in market variables as controls} \\ &+ \text{ lagged changes in the log settlement price as controls},\end{aligned} \tag{20}$$

where the market variables include changes in turnover $\Delta\tau_t$ and price volatility ΔX_t as well as the imbalance measure Z_t. For the dependent variable, we can consider different prices given the presence of multiple contracts. One choice is the price of the nearest contract, $P_t^{(1)}$. The reason for this choice is that the nearest contract is the most liquid and its price is

TABLE 10 Regression of the equally weighted price change on contemporaneous changes in market variables

c	−0.000 [−1.516]	−0.000 [1.654]	−0.000 [−1.662]	0.000 [1.647]	0.000 [1.619]
$\widehat{\Delta\tau_t}$		**−0.020** [2.708]			**−0.015** [−2.308]
$\hat{u}_{t,\Delta\tau_t}$		0.012 [1.000]			
$\widehat{\Delta X_t}$			**−0.094** [−1.965]		
$\hat{u}_{t,\Delta X_t}$			**0.178** [2.538]		0.119 [1.926]
Z_t				**0.100** [18.025]	**0.098** [17.575]
$\Delta \ln P_{t-1}$	**0.053** [2.644]	**0.047** [2.266]	**0.044** [2.078]	0.012 [0.608]	0.006 [0.328]
$\Delta \ln P_{t-2}$	0.018 [0.829]	0.012 [0.558]	0.013 [0.596]	−0.012 [−0.634]	−0.017 [−0.907]
$\Delta \ln P_{t-3}$	−0.031 [−1.435]	−0.038 [−1.758]	−0.039 [−1.798]	−0.034 [−1.808]	−0.039 [2.112]
$\Delta \ln P_{t-5}$	**−0.051** [−2.131]	**−0.052** [−2.189]	**−0.049** [−2.080]	**−0.048** [−2.338]	**−0.049** [−2.438]
R^2	0.539	0.883	1.485	18.316	18.794

Note: The results for the regression of the equally weighted price change ($\Delta \ln P_t$) on contemporaneous changes in market variables (decomposed into expected and unexpected parts). The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets. Bold face means significative at the 5% level.

probably the most efficient in reflecting the information in the market. Another is the average price of the first four nearest contracts, P_t. The advantage of using this price is that through averaging it reduces the potential noise in the price of a particular contract[35]. To be consistent with the rest of the paper, we use the average price over the four nearest contacts. The results are similar when we use the price of the nearest contact, which are given in Appendix A.4 as a robustness check.

Table 10 describes how market variables, as well as lagged price changes, explain the contemporaneous price changes. It is clear that lagged price changes yield significant coefficients, especially price changes with 1-day lag and 5-day lag. However, they yield very low R^2's, overall less than 1%. Expected changes in turnover, $\widehat{\Delta\tau_t}$, and unexpected change in high minus low prices, $\hat{u}_{\Delta X,t}$, also have significant coefficients when appear alone in regression (20). But the marginal contribution to R^2 is quite small, less than 1%. The imbalance measure, Z_t, however, leads to significant explanatory power, increasing the R^2 by more than 15%. In fact, it dominates all the other variables by far in terms of explaining the daily price changes. Given these results, in our future analysis, we will keep these variables, especially Z_t, in our analysis, in particular as controls.

In Table 11, we report the results with the changes in different CTI groups' interday positions. Controlling for the market variables and lagged price changes, changes in the interday position of the group CTI2 and CTI4 exhibit significant explanatory power (at 5% level) over the contemporaneous price changes. In particular, changes in the interday position of group CTI2 have a negative coefficient, while changes in the interday positions of group CTI4 exhibit a positive coefficient. If both are included in the regression, only changes in the interday positions of group CTI4 remains significant at the 5% level. It increases the (adjusted) R^2 by about 1.5%.[36] This result is consitent with those presented in Figure 2d and Table 7, which show that CTI4 trades heavily intraday, and consumes liquidity, loses money. Market prices move to reflect fundamental external information they bring to the market.

[35]Of course, other than the simple average, one could consider other forms of average price such as averages weighted by open interest or trading volume. Our results are quite robust with respect to these different choices of averages.

[36]If one is willing to assume that the missing trades in our sample do not change the results, then we interpret the signs in CTI groups' interday trades, that is, CTI2 group are overall contrarians/hedgers and CTI4 momentum traders. Considering both together, the momentum trading prevails.

TABLE 11 Regression of the equally weighted price change on contemporaneous changes in overnight position of each CTI group and market variables

c	0.000 [1.619]	0.000 [1.123]	**0.000** [1.958]	**0.000** [2.059]	**0.000** [2.394]	**0.000** [2.368]
$\Delta N_{1,t}$		−0.340 [−1.769]				
$\Delta N_{2,t}$			**−0.732** [−3.381]			−0.477 [−1.914]
$\Delta N_{3,t}$				−0.219 [−1.674]		
$\Delta N_{4,t}$					**0.399** [4.442]	**0.274** [2.530]
$\widehat{\Delta\tau_t}$	**−0.015** [−2.308]	**−0.015** [−2.270]	**−0.014** [−2.148]	**−0.015** [−2.203]	**−0.013** [−1.991]	**−0.013** [−1.984]
$\hat{u}_{t,\Delta X_t}$	0.119 [1.926]	0.117 [1.893]	**0.127** [2.083]	0.115 [1.846]	0.112 [1.830]	**0.120** [1.959]
Z_t	**0.098** [17.575]	**0.098** [17.502]	**0.097** [17.097]	**0.099** [17.677]	**0.098** [17.128]	**0.097** [17.062]
$\Delta \ln P_{t-1}$	0.006 [0.328]	−0.000 [−0.011]	0.003 [0.156]	0.008 [0.440]	0.001 [0.033]	0.000 [0.015]
$\Delta \ln P_{t-2}$	−0.017 [−0.907]	−0.016 [−0.885]	−0.019 [−1.052]	0.018 [−0.984]	−0.021 [−1.103]	−0.021 [−1.135]
$\Delta \ln P_{t-3}$	**−0.039** [−2.112]	**−0.040** [−2.131]	**−0.039** [−2.129]	**−0.043** [−2.326]	**−0.048** [−2.527]	**−0.045** [−2.434]
$\Delta \ln P_{t-5}$	**−0.049** [−2.438]	**−0.050** [−2.465]	**−0.050** [−2.469]	**−0.051** [−2.481]	**−0.053** [−2.574]	**−0.052** [−2.557]
R^2	18.794	19.023	20.017	18.951	20.088	20.462

Note: The regression of the equally weighted price change ($\Delta \ln P_t$) on contemporaneous changes in overnight position of each CTI group ($\Delta N_{i,t}$) and market variables (decomposed into expected and unexpected parts) are shown. The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets. Boldface means significative at the 5% level.

For Treasury futures, Brandt et al. (2007) find that net order flow (changes in overnight positions) for CTI4 and CTI2 plus CTI3 are significantly correlated with contemporaneous price changes. They only consider order flows in the nearest contract and do not control for market variables and lagged prices, although they do use detrended variables and control for macroeconomic announcements.[37] Manaster and Mann (1996) report a similar relationship between change in prices and group CTI4's net order flow for CME traded futures.[38] Our results are in general consistent with these findings but further, sharpen them. In particular, we control for other market variables that contribute significantly to daily price changes. This allows us to better isolate the marginal contribution of each CTI group's daily net order flow to price formation. By looking at groups CTI2 and CTI3 separately, we can better identify their own impact.

5 | CTI TRADING AND PRICE FORECASTING

The previous section studies the contemporaneous relationship between changes in different CTI groups' overnight positions and price changes. In this section, we explore what useful information, if any, changes in the overnight positions of different CTI groups may contain about future price movements. We consider

[37]See also Menkveld et al. (2012). Using similar methodology and data, they also find positive correlation between net customer order flow (net trades of group CTI4) and price changes of the nearest contract during the beginning of each trading day.

[38]Using 250 days of transaction data in 2005 for the E-mini S&P 500, Nasdaq-100 and Russell 2000 index futures (all electronically traded), Kurov (2008) examines the price formation process within a trading day. He finds that trades initiated by exchange members (CTI1 and CTI3) and CTI2 contribute more to price discovery than do off-exchange (CTI4) initiated trades. Clearly, the contributions from different trader groups to the price formation process at high frequency (intraday) are different from those at the low frequency. Also see Daigler and Wiley (1999), Kurov and Lasser (2004), and Locke and Onayev (2007).

this question from two different perspectives. We first use simple linear models to examine if overnight positions of each CTI group can forecast the price changes the next day. Next, we use nonparametric methods to analyze if their overnight positions reveal more information about the overall distribution of next-day price changes.

5.1 | Price forecasting with CTI overnight positions

We follow the similar procedure as in the previous section, with now the next-day price change as the dependent variable. The general specification takes the following form:

$$\begin{aligned}\Delta \ln P_{t+1} = a &+ b_1 \Delta N_{1,t} + b_2 \Delta N_{2,t} + b_3 \Delta N_{3,t} + b_4 \Delta N_{4,t} \\ &+ \text{ lagged changes in market variables as controls} \\ &+ \text{ lagged changes in the log settlement price as controls}.\end{aligned} \tag{21}$$

As before, the market variables include the changes in turnover and volatility with both the predicted part and the unpredicted part from an MA(2) model. Market imbalance and lagged price changes are also included as additional controls. We follow the same procedure as before: First, we choose the relevant control variables and then include the changes in positions of the CTIs. We only include the market variables that are significant in univariate regressions. For brevity, we do not detail these results here but only give a brief summary. Lagged price changes, in particular with lag 1 and 5, show up significantly as predictors future price changes. However, the R^2 is quite small, less than 1%. This, of course, should not be surprising given the high informational efficiency of the corn futures market. Two components of the market variables are significant in forecasting future price changes, unexpected changes in market turnover $\hat{u}_{\Delta\tau,t}$ and imbalance measure Z_t. Their marginal contribution to the R^2 is rather small, being less than 0.5% altogether.

Table 12 describes the results from the regression in Equation (21). Obviously, only the overnight positions of group CTI3 exhibit additional significant predictive power, over the controls. The marginal contribution of R^2, which is around 0.25%, remains small for reasons mentioned above. It is interesting to note that this result is consistent with the results presented in Table 5, which shows that CTI3 is the only group whose overnight positions are positively correlated across contracts with different maturities. It is also worth noting that the predictability disappears beyond 1 day.

The result that changes in CTI3's overnight position can predict next day price changes contrasts interestingly with those from previous work. For Treasury futures, Brandt et al. (2007) find no significant predicative power for net order flows of different CTI groups at the daily interval. This may simply reflect the fact that for financial futures, which are linked to market-wide prices, no CTI group possesses superior information over the market. As mentioned before, this is also the main motivation for us to look at corn futures in this paper.[39] Fishe and Smith (2012) do find floor brokers/traders are over-represented among the traders who have positive performance on overnight positions. Their results are consistent with our findings, with two important distinctions. First, market variables can predict future price changes, as we show here, which implies that expected returns on futures are time-varying. Thus, positive performance can come from two sources, trading on time-varying expected returns and on private information. Without properly adjusting the effect of time-varying expected returns, it is hard to identify the potential effect of private information. Second, floor brokers and floor traders may have different trading motives and private information. Pooling them together will prevent us from further identify these differences. By controlling for the impact of market variables that can predict future price changes and examining CTI2 and CTI3 groups separately, we are able to better identify the potential information advantage of group CTI3.

Locke and Onayev (2007) look at 5-min intervals and find a positive relationship between price changes and only customer order flows (group CTI4) beyond several lags. Using weekly trader positions in currency futures, Wang (2004) finds that positions of speculators (noncommercial traders) have a positive influence on future returns while positions of hedgers (commercial traders) have a negative influence.[40] It should be clear that the potential information advantage

[39]Dusak (1973), Bessembinder (1992), and de Roon et al. (2000) find different relationships between agricultural futures and systematic market risk variables than those for financial futures.

[40]Using weekly trader positions from the CFTC (commercial and noncommercial) for five foreign currency futures markets, January 1993 through March 2010, finds that speculators (noncommercial) positive influence future returns and hedgers (commercial) negatively influence future returns. See also Bessembinder (1992) and de Roon et al. (2000).

TABLE 12 Regression of the equally weighted price change on lagged changes in overnight position of each CTI group and lagged change in market variables

c	−0.000 [−1.211]	−0.000 [−1.244]	−0.000 [−1.194]	−0.000 [−1.791]	−0.000 [−1.386]
$\Delta N_{1,t}$		−0.061 [−0.522]			
$\Delta N_{2,t}$			−0.027 [−0.279]		
$\Delta N_{3,t}$				**0.297** [2.731]	
$\Delta N_{4,t}$					−0.117 [−1.734]
$\hat{u}_{t,\Delta\tau_t}$	**0.014** [2.604]	**0.014** [2.599]	**0.014** [2.610]	**0.015** [2.692]	**0.014** [2.586]
Z_t	**0.013** [2.542]	**0.013** [2.545]	**0.013** [2.541]	**0.012** [2.390]	**0.013** [2.479]
$\Delta \ln P_t$	0.028 [1.317]	0.028 [1.288]	0.028 [1.271]	0.031 [1.483]	0.033 [1.517]
$\Delta \ln P_{t-1}$	0.010 [0.440]	0.008 [0.387]	0.010 [0.435]	0.007 [0.337]	0.012 [0.525]
$\Delta \ln P_{t-2}$	−0.039 [−1.772]	−0.039 [−1.770]	−0.039 [−1.779]	−0.036 [−1.649]	−0.037 [−1.700]
$\Delta \ln P_{t-4}$	**−0.050** [−2.101]	**−0.050** [−2.106]	**−0.050** [−2.104]	−0.045 [−1.895]	**−0.048** [−2.001]
R^2	0.953	0.926	0.920	1.263	1.031

Note: The regression of the equally weighted price change ($\Delta \ln P_{t+1}$) on lagged changes in overnight position ($\Delta N_{i,t}$) and lagged change in market variables (decomposed into expected and unexpected parts) are shown. The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets. Boldface means significative at the 5% level.

over very short or long horizons documented in these studies is different from what we identified at daily frequencies (but not beyond).

5.2 | Non parametric analysis

The regression analysis above focuses on how overnight positions of each CTI group may forecast next-day price changes, on average. It should be recognized that their positions may contain other useful information about the next-day price beyond what is capture in its mean change. It may convey information about other aspects of the next-day price, such as the distribution of its change. In this section, we use nonparametric statistics to examine the extent of this information.

First, for each CTI group, we sort the sample into two cases: days when $\Delta N_{i,t} > 0$ and days when $\Delta N_{i,t} < 0$ (we ignore the days with $\Delta N_{i,t} = 0$). We then consider the price change in the following day. If changes in the CTI group's overnight positions are not informative about future price changes, we should see the same distribution of next-day price changes between these two cases. In other words, the distributions of future price changes conditional on the CTI group's buying and selling of the market, respectively, should be the same when these trades are uninformative. Otherwise, the distributions will be different.

Next, we compare the moments of these two conditional distributions. The first four rows of panels A–D in Table 13 report the results for CTI group 1–4. In particular, we report in columns 2 and 3 the the mean, standard deviation, skewness, and kurtosis of the two conditional distributions of $\Delta \ln P_{t+1}$, conditional on buying and selling of a CTI group on date t. The last column reports the p value for the test of the null hypothesis that two conditional moments are the same, that is, $H_0 : \theta_+ = \theta_-$, where θ_+ and θ_- denote the moment conditional on buying and selling respectively. Here, the p value are obtained using bootstrap statistics with 5,000 simulations.

TABLE 13 Nonparametric tests on the distribution of future price changes of the equally weighted price conditional on changes in overnight positions of each CTI group

$\Delta \ln P_{t+1}$			
Panel A	$\Delta N_{1,t} > 0$	$\Delta N_{1,t} < 0$	***p* Value** $(\theta_+ = \theta_-)$
Mean	−0.000	−0.000	0.952
SD	0.013	0.012	
Skewness	0.257	0.178	0.676
Kurtosis	4.992	4.439	0.322
No. of obs.	1,239	1,587	
K–S test	F_+	F_-	0.989
Panel B	$\Delta N_{2,t} > 0$	$\Delta N_{2,t} < 0$	***p* Value** $(\theta_+ = \theta_-)$
Mean	−0.000	−0.001	0.531
SD	0.012	0.013	
Skewness	0.493	−0.044	**0.004**
Kurtosis	5.453	4.025	**0.009**
No. of obs.	1,474	1,343	
K–S test	F_+	F_-	0.526
Panel C	$\Delta N_{3,t} > 0$	$\Delta N_{3,t} < 0$	***p* Value** $(\theta_+ = \theta_-)$
Mean	0.000	−0.001	**0.001**
SD	0.012	0.013	
Skewness	0.435	−0.067	**0.008**
Kurtosis	5.290	3.884	**0.012**
No. of obs.	1,719	1,105	
K–S test	F_+	F_-	**0.036**
Panel D	$\Delta N_{4,t} > 0$	$\Delta N_{4,t} < 0$	***p* Value** $(\theta_+ = \theta_-)$
Mean	−0.001	−0.000	0.075
SD	0.013	0.013	
Skewness	0.022	0.362	0.077
Kurtosis	4.102	5.175	0.056
No. of obs.	1,177	1,646	
K–S test	F_+	F_-	0.421

Note: The nonparametric tests on the distribution of future price changes of the equally weighted price ($\Delta \ln P_{t+1}$) conditional on changes in overnight positions of each CTI group ($\Delta N_{i,t}$, $i = 1, 2, 3, 4$). Panel A is for CTI1 group, Panel B is for CTI2 group, Panel C is for CTI3 group, and Panel D is for CTI4 group. The K-S test means the Kolmogorov–Smirnov test of the equality of the two conditional future price changes distributions. The sample period is 1995–2006. The *p* values are obtained using bootstrap statistics with 5,000 simulations. Bold face means different from zero at the 5% level.

Consistent with the results from the forecast regressions, the mean of the two conditional distributions are statistically different only for group CTI3. What is interesting is that the trading of group CTI3 also contains information about the other moments of future price change. For example, conditional on the buying of group CTI3, the skewness of the price change the next day is 0.435. But conditional on their selling, the skewness of the price change the next day becomes −0.067. The difference is significant at the 1% level. The kurtosis in the two cases are also different, 5.290 for the former and 3.884 for the latter.

It should also be noted that although the trading of group CTI2 is not informative about the mean of the future price change, it is informative about its skewness. For both groups CTI2 and CTI3, buying today predicts a positive skewness for next-day's price change, while selling predicts a negative skewness.

Finally, we use the Kolmogorov–Smirnov test (K–S test) to detect the difference between the two conditional distributions. The results are presented in the last rows of Panels A–D in Table 13 for each CTI group, respectively. We

find that the distributions conditional the buying and selling are different only for group CTI3. We thus conclude that at the CTI group level, only group CTI3's trading contains information about future price changes.

It is widely believed among researchers that the CTI3 group consists of mostly hedgers, who trade in other products for their own accounts and use the futures for hedging reasons. For example, Ferguson and Mann (1999) states that "Trades executed on behalf of other members (CTI3) are most often trades done for hedging purposes."[41] Such a perception has lead most prior work to assume that group CTI3 does not possess superior information, at least beyond very short horizons (e.g., several minutes). It was then often bundled either with group CTI2 as hedgers or with CTI1 as floor brokers/traders (FBT). Even though Fishe and Smith (2012) have find a more likely positive correlation between the overnight positions of FBTs as a group, they interpret this information advantage as coming from better understanding order flows rather coming from private information about more fundamental factors.[42] Our analysis suggests that group CTI3 may have a distinctive information advantage over other groups at daily horizons. This information is unlikely to come from overall order flows, for which group CTI1 not CTI3 is in the best position to observe.[43] Thus, group CTI3 may play a unique role in the futures market, especially in bringing in information in the price formation process. It would be interesting to further explore the potential source of this information.

6 | CONCLUSION

In this paper, we examine the trading behavior of different CTI groups in the corn futures market. Despite the fact that each group consists of a rich set of market participants with potentially different trading motives, we have several interesting findings. First, there are clear differences in the trading behavior of the different CTI groups. In particular, most of the trading volume is attributed to the trading of groups CTI1 and CTI4. In addition, most of their volume is intraday. In terms of interday trading, it is coming mostly from groups CTI2 and CTI3. This is consistent with the broad definition of each CTI group. Group CTI4 is the main liquidity demander, especially on an intraday basis, while group CTI1 is the main liquidity supplier. Groups CIT2 and CTI3 trade mostly for hedging or speculative reasons. The half-life of their positions tend to be longer—they carry nontrivial amount of overnight positions.

Consistent with this picture, we observe that on intraday trading, group CTI4 consistently loses money, which reflects the price they pay for consuming liquidity. On the contrary, group CTI1 consistently makes money on their intraday trading, which reflects the compensation for their liquidity provision. The Sharpe ratio for group CTI1's intraday liquidity provision amounts to 0.791, which is quite substantial. Furthermore, we find that group CTI1 P&L from intraday trading is positively correlated with the unexpected change in turnover and high minus low price, which captures the intraday price volatility. These results are consistent with the interpretation that their P&L compensates for the risk of market-making.

The different trading behavior of different CTI groups further allow us to examine their role in the price discovery process of the market. Here, we find several informative results. First, we noticed that daily changes in several market variables, especially the expected level of turnover and unexpected change in high minus low price during the day, are significantly correlated with contemporaneous changes in price. More important, we find that a measure of trade imbalance, a normalized difference between the volume at the top price interval and the bottom price interval, exhibits significant explanatory power on contemporaneous price changes. Its contribution to R^2 exceed 15%, while the other variables including lagged price changes can only yield a R^2 of no more than 1.5% even when they are statistically significant. Next, we show that changes in the interday positions of only groups CTI2 and CTI4 also have significant explanatory power for the contemporaneous price. The former has a negative coefficient, indicating that they trade against the market, while the latter has a positive coefficient. When we include both changes in the regression, only changes in the interday positions of group CTI4 remains significant, with a positive coefficient. In other words, when group CTI4 collectively buys during a day, the futures price tends to increase for the same day.

[41]See also Manaster and Mann (1996). Both papers further attribute the hedging needs to the delta hedging of CTI3 traders' positions in futures options.

[42]As mentioned above, Fishe and Smith (2012)'s results on positive performance from overnight positions of FBTs may come from risk-taking rather than better information.

[43]In fact, Dewally et al. (2013) actually show that the group composed of market makers and floor traders, based on CFTC's LTRS data, on average lose money on their overnight positions.

We further explore if changes in the overnight positions of different CTI groups can predict future price changes. We find that only the changes in group CTI3's interday positions can predict the change in futures price next day. In addition, using nonparametric statistics, we also find that group CTI3's interday positions contains information about other aspects of the distribution of next-day price changes such as skewness and kurtosis.

Overall, we conclude that in the corn futures market, group CTI1 mainly act as market makers for intraday transactions. Groups CTI2 and CTI4 contribute to the price discovery through their intraday trading. Group CTI3, however, may possess private information about futures prices beyond the intraday horizon.

ACKNOWLEDGMENTS

The authors acknowledge CBOT for providing part of the data. Sergey Iskoz contributed in the early stage of this study. The authors thank Steve Figlewski, Yingzi Zhu, Tingting Liu, Diego Rodriguez, the Editor (Robert Webb), and participants at the China International Conference in Finance 2015, CFTC, 24th Finance Forum, Cunef, UIB, EFA-2018, 4th FMND-Paris, and 3rd Annual JPMCC-CU-Denver meetings/seminars for valuable comments, and Robert Almgren, Raymond P.H. Fishe, Esen Onur, and Donald Jones for helpful exchanges in understanding the structure of the market. Llorente thanks the MIT Sloan School of Management for its hospitality during his stays at MIT during different stages of this project. Llorente acknowledges financial support from the BBVA Foundation, and Grants ECO02012-32554 and ECO2017-85356-P from the Spanish Ministerio de Economia and Competitividad.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Guillermo Llorente http://orcid.org/0000-0003-3570-3067

REFERENCES

Bessembinder, H. (1992). Systematic risk, hedging pressure, and risk premiums in futures markets. *Review of Financial Studies*, *5*(4), 637–667.

Bessembinder, H., & Chan, K. (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*, *32*(2), 169–193.

Bessembinder, H., & Seguin, P. J. (1993). Price volatility, trading volume, and market depth: Evidence from futures markets. *Journal of Financial and Quantitative Analysis*, *28*, 21–39.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, *3*, 167–179.

Boyd, N. E., & Kurov, A. (2012). Trader survival: Evidence from the energy futures markets. *Journal of Futures Markets*, *32*(9), 809–836.

Brandt, M. W., Kavajecz, K. A., & Underwood, S. E. (2007). Price discovery in the treasury futures markets. *The Journal of Futures Markets*, *27*(11), 1021–1051.

Brunetti, C., Buyuksahin, B., & Harris, J. H. (2016). Speculators, prices and market volatility. *Journal of Financial and Quantitative Analysis*, *51*, 1545–1574. http://ssrn.com/abstract=1736737

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

Chang, E. C. (1985). Returns to speculators and the theory of normal backwardation. *Journal of Finance*, *40*(1), 193–208.

Daigler, R. T., & Wiley, M. K. (1999). The impact of trader type on the futures volatility volume relation. *Journa of Finance*, *54*(6), 2297–2316.

de Roon, F., Nijman, T. E., & Veld, C. (2000). Hedging pressure effects in futures markets. *Journal of Finance*, *55*(3), 1437–1456.

Dewally, M., Ederington, L. H., & Fernando, C. S. (2013). Determinants of trader profits in commodity futures markets. *Review of Financial Studies*, *26*(10), 2648–2683.

Dusak, K. (1973). Futures trading and investors returns: An investigation of commodity market risk premiums. *Journal of Political Economy*, *81*(6), 1387–1406.

Ederington, L., & Lee, J. H. (2002). Who trades futures and how: Evidence from the heating oil futures market. *Journal of Business*, *75*(2), 353–373.

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, *62*(2), 69–97.

Evans, M. D. D., & Lyons, R. K. (2002). Order flow and exchange rate dynamics. *Journal of Political Economy*, *110*(1), 170–180.

Ferguson, M., & Mann, S. C. (1999). *The impact of off-exchange customer trades on prices: evidence from the futures markets*. Working Paper.

Fishe, R. P. H., & Smith, A. D. (2012). Identifying informed traders in futures markets. *Journal of Financial Markets*, *15*, 329–359.

Fishman, M. J., & Longstaff, F. A. (1992). Dual trading in futures markets. *Journal of Finance*, *47*(2), 643–671.

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47–68.

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance*, *17*(1), 35–105.

Graman, M., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, *53*, 67–78.

Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *Journal of Finance*, *31*(2), 573–585.

Grossman, S. J., & Miller, M. H. (1988). Liquidity and market structure. *Journal of Finance*, *43*(3), 617–633.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408.

Haigh, M. S., Hranaiova, J., & Overdahl, J. A. (2005). *Price dynamics, price discovery and large futures trader interactions in the energy complex*. U.S. CFTC Working Paper, pp. 1–44.

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

Hartzmark, M. L. (1987). Returns to individual traders of futures: Aggregate results. *Journal of Political Economy*, *95*(6), 1292–1306.

Hartzmark, M. L. (1991). Luck versus forecast ability: Determinants of trader performance in futures markets. *Journal of Business*, *64*(1), 49–74.

Hicks, J. R. (1939). *Value and Capital*. Oxford and New York: Oxford University Press.

Hirshleifer, D. (1988). Residual risk, trading costs, and commodity futures risk premia. *Review of Financial Studies*, *1*(2), 173–193.

Hirshleifer, D. (1989). Determinants of hedging and risk premia in commodity futures markets. *Journal of Financial and Quantitative Analysis*, *24*(3), 313–331.

Hong, H., & Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, *105*, 473–490.

Houthakker, H. S. (1957). Can speculators forecast prices? *Review of Economics and Statistics*, *39*(2), 143–151.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economics Studies*, *7*(1), 1–27.

Kang, W., Rouwenhorst, K. G., & Tang, K. (2014). A tale of two premiums: The role of hedgers and speculators in liquidity provision to commodity futures markets. *Journal of Finance*, forthcoming, 1–53. http://ssrn.com/abstract=2449315

Keynes, J. M. (1923). "Some aspects of commodity markets". manchester guardian commercial, reconstruction supplement, *The Collected Writings of John Maynard Keynes* (p. 12). London: Macmillan.

Kucher, O., & Kurov, A. (2014). Business cycle, storage, and energy prices. *Review of Financial Economics*, *23*, 217–226.

Kurov, A. (2008). Information and noise in financial markets: Evidence from the E-mini index futures. *The Journal of Financial Research Fall*, *31*(3), 247–270.

Kurov, A., & Lasser, D. J. (2004). Price dynamics in the regular and E-mini futures markets. *Journal of Financial and Quantitative Analysis*, *39*(2), 365–384.

Kuserk, G. J., & Locke, P. R. (1993). Scalper behavior futures markets: An empirical analysis. *Journal of Futures Markets*, *13*(4), 409–431.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1336.

Labuszewski, J., Nyhoff, J. E., Co, R., & Peterson, P. (2010). *The CME Group Risk Management Handbook, Products and Applications*. Hoboken, NJ: John Wiley and Sons, Inc.

Leuthold, R. M., Garcia, P., & Lu, R. (1994). The returns and forecasting ability of large traders in the forzen pork bellies futures market. *Journal of Business*, *67*(3), 459–473.

Leuthold, R. M., & Kim, M.-K. (2000). Managing overnight corn price risks: E*hedging versus tokyo. *Journal of Agribusiness*, *18*(3), 275–288.

Locke, P., & Onayev, Z. (2007). Order flow, dealer profitability, and price formation. *Journal of Financial Economics*, *85*, 857–887.

Manaster, S., & Mann, S. (1996). Life in the pits: Competitive market making and inventory control. *Review of Financial Studies*, *9*, 953–975.

Manaster, S., & Mann, S. C. (1999). *Sources of market making profits: Man does not live by spread alone*. Working Paper.

Menkveld, A. J., Sarkar, A., & van der Wel, M. (2012). Customer order flow, interemediaries, and discovery of the equilibrium risk-free rate. *Journal of Financial and Quantitative Analysis*, *47*(4), 821–849.

Rouwenhorst, K. G., & Tang, K. (2012). Commodity investing. *Annual Review of Financial Economics*, *4*, 447–467.

Sanders, D. R., Irwin, S. H., & Merrin, R. P. (2009). Smart money: The forecasting ability of CFTC large traders in agricultural futures markets. *Journal of Agricultural and Resource Economics*, *34*(2), 276–296.

Silber, W. L. (1984). Marketmarket behavior in an auction market: An analysis of scalpers in futures markets. *Journa of Finance*, *39*, 937–953.

Singleton, K. J. (2014). Investor flows and the 2008 boom/bust in oil prices. *Management Science*, *60*(2), 300–318.

Stoll, H. R. (1979). Commodity futures and spot price determiantion and hedging in capital market equilibrium. *Journal fo Financial and Quantitative Analysis*, *14*(4), 873–895.

Strong, R. A. (2009). *Portfolio Construction, Management, and Protection* (5th ed.). South Western Cengage Learning.

Vayanos, D., & Wang, J. (2012). Theories of liquidity. *Foundations and Trends in Finance*, *6*, 221–317.
Wang, C. (2003). The behavior and performance of major types of futures traders. *Journal of Futures Markets*, *23*, 1–31.
Wang, C. (2004). Futures trading activity and predictable foreign exchange market movements. *Journal of Banking and Finance*, *28*, 1023–1041.
Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy*, *102*(1), 127–168.
Wiley, M. K., & Daigler, R. T. (1998). Volume relationships among types of traders in the financial futures markets. *Journal of Futures Markets*, *18*, 91–113.
Working, H. (1949). The theory of price storage. *American Economic Review*, *39*(6), 1254–1262.

How to cite this article: Llorente G, Wang J. Trading and information in futures markets. *J Futures Markets*. 2020;40:1231–1263. https://doi.org/10.1002/fut.22079

APPENDIX A

A.1 Volume from the end of the day (EOD) and Liquidity Data Bank (LDB) files

This section presents the differences between the daily total volume from the EOD and LDB data files. Figure A1 Panel (a) presents the EOD volume, Panel (b) the LDB volume, Panel (c) present the difference between both volumes and

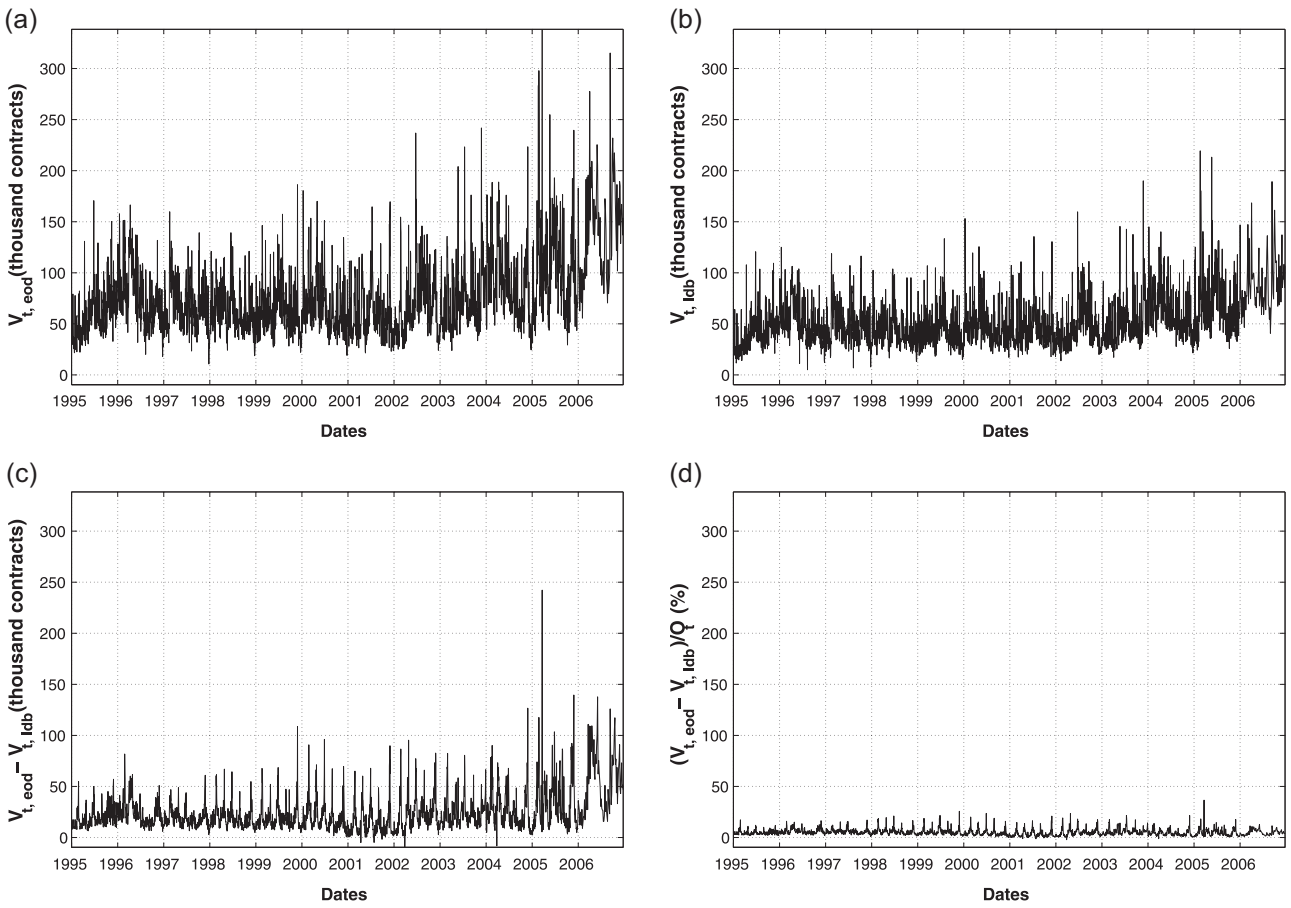

FIGURE A1 Daily market volume comparison between EOD volume and LDB volume. The differences between the reported market volume in the EOD files and the market volume calculated from LDB files is shown. (a) The total market volume in thousand contracts from the EOD files, (b) the total market volume in thousand contracts calculated from the LDB files, (c) difference between EDO and LDB volumes in thousand contracts, and (d) the difference between EOD and LDB volume as a percentage of the total open interest, respectively. EOD, end of the day, LDB, Liquidity Data Bank

TABLE A1 Descriptive statistics for the total market volume figures either from the end of day (EOD) files or from the Liquidity Data Bank (LDB) files

	Mean	Median	*SD*	Skew	Kurt
$V_{t,\mathrm{EOD}}/1{,}000$	74.048	65.676	35.373	1.667	7.849
$V_{t,\mathrm{LDB}}/1{,}000$	51.112	45.778	23.553	1.575	7.310
$(V_{t,\mathrm{EOD}} - V_{t,\mathrm{LDB}})/1{,}000$	22.936	17.833	17.667	2.621	16.754
$(V_{t,\mathrm{EOD}} - V_{t,\mathrm{LDB}})/\mathrm{OI}_t(\%)$	5.053	4.325	3.277	1.870	9.968

Note: This table presents the descriptive statistics for the total market volume figures either from the end of day (EOD) files or from the Liquidity Data Bank (LDB) files. The figures are measured in the number of contracts. The sample period is 2003–2006.

Panel (d) the difference normalized by the daily open interest. Table A1 presents the descriptive statistics of the graphs in Figure A1.

The numbers in the LDB files are made public by the Chicago Board of Trade (CBOT) after the second reconciliation of the unmatched trades. There is some controversy about how many unreconciliated trades are missing. CBOT says that 90% of the out trades are cleared in the second reconciliation (Strong, 2009, p. 580). Taking into account these references and the numbers in Table A1, we conclude that in the LDB files about 31% of the total EOD volume is missing, or 5% in terms of Open Interest. As it was pointed in the main text the correlation between both volume measures (EOD and LDB) is 0.983.

As mentioned in the text, for the period from January 2, 2003 to the end of our sample, May 8, 2006, CBOT also provided us the data on the electronic and spread volume. Table A2 shows the mean daily contracts traded on each venue during these last years of the sample. The numbers are mean daily trading contracts in thousands.

A.2 Customer type indicator (CTI) definitions

The following CTI definitions were taken from the CBOT web site before it merged with the CME to form the CME group. They correspond to what was rule number 332.03 in the CBOT Rule Book. Now the CTI definitions are in Rule 536.D on the Rule Book of the CME Group (see www.cmegroup.com web site).

RULE 332.03 Customer Type Indicator (CTI) Codes.[44] Each clearing member must identify each transaction executed on the Floor of the Exchange or on e-cbot, on the record of transactions that is submits to the Clearing Services Provider, with the correct CTI code. The CTI codes are as follows:

CTI: Electronic Trading and Open Auction: Applies to transactions initiated and executed by an individual member for his/her own account, for an account he/she controls, or for an account in which he/she has an ownership or financial interest. However, transactions initiated and executed by a member for the proprietary account of a member firm must be designated as CTI2 transactions.

CTI: Electronic Trading and Open Auction: Applies to orders entered/trades executed for the proprietary accounts of a member firm or Qualified Affiliate, as defined in Regulations 230.02 or 230.03. Members initiating and executing trades by open auction for the proprietary accounts of a member firm may participate in the Exchange CTI 2/1 Conversion Program.

CTI: Applies to orders entered by a member or a nonmember Registered User for the account of another individual member or an account controlled by such other individual member.

CTI: Applies to orders that a member executes on behalf of another individual member, or for an account such other member controls or in which such other member has an ownership or financial interest.

CTI: Applies to all orders/transactions not included in CTI categories 1, 2, or 3. These typically are orders entered by or on behalf of nonmember entities.

[44]Source: CBOT Rules, www.cbot.com.

TABLE A2 Mean daily contracts traded in the EOD file and LDB files broken by venue and type of trade

	Mean
$V_{t,\text{EOD}}$ (in thousands)	93.948
$V_{t,\text{LDB}}$ (in thousands)	66.386
$V_{t,\text{ no spread-LDB-pit}}$ (in thousands)	61.659
$V_{t,\text{ no spread-LDB-electro}}$ (in thousands)	1.496
$V_{t,\text{ spread-LDB-pit}}$ (in thousands)	2.319
$V_{t,\text{ spread-LDB-electro}}$ (in thousands)	0.912
$(V_{t,\text{EOD}} - V_{t,\text{ no-spread-LDB-pit}})$ (in thousands)	32.289
$(V_{t,\text{EOD}} - V_{t,\text{ no-spread-LDB-pit}})/OI_t$ (%)	5.226

Note: This table presents the descriptive statistics for the total market volume figures either from the End of Day (EOD) files or from the Liquidity Data Bank (LDB) files broken by venue (pit vs. electronic) and type of trade (no spread vs. spread). The figures are measured in number of contracts. The sample period is 1995–2006.

As it was already said in the main text, on December 1, 2004, took place the agreement among the Exchanges and the National Futures Association to "Harmonize the set of CTI code definitions for the industry." The main differences from the previous CBOT classification was:

- CTI2 includes all trades from proprietary account. Previously it was exclusively reserved for the proprietary account of clearing members. It will also include members receiving preferential rates for their trading, they were previously classified as CTI4.
- CTI3 will also include open outcry orders entered for a member's account through the normal customer order flow process, previously they were coded as CTI4.

That means that after December 1, 2004, some of the trades previously coded in the open outcry trading as CTI4, afterward they were coded either as CTI2 or CTI3.

A.3 Dynamics of market variables and P&L

As shown in Table 1, changes in market variables such as turnover and normalized high minus low prices, exhibit significant persistence. Table A3 presents the MA(2) estimation of the these variables and Table A4 present the characteristics of the residuals in the MA(2) estimation.

Table A5 presents the correlations among the residuals of the MA(2) for the market variables.

As noted in the main text, given the persistence in P&L, we consider changes in P&L, which still exhibits nontrivial autocorrelation. Following the same approach as for the market variables, we decompose the changes in P&L into a predictable part using an MA(2) model and a residual (Table A6 and Table A7 present the results of the MA(2) estimations for the change in the P&L and the corresponding residuals):

TABLE A3 Parameter estimates for market variables

	c	θ_1	θ_2	R^2
$\Delta\tau_t$	−0.000 [−0.035]	**−0.698** [−26.006]	**−0.156** [−6.333]	31.411
ΔX_t	0.000 [0.018]	**−0.828** [−35.299]	−0.016 [−0.699]	40.567

Note: This table presents the parameter estimates for changes in market variables using an MA(2) model. The sample period is 1995–2006. The t statistics are inside brackets. Boldface means significative at the 5% level.

TABLE A4 Descriptive statistics for the unexpected parts for market variables

	Mean	*SD*	Skew	Kurt	ρ_1	ρ_2	ρ_5	ρ_{10}
$\hat{u}_{t,\Delta\tau_t}$	−0.000	0.044	1.735	8.575	0.019 (.325)	0.101 (.000)	−0.004 (.842)	−0.009 (.618)
$\hat{u}_{t,\Delta X_t}$	−0.000	0.007	1.433	7.483	0.001 (.966)	0.040 (.033)	−0.008 (.677)	0.014 (.453)

Note: This table presents the descriptive statistics of unexpected parts for market variables. The unexpected parts are calculated as the difference between the observed values and the forecasted (expected) parts of an MA(2) model for the considered variable. ρ_i is the autocorrelation coefficient for lag $i = 1, 2, 5, 10$. The sample period is 1995–2006. p values are inside the parenthesis.

TABLE A5 Correlations among the unexpected parts of the changes in the market variables

	$\hat{u}_{t,\Delta\tau_t}$	$\hat{u}_{t,\Delta X_t}$
$\hat{u}_{t,\Delta\tau_t}$	1.000 (.000)	0.673 (.000)
$\hat{u}_{t,\Delta X_t}$		1.000 (.000)

Note: This table presents the correlations among the unexpected parts of the changes in market variables. The unexpected parts are calculated as the difference between the observed values and the forecasted (expected) parts from an MA(2) model for the considered variable. The sample period is 1995–2006. The p values are inside the parenthesis.

TABLE A6 Parameter estimates for the change in the intraday P&L for the groups CTI1 and CTI4

	c	θ_1	θ_2	R^2
$\Delta\text{PNL}^0_{1,t}$	0.006	**−0.855**	**−0.062**	42.443
	[0.114]	[−28.667]	[−2.257]	
$\Delta\text{PNL}^0_{4,t}$	−0.011	**−0.936**	−0.009	46.650
	[−0.239]	[−35.046]	[−0.413]	

Note: This table presents the parameter estimates from an MA(2) model for the change in the intraday P&L for the groups CTI1 and CTI4 ($\Delta\text{PNL}^0_{i,t}$, $i = 1, 4$). The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets.

TABLE A7 Descriptive statistics for the unexpected parts of the changes in the intraday P&L of groups CTI1 and CTI4

	Mean	*SD*	Skew	Kurt	ρ_1	ρ_2	ρ_5	ρ_{10}
$\hat{u}^0_{t,\Delta\text{PNL}_{1,t}}$	−0.012	33.374	2.585	26.755	−0.001 (.963)	−0.011 (.561)	−0.008 (.688)	0.015 (.431)
$\hat{u}^0_{t,\Delta\text{PNL}_{4,t}}$	0.006	41.834	−2.378	26.931	0.000 (.998)	0.007 (.712)	0.051 (.006)	0.006 (.759)

Note: This table presents the descriptive statistics of the unexpected parts of the changes in the intraday P&L of groups CTI1 and CTI4 ($\hat{u}^0_{t,\Delta\text{PNL}_{i,t}}$, $i = 1, 4$). ρ_i is the autocorrelation coefficient for lag $i = 1, 2, 5, 10$. The sample period is 1995–2006. p Values are inside the parenthesis.

TABLE A8 Regression of the price change of the nearest contract on contemporaneous changes in overnight position of each CTI group and market variables

c	0.000 [1.230]	0.000 [0.742]	0.000 [1.563]	0.000 [1.655]	**0.000** [1.982]	**0.000** [1.958]
$\Delta N_{1,t}$		−0.381 [−1.875]				
$\Delta N_{2,t}$			**−0.782** [−3.269]			−0.512 [−1.875]

(Continues)

$\Delta N_{3,t}$				−0.219 [−1.572]		
$\Delta N_{4,t}$					**0.424** [4.308]	**0.290** [2.483]
$\widehat{\Delta\tau_t}$	**−0.017** [−2.280]	**−0.017** [−2.235]	**−0.016** [−2.124]	**−0.016** [−2.183]	**−0.015** [−1.963]	**−0.015** [−1.958]
$\hat{u}_{t,\Delta X_t}$	0.125 [1.899]	0.122 [1.863]	**0.133** [2.059]	0.120 [1.825]	0.118 [1.803]	0.125 [1.935]
Z_t	**0.103** [17.597]	**0.103** [17.539]	**0.102** [17.081]	**0.104** [17.710]	**0.103** [17.131]	**0.102** [17.031]
$\Delta \ln P_{t-1}$	0.029 [1.448]	0.022 [1.107]	0.025 [1.260]	0.031 [1.555]	0.023 [1.161]	0.023 [1.130]
$\Delta \ln P_{t-2}$	−0.021 [−1.117]	−0.020 [−1.091]	−0.023 [−1.224]	−0.022 [−1.179]	−0.024 [−1.268]	−0.024 [−1.289]
$\Delta \ln P_{t-3}$	−0.036 [−1.914]	−0.036 [−1.933]	−0.036 [−1.941]	**−0.040** [−2.100]	**−0.044** [−2.301]	**−0.042** [−2.223]
$\Delta \ln P_{t-5}$	**−0.050** [−2.541]	**−0.051** [−2.577]	**−0.050** [−2.569]	**−0.052** [−2.570]	**−0.054** [−2.668]	**−0.053** [−2.651]
R^2	17.820	18.065	19.002	17.949	19.058	19.423

Note: This table presents the regression of the price change of the nearest contract ($\Delta \ln P_t^{(1)}$) on contemporaneous changes in overnight position of each CTI group ($\Delta N_{i,t}$) and market variables (decomposed into expected and unexpected parts). The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The t statistics are inside brackets.

$$\Delta \mathrm{PNL}_{i,t}^0 = \widehat{\Delta \mathrm{PNL}}_{i,t}^0 + \hat{u}_{i,\Delta \mathrm{PNL},t}^0, \tag{A1}$$

where $\widehat{\Delta \mathrm{PNL}}_{i,t}^0$ is the forecast of $\Delta \mathrm{PNL}_{i,t}^0$ from the estimated MA(2) model, which is summarized in Table A6.

A.4 Robustness

The robustness checks refer to the price discovery and price forecasting when using as objective either discover or forecast the price of the nearest contract instead of the average of the four nearest contracts to mature, as it was presented in the main text. Table A8 corresponds to the price discovery, Table A9 refers to the price forecasting using regressions and Table A10 to the forecasting using nonparametric analysis. As it was already pointed the conclusions do not change.

TABLE A9 Regression of the nearest price change on lagged changes in position and lagged changes in market variables

c	−0.000 [−1.422]	−0.000 [−1.399]	−0.000 [−1.409]	−0.001 [−1.953]	−0.000 [−1.606]
$\Delta N_{1,t}$		−0.013 [−0.100]			
$\Delta N_{2,t}$			−0.012 [−0.109]		
$\Delta N_{3,t}$				0.306 [**2.592**]	
$\Delta N_{4,t}$					−0.138 [−1.904]
$\hat{u}_{t,\Delta\tau}$	**0.015** [2.533]	**0.015** [2.532]	**0.015** [2.532]	**0.016** [2.615]	**0.015** [2.511]
Z_t	**0.014** [2.585]	**0.014** [2.585]	**0.014** [2.584]	**0.013** [2.442]	**0.014** [2.522]

(Continues)

$\Delta \ln P_t$	**0.052** [2.435]	**0.052** [2.428]	**0.052** [2.382]	**0.055** [2.585]	**0.057** [2.621]
$\Delta \ln P_{t-1}$	0.004 [0.180]	0.004 [0.171]	0.004 [0.178]	0.002 [0.074]	0.006 [0.264]
$\Delta \ln P_{t-2}$	−0.035 [−1.627]	−0.035 [−1.628]	−0.035 [−1.630]	−0.033 [−1.518]	−0.034 [−1.560]
$\Delta \ln P_{t-4}$	**−0.052** [−2.288]	**−0.052** [−2.288]	**−0.052** [−2.290]	**−0.047** [−2.081]	**−0.049** [−2.177]
R^2	1.229	1.194	1.194	1.504	1.328

Note: This table presents the regression of the nearest price change ($\Delta \ln P_{t+1}^{(1)}$) on lagged changes in position ($\Delta N_{i,t}$) and lagged changes in market variables (decomposed into expected and unexpected parts). The R^2 is adjusted by degrees of freedom and in percentage. The sample period is 1995–2006. The *t* statistics are inside brackets.

TABLE A10 Nonparametric tests on the distribution of future price changes of the nearest contract conditional on changes in overnight positions of each CTI group

	$\Delta \ln P_{t+1}$		
Panel A	$\Delta N_{1,t} > 0$	$\Delta N_{1,t} < 0$	***p* Value** ($\theta_+ = \theta_-$)
Mean	−0.000	−0.001	0.625
SD	0.014	0.013	
Skewness	0.196	0.126	0.740
Kurtosis	5.361	4.328	0.089
No. of obs.	1,239	1,587	
K–S test	F_+	F_-	0.559
Panel B	$\Delta N_{2,t} > 0$	$\Delta N_{2,t} < 0$	***p* Value** ($\theta_+ = \theta_-$)
Mean	−0.000	−0.001	0.654
SD	0.013	0.014	
Skewness	0.394	−0.055	**0.030**
Kurtosis	5.175	4.544	0.305
No. of obs.	1,474	1,343	
K–S test	F_+	F_-	0.602
Panel C	$\Delta N_{3,t} > 0$	$\Delta N_{3,t} < 0$	***p* Value** ($\theta_+ = \theta_-$)
Mean	0.000	−0.001	**0.002**
SD	0.013	0.014	
Skewness	0.428	−0.161	**0.005**
Kurtosis	5.316	4.172	**0.070**
No. of obs.	1,719	1,105	
K–S test	F_+	F_-	**0.043**
Panel D	$\Delta N_{4,t} > 0$	$\Delta N_{4,t} < 0$	***p* Value** ($\theta_+ = \theta_-$)
Mean	−0.001	−0.000	0.053
SD	0.014	0.013	
Skewness	−0.076	0.358	**0.043**
Kurtosis	4.406	5.203	0.200
No. of obs.	1,177	1,646	
K–S test	F_+	F_-	0.349

Note: This table shows the nonparametric tests on the distribution of future price changes of the nearest contract ($\Delta \ln P_{t+1}^{(1)}$) conditional on changes in overnight positions of each CTI group($\Delta N_{i,t}$, $i = 1, 2, 3, 4$). Panel A is for CTI1 group, Panel B is for CTI2 group, Panel C is for CTI3 group, and Panel D is for CTI4 group. The K-S test means the Kolmogorov–Smirnov test of the equality of the two conditional future price changes distributions. The *p* values are obtained using bootstrap statistics with 5,000 simulations. Boldface signifies different from zero at the 5% level. The sample period is 1995–2006.