Do old limit orders contain information?

By

Daniel Justin Daniels

A Thesis submitted to the Faculty of Graduate Studies

The University of Manitoba

In partial fulfillment of the requirements of the degree of

Master of Science

Agribusiness and Agricultural Economics

University of Manitoba

Winnipeg

Abstract

Studies in agricultural futures markets have shown that limit orders carry significant information and contribute to the price discovery process. This means that orders arriving to the exchange contain information coming from different sources. Public information is easily tractable and orders submitted to the exchange after a public announcement presumably contain similar information. However, private information is hard to track and differs across sources. Until recently, it was not possible to identify individual orders and assess their information content. With new data availability, it is now possible to identify and track individual orders throughout their lifespan and further our understanding of the private information carried in limit orders and their role in the price discovery process. The study of the microstructure of the agricultural futures markets has been significantly improved with the advent of order level data with nanosecond precision timestamps. We use the new available dataset for CME corn and soybean futures markets from December 2018 to December 2019 and categorize each event in the life of an order to define the orders' lifespan. We then calculate the Modified Information Share on a series of age categories from 10 minutes to 80 minutes to study the effect that age has on the contribution to price discovery. We find that younger orders contain more information than older orders, and this effect is persistent beyond 80 minutes. This finding shows that private information is much longer lived than public information.

Acknowledgements

I would like to sincerely thank Dr. Julieta Frank for her time, effort and patience in working with me throughout my master's program. Her mentorship and guidance were valuable in helping me to accomplish my goals as it allowed me to see the strengths and weaknesses of my efforts. I would also like to thank Dr. Mehdi Arzandeh for the many hours of meetings and insightful questions and direction that he provided, and Dr. Barry Coyle for his thought-provoking feedback and Dr. Ali Kerrache for his help with the Grex High Performance Computing (HPC) system. This paper would not have been possible without the experience of my committee; for this I am immensely grateful.

Financial support from the Natural Sciences and Engineering Research Council (NSERC), the Graduate Enhancement of Tri-Agency Stipends (GETS), and the Winnipeg Commodity Exchange Fellowship are gratefully acknowledged. This research was enabled in part by research computing support provided by Grex at the University of Manitoba HPC Centre.

I would also like to thank my family and especially Leanne for putting up with all the time I spent exploring rabbit holes, and my fellow students who inspired me to finish.

Dedicated to my parents for encouraging me to always explore new things.

List of Tables

Table of Figures

Table of Contents

1. Introduction

Historically, research in agricultural futures markets has been performed using the open, high, low, and close prices which were manually collected, published, and archived daily. With the advent of electronic trading data capture and dissemination have increased exponentially to include every order entered and every individual trade, all marked with nanosecond (10^{-9}) timestamps.[1] These data has opened new opportunities to study price discovery of a single commodity (asset); including an examination of the market's action that results from order flow, day traders, liquidity providers, and even algorithms (Hasbrouck 2021). Lehecka, Wang, and Garcia (2014) showed the value of the intraday dataset when studying the speed at which the market incorporated new public information. According to Holder and Lagrehr (2001), pit traders and brokers working on the floor often developed a "feel" and a "sense" for the market that was only available to participants that were close to the paper flow. Indeed, it may be that the "feel" and "sense" for the market that many open outcry participants reported comes from a combination of last price and order-flow.

This paper utilizes the newly available Market by Order (MBO) dataset from the Chicago Mercantile Exchange Group (CME) to calculate information shares on subgroups of individual orders based on their age.[2] To our knowledge, data with this level of aggregation has not been used to examine the price discovery process. We show that limit orders less than 60 minutes old and resting in steps 1 through 15 of the limit order book (LOB) carry a substantial amount of information, and that this information is relatively persistent. In this paper, we utilize the recently available data from the CME to investigate the role that the age of a limit order plays in price discovery relative to the last trade price. The MBO dataset is parsed, prepared, and analyzed using the Modified Information Share metric developed by Lien and Shrestha (2009) to determine the impact of a partial LOB (pLOB) on price discovery. Partial LOBs are constructed using different ranges of order ages and LOB steps.

[1] Before electronic trading, no dataset of limit orders existed as limit orders were not allowed to be disseminated beyond the best bid-ask (CME rule 532). In Chicago futures pits, orders were received by brokers and marked with a 15-minute interval for the timestamp, and then only formed part of the market when the order matched the best bid or best ask.

[2] The CME began disseminating order-based data in 2017 and these data are available for purchase at https://www.cmegroup.com/confluence/display/EPICSANDBOX/MBO+FIX

2. Motivation

Before 2018, the CME disseminated Market by Price (MBP) data, with the limit orders being aggregated by price onto "steps" in the LOB, and then disseminated to trading platforms. These data are limited to a set number of steps depending on the commodity, for example, corn and soybean data are limited to 10 steps while lean hog data are limited to 5 steps. Cao, Hansch, and Wang (2009) and Arzandeh and Frank (2019) studied the contribution from orders on various steps of the LOB and found that the limit orders play an important role in the price discovery process.

In 2017, the CME began disseminating MBO data. These data forms part of the digital information that is sent to trading applications, as well as being archived and available for historical analysis. This dataset includes several enhancements compared to the MBP data, including order level information allowing individual orders to be identified and tracked. Orders can be classified as limit orders and market orders, with limit orders having a lifespan, and market orders triggering an immediate trade. It is also possible to identify and classify events in an order's life including initial order placement, any order amendment, partial fill, complete fill, and order cancellation. Also, the dataset includes all orders as opposed to just the first 5 or 10 steps which were available in the MBP dataset. By using the time when an individual order is submitted, the age of each order can be calculated at any point in time, and this can be used along with the LOB steps to group limit orders in subsets of the LOB for analysis.

For this study, we construct books of up to 15 steps with orders of various ages. We aggregate the new and old LOB into distinct time series of quantity-weighted average prices. Snapshots of the trades and orders are taken every ten seconds of the daytime session, beginning 0.5 seconds after the opening and ending 0.5 seconds before the closing of trade.

3. Objective

Arzandeh and Frank (2019) showed that the LOB beyond the best bid and ask for several agricultural commodities is informative, however their analysis included all open orders on each step regardless of age and it was limited to the first 10 steps of the LOB due to the MBP dataset. The objectives of this study are to:

a) Estimate the informational content of limit orders of different ages in the LOB, and;
b) Assess the contribution of limit orders further away in the LOB, beyond the first 10 steps used in previous studies, to the price discovery process.

3.1 Informational content of orders of different ages

Historically, price discovery models were analysed using the last trade price and a subset of bid and ask prices that were dependent only on the steps that orders appeared on. However, at any point in time orders of various ages make up the LOB. We restrict the orders that can be included in the analysis using various, arbitrary, age limits and define two subgroups, the "new" partial LOB (pLOB) and the "old" pLOB. The "new" pLOB changes when traders add new orders, amend orders, or when new orders either trade or become "old"; while the "old" pLOB changes when orders enter the category from the "new" category as a result of the order age, orders leaving the category as a result of traders canceling orders, traders amending orders, and orders trading. In both subgroupings, the aggregate pLOB is continually updated based on new activities (both the "new" and "old" subgroups are updated as a result of new input from market participants and from market activities). We estimate the information share for the "new" and "old" pLOBs. Our expectation is that new and old orders contribute differently to price discovery. To our knowledge no studies have been performed analysing the impact of order age on the information content of the LOB.

3.2 Information content of a large LOB

Order information from the MBO files can be categorized with an order type of bid, ask, or trade, and an update type of new, amend, or cancel. At any point in time, orders that have been entered (new) but have not been canceled or traded form the LOB. A complete LOB is made up of the complete set of limit orders that are open at any point in time. A pLOB is made up of all orders that meet a certain set of criteria. Often the criteria are as simple as limiting the pLOB to a certain number of steps. In addition to building two pLOBs, one made from orders that are younger than certain age and one from orders that are older than that age, we include additional open orders relative to the 10-step maximum that was possible with the MBP data.

3.3 Validation

Arzandeh and Frank (2019) calculate the information shares of the LOB using MBP data. We perform a similar analysis to validate our process of parsing the MBO files, aggregating pLOBs and calculating the Modified Information Shares. This is an important check since the MBO files have not been used to recreate the LOB nor has the age of orders been used to aggregate age dependent pLOBs.

4. Description of the data

The CME offers trading in standardized futures contracts including the corn and soybean futures used for this study. Futures contracts trade in highly standardized units that include all specifications except price and quantity. The standardized contracts allow bids and asks (offers) to be submitted with only three pieces of information; the product identifier, the quantity, and the price. The corn and soybean futures that comprise our dataset use a 0.0025 $/bushel price tick and a 5,000 bushel/contract unit size. As discussed by Hasbrouck (2021), the price tick constraint results in individual bid, ask, and trade prices being confined to a minimum grid. The last trade price at any point in time maintains this grid, while the process of aggregating the bid and ask prices into a single pLOB price results in much finer detail. The CME provides the electronic platform to accept orders, the matching algorithm that generates trade, and a data dissemination mechanism to inform participants of the market activity (including the state of previously submitted orders).

The CME trading day consists of two distinct periods of trade preceded by a short period for pre-market order entry, during which trades do not occur (Table 4.1). Both the corn and soybean contracts have identical hours. The trading session opens at 16:45 the day before the official trading day with pre-market order placement, followed by trade beginning at 19:00 translating to 24:00 UTC during daylight savings time and 01:00 during standard time. Data for this study were exposed to standard time (CST is UTC minus 6 and CDT is UTC minus 5).

Table 4.1: Trading schedule for CME corn and soybean futures

	Central Daylight Saving Time (CDT)	Coordinated Universal Time (UTC)
Pre-Open CME Globex (Trading Day minus 1)	4:45pm - 7:00pm 16:45 - 19:00	9:45pm – 12:00am 21:45 - 00:00
CME Globex Trading Session "Evening session"	7:00pm - 7:45am 19:00 - 07:45	12:00am – 12:45pm 00:00 - 12:45
Pre-Open for daytime session	8:00am – 8:30am 08:00 - 08:30	1:00pm – 1:30pm 13:00 - 13:30
Trading Session "Daytime Session"	8:30am – 1:20pm 08:30 – 13:20	1:30pm - 6:20pm 13:30 - 18:20

4.1 Market by Order (MBO) dataset

The CME MBO data files consist of a single Financial Information eXchange (FIX) format file for each day (including Sunday, excluding Saturday) for each product. The CME provides the datasets in compressed files that average 46 MB for the weekday files and 354MB when uncompressed. The files are readable by standard text readers, with each line representing a FIX packet. FIX packets contain all the information needed by trading and reporting applications, including market status information, MBP information and MBO information. The FIX data files include every single order submitted, amended, and canceled by traders and by the trading algorithm, and all these orders are marked with a unique identifier that can be used to track an individual order throughout its lifespan. We parse all individual orders from the FIX files.

This study uses MBO data from December 02, 2018, to December 27, 2019 (56 weeks) for the CME corn and soybean futures contracts. During the study period, there were 10 holidays (Table 4.1.1) and 4 days with shortened trading sessions (Table 4.1.2) when the market closes at 12:15 pm instead of the normal 1:30 pm. The dataset included 21.027 million corn orders and 29.438 million soybean orders.

Table 4.1.1: Trading holidays

Date	**Description**
December 25, 2018	Christmas
January 1, 2019	New Years
January 21, 2019	Martin Luther King Jr. Day
February 18, 2019	Presidents Day
April 19, 2019	Good Friday
May 27, 2019	Memorial Day
July 4, 2019	Independence Day
September 2, 2019	Labour Day
November 28, 2019	Thanksgiving
December 25, 2019	Christmas

Table 4.1.2: Shortened trading days

Date	**Description**
December 24, 2018	Christmas Eve
July 3, 2019	Day before Independence Day
November 29, 2019	Day after Thanksgiving
December 24, 2019	Christmas Eve

4.2 Frequency of observations

Time series analysis is usually performed using data with uniform intervals between observations. In the futures market, however, orders are submitted to the market in nearly continuous time and recorded with nanosecond precision, resulting in a stream of observations that are not separated by regular intervals. Hasbrouck (2007, p90) discusses the use of "Wall-Clock" time and "Event" time in microstructure studies. Event time takes observations as a result of one or more of the variables changing. Wall-Clock time takes observations at equal time increments. Both mechanisms require an arbitrary decision on how to define the time series increment. Cao, Hansch, and Wang (2009) and Arzandeh and Frank (2019) constructed their time series based on regular time intervals, the later choosing the average time between trades of different prices. Regardless of the method chosen, the critical impact on the analysis is that the observations be chosen to keep as many observations as possible while minimizing the observations when nothing is changing. This is particularly problematic in time series where repeated observations in all the variables in the model cause collinearity, in which case the estimation process will fail to produce consistent results if there are instances where the observations do not change for longer than the lags in the model. There is a balance between frequent updates to ensure the microstructure data is optimally integrated into the model and having the model provide consistent results. We choose to use 10-second snapshots, as any shorter time between snapshots complicates model estimation. The selected 10-second snapshot is also consistent with Arzandeh and Frank (2019) who used 8.63 and 7.60 seconds for corn and soybean, respectively.

Any interval that is chosen will miss updates of some variables; indeed, this is a feature of time series datasets that deal with non-uniform update intervals. With microstructure data the intervals of updates for each variable of interest are extremely non-uniform, that is, at time some variables will undergo changes much more rapidly than others, and at other times different variable will undergo changes much more rapidly. This issue cannot be addressed by switching between event time and clock time. Indeed, with microstructure data, the issue of repeated observations needs to be balanced against the issue of missed observational updates.

Another characteristic of the data is trade prices that bounce up and down between the best bid and best ask price. If the snapshot frequency is set too low, then the entire lag structure of the model may be dominated by unchanged prices of random duration.

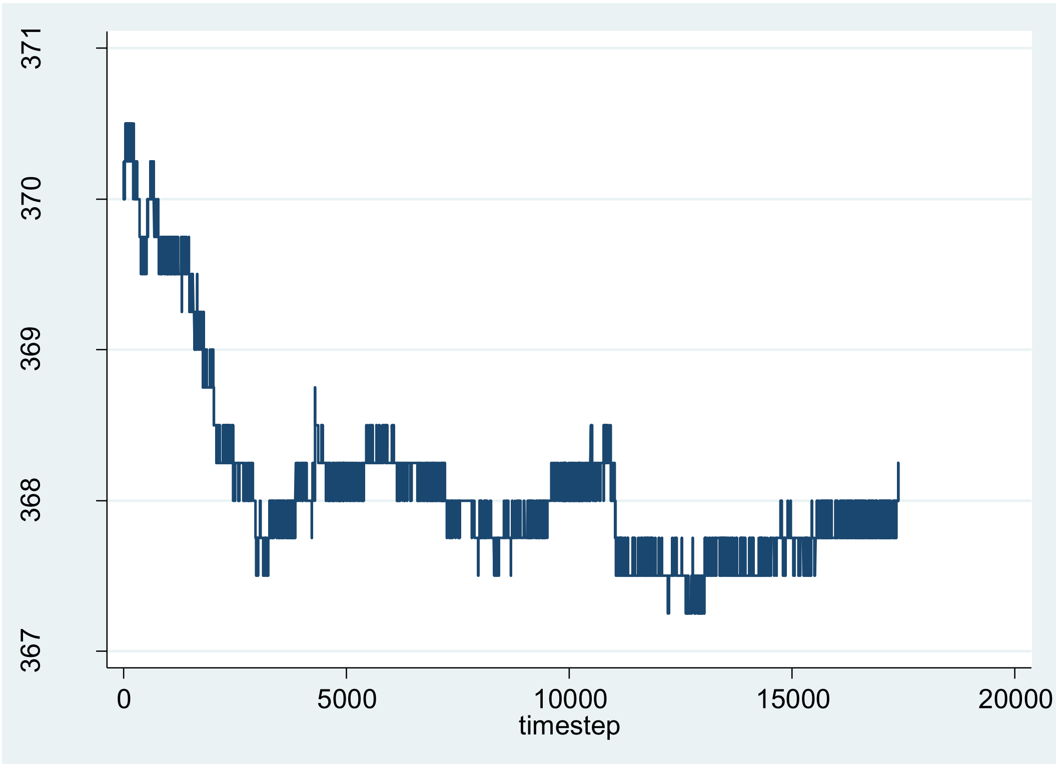
Our analysis includes all orders regardless of when the order was sent. Snapshots of the available (current) orders observed in the LOB were only taken during the daytime session, as these data includes more frequent updates (trades and order updates). This provides one timeseries per market day with no gap in the individual time series.

4.3 Limit orders and the LOB

Limit orders are placed in the LOB by specifying the price and the quantity, whereas market orders trade immediately by specifying the quantity of the order only. Market orders trade at the best possible price available when the order is entered, as opposed to limit orders, where the tradeable price is limited to the price, entered on the order. The FIX dataset does not identify the order type, however since we are interested in orders that compose the LOB, in this study we define a limit order as an order that does not instantaneously execute; in other words, we define limit orders as orders with measurable lifespan in the LOB.

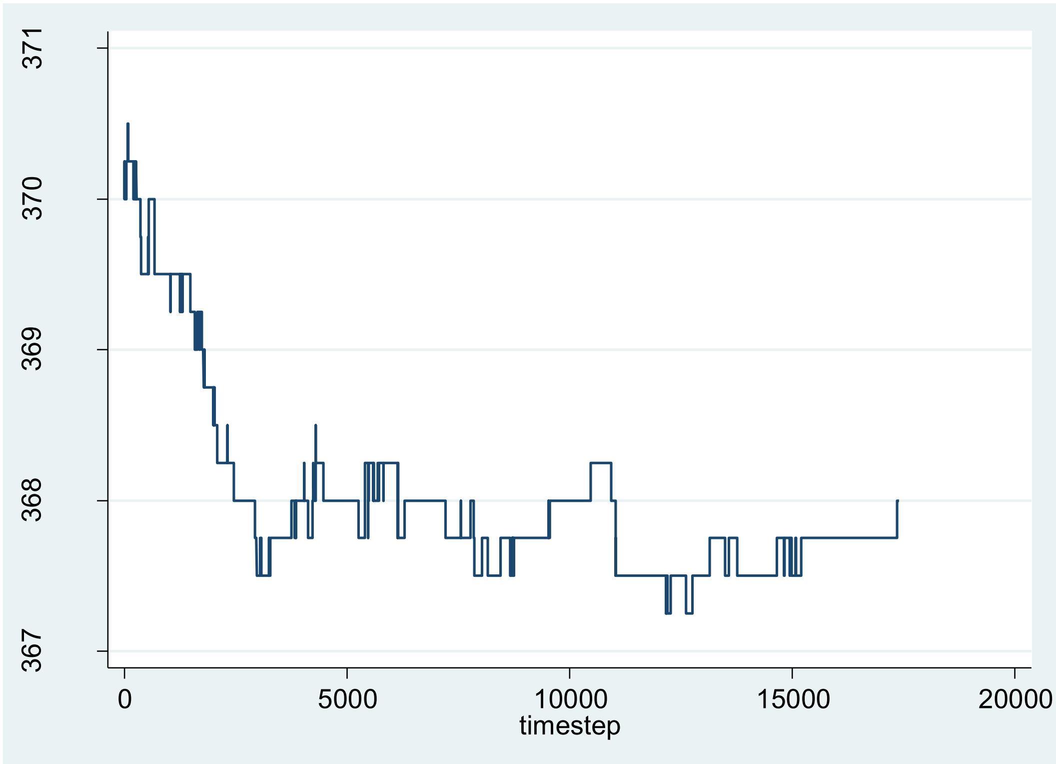
Trade occurs when a bid-side order is matched with a sell-side order. A matching algorithm determines the priority of the orders that are filled when a trade occurs. Both order entry and trade execution occur at the nanosecond timescale. Figure 4.3.1 shows trade prices throughout a single day with 1-second snapshots; of note is the bid-ask bounce where the last trade price moves up and down in a 1 tick range from the best bid to the best ask as trade often occurs in small amounts as new orders are executed (matched) at the best bid or the best ask. This bid-ask bounce is a result of one or more market orders to buy followed by one or more market orders to sell, followed by market order(s) to buy and so on, generally the last tade price bounces between the step 1 bid and the step 1 ask price, and these prices only change if the market buy (sell) orders consume all of the limit sell (buy) orders. This bid-ask bounce is discussed by Engle and Russel (1998).

Figure 4.3.1: Trade price at 1-second snapshots



Similarly, a chart of the best bid price is presented in Figure 4.3.2. This chart indicates much less “bounce” but also fewer price changes. The quantity of the bid orders can be aggregated at each price step, therefore when a trade occurs at the best bid (step 1) it does not necessarily result in a new bid price, unless it consumes all the quantity of that step. Similarly, if additional bid orders arrive at the step 1 price, the bid price remains unchanged. Therefore, the price of the step 1 bid only changes if a bid arrives with a more aggressive price than the previous step 1 bid price (resulting in a new lower step 1 bid price), if market sell orders consume all step 1 bid quantity (resulting in a new higher step 1 bid price), or if traders cancel all step 1 bid quantity (resulting in a new higher step 1 bid price).

Figure 4.3.2: Best bids at 1-second snapshots

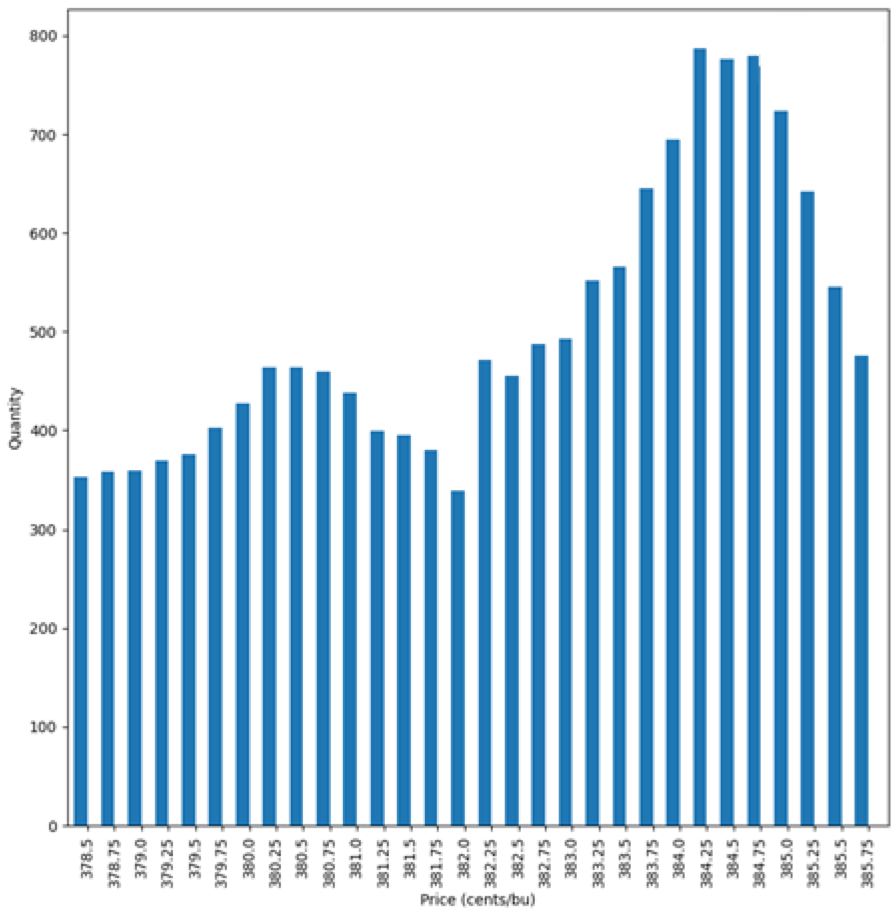


At any point in time, we can mark the last trade price and we can take a snapshot of all orders and use these to build the steps of an aggregate LOB. Additionally, we can select a subset of the bids and ask prices that meet certain criteria and build the steps of an aggregate pLOB.

A challenging feature of these data is the vast difference between the nanosecond precision that the orders and trades occur and the relatively large variation in the time between order and trade updates. Although it may be possible to take snapshots at the nanosecond or microsecond frequency, market updates often occur at a much slower pace, and therefore care must be taken to ensure the snapshots do not occur so frequently that the prices remain unchanged for consecutive snapshots beyond the lag order structure of the model, which results in severe collinearity caused by reoccurring data.

Figure 4.3.3 shows the average LOB for the complete LOB for the March 2019 corn data that was used in this study (December 02, 2018, to December 27, 2019). Note, the sell side (higher prices) of the LOB had more orders on average for this period. This will have the effect of raising the LOB aggregate price (this calculation is explained in section 6.1) but does not impact the information share analysis. An additional notable feature of this data is the lower quantities that occur in the middle of the LOB where trading activity occurs. This results in the orders in the middle of the book (particularly on the step representing the best bid and best ask) disappearing through active trade before larger quantities can accumulate.

Figure 4.3.3: Average limit order book (prices and quantities) for the March corn 2019.



5. Partial limit order books (pLOBs)

5.1 Lifespan of limit orders

The order data parsed from the FIX file contains a nanosecond precision (10^{-9}) timestamp that indicates when the order information is processed.[3] Order updates in the dataset occur as one of three update types: new, amend, or cancel. These can be used to determine when an order lifespan begins, and when it ends by finding the next instance of the order and determining if this instance represents the end of the order lifespan. We define an order lifespan as beginning when a trader enters a new or amended order, and an order lifespan ending when the trader either amends or cancels the order or when the order is fully traded. As a result, order updates that result from partial trade executions do not impact order lifespan, and individual orders amended by the user can have more than one lifespan.

Additionally, orders that remain open over the weekend are resent in a new FIX file with a new order ID that is sent on Sunday afternoon. These order updates appear in the dataset as "new" (as opposed to "amend" or "cancel") and were arbitrarily assigned a start of lifespan on Sunday.[4] As a result, the lifespan of orders is a maximum of 5 days, and orders with a lifespan exceeding one week were always assigned to the "old" pLOB.

For our study period, corn orders have an average lifespan of 201 minutes, and a median lifespan of 7.9 seconds, while soybean orders have an average lifespan of 72:40 minutes and a median lifespan of 2.8 seconds (Table 5.1.1). The lifespan of orders ranges from the nanosecond up to 5 days ($1.157x10^{-14}$ days to 5 days). The distribution of order lifespan is highly skewed with most orders having a very short lifespan.

[3] This study uses FIX Tag 60 (processing time), as opposed to Tag 52 (sending time) which may be used in other studies. Generally processing time precedes sending time. Although the official title for Tag 52 is sending time, it might be better described as 'data dissemination time.'

[4] No attempt was made to match the open orders on Friday to the following resent open order on Sunday, although this may be possible.

Table 5.1.1: Lifespan of orders

Commodity	Observations (count of orders)	Mean Lifespan (seconds)	Median Lifespan (seconds)
Corn	21,026,838	12,035.4	7.9
Soybean	29,438,180	4,360.5	2.8

An order has a lifespan that is bound by 0 and 5 days. The lifespan of orders is extremely skewed with the mean being significantly larger than the median. We measured lifespan in days, therefore a lifespan of 10^{-5} equates to 0.864 seconds, and a lifespan of $e^{-11.36}$ equates to approximately 1 second. Figure 5.1.1 and Figure 5.1.2 show a significant concentration of orders that occur with very small lifespans on the order of 10^5 to 10^7 nanoseconds.

Figure 5.1.1: Lifespan of corn orders

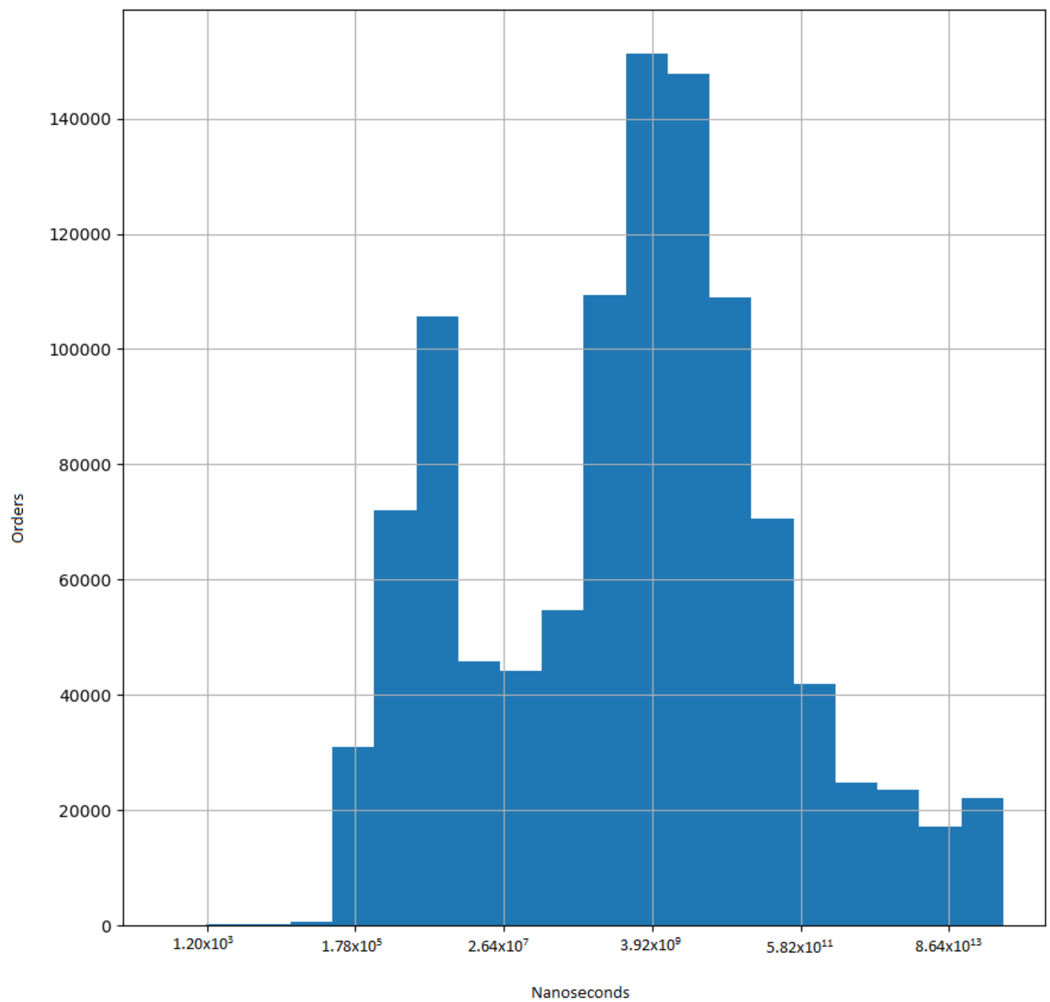
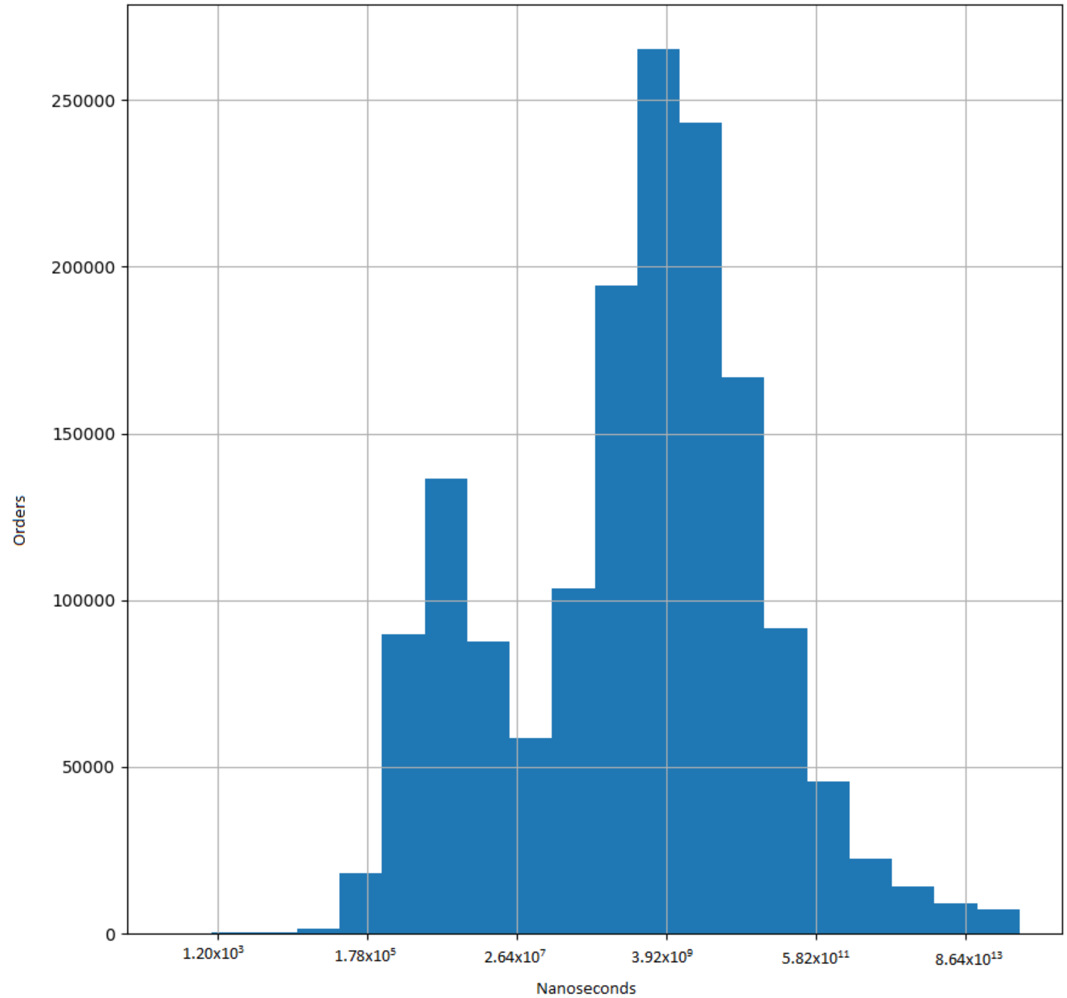


Figure 5.1.2: Lifespan of soybean orders



5.2 Splitting the LOB

The MBO data is reset every week with all open orders being resent with new order identifiers on Sunday each week. These order identifiers are maintained for the duration of the week, as a result, it is possible to track orders from Sunday to Friday's settlement, resulting in the maximum lifespan of orders being 5 days. Orders dated Sunday represent the open orders from the previous week. We can recreate the original (complete) LOB at time t by finding all limit orders that have a start time before (less than) t and an end time greater than or equal to t. Since the MBO dataset includes all orders, we can recreate the entire LOB. To ease processing, we only processed up to 15 steps on each side of the best bid-ask with the expectation that previous studies were limited to 10 steps due to the MBP constraints and can now be extended to include additional steps.

The process to split the LOB by age uses the following process. Consider the following three points in time for each order.

t_1 time the order is entered (order lifespan begins)

t_2 time the order is entered plus the increment (age) used to split orders into "new" and "old"

t_3 time the order is either amended, cancelled by the user, or when the order is fully executed (order lifespan ends)

The three points in time are generally arranged in order, so that $t_1 < t_2 < t_3$ however there are many instances where $t_1 < t_3 < t_2$ where the lifespan is shorter than the time increment (age) used to split the orders into "new" and "old". Using the above notation, we can recreate the entire LOB at any point in time t by filtering for all orders where $t_1 < t < t_3$. Similarly, we can create a pLOB with "new" orders at any point in time by filtering for all orders where both $t_1 < t < t_2$ and $t < t_3$. Additionally, we can create a pLOB with "old" orders at any point in time by filtering for all orders where $t_2 < t < t_3$. That is, we construct two pLOBs, one made up of "new" orders and one made up of "old" orders.

5.3 New and old pLOBs

Traders often place limit orders near the best bid-ask (near the center of the LOB), however as orders age, they have a natural tendency to disappear from the center of the LOB; this occurs due to orders at the center of the LOB having a greater propensity to trade and therefore trigger an end of the lifespan. Using MBO data, we split the LOB into two pLOBs based on the age of the orders. The following figures show the impact of orders disappearing from the center of the LOB as orders age. These figures are an average of the November 2019 soybean data used in our study with Figure 5.3.1 showing both the new and old pLOBs using an age split of 10 minutes. Notably, the new pLOB contains wide price gaps in the steps further from the centre of the book and the old pLOB contains a wide bid-ask price gap.

Figure 5.3.1: Average new and old pLOB using a 10-minute split

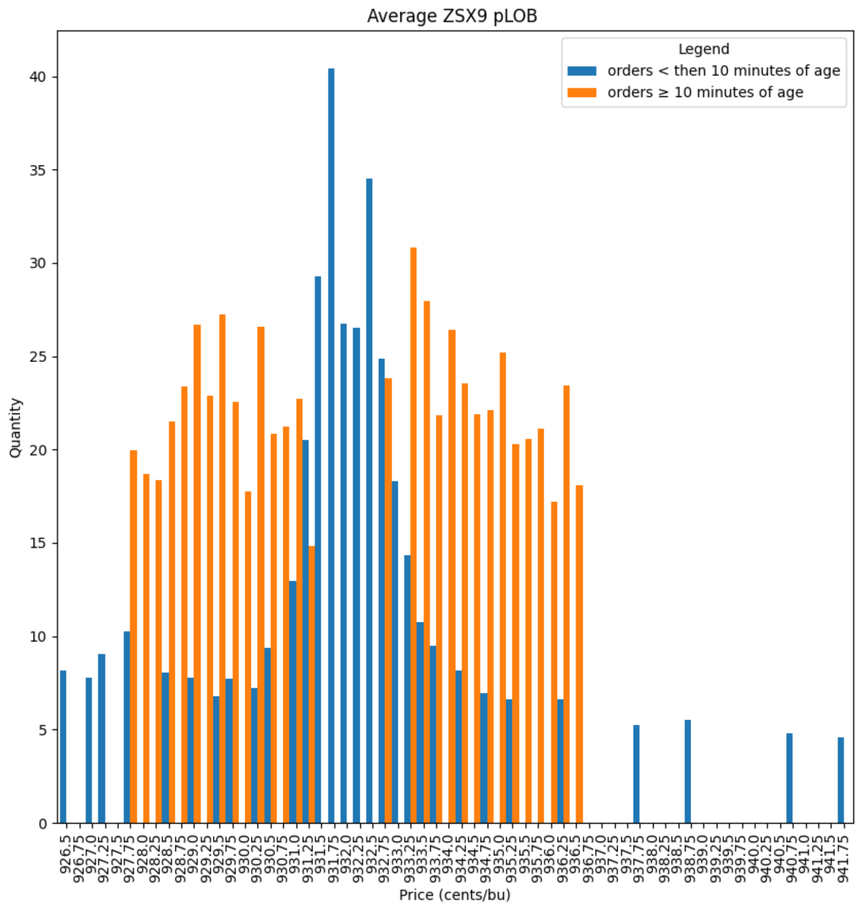
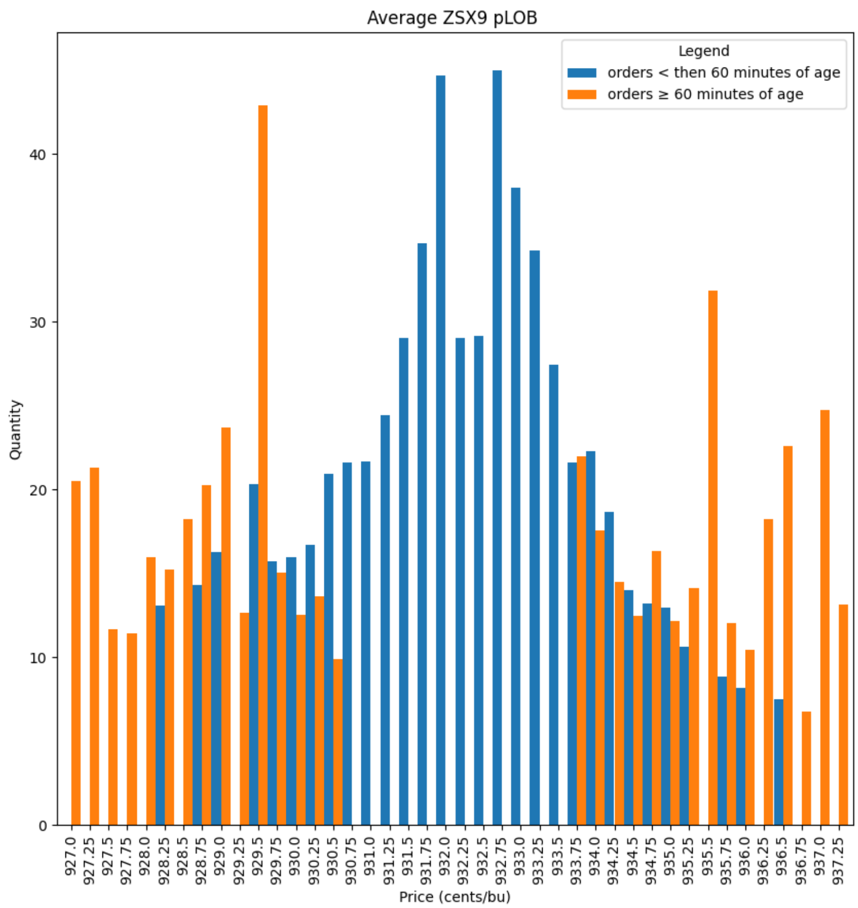


Figure 5.3.2 shows the new and old pLOB using a 60-minute split. This demonstrates the widening of the bid-ask spread that occurs in the old pLOB, particularly notable as the new pLOB takes more of the orders with longer time intervals, and orders have more time to be traded from the center of the old pLOB. In comparison to Figure 5.3.1, the new pLOB has fewer price gaps in the 60-minute new pLOB compared to the 10-minute pLOB.

Figure 5.3.2: Average new and old pLOB using a 60-minute split



As previously seen in the complete LOB (Figure 4.3.3) generally the center of the complete LOB contains orders for each step. As a result of splitting the orders into two subgroups, two properties become evident: 1) new orders are generally placed near the center of the LOB and as a result, the steps further from the centre may be skipping price ticks (average price increments greater than 1 tick), and 2) old orders are less likely to occur at the center of the LOB. This is due to the market oscillations clearing out the center of the LOB before the orders transition to being "old".

There are a few notable characteristics of the LOB described above. The “new” pLOB holds most of the concentration of quantity near the best bid-ask. While the “old” pLOB is much more uniformly spread throughout the 15 steps. Also, the “new” and “old” pLOBs do not necessarily sum to the complete LOB at each step, since the two LOBs can have different bid-ask spread, and they can have different price spreads between each step. Finally, the new pLOB tends to have a narrower bid-ask spread, and the “old” pLOB tends to have narrower spreads between steps compared to the new pLOB after step 4 (Tables 5.3.1 and 5.3.3). For corn futures during our study period, the steps in the LOB are often separated by 1 tick, however, there are short periods where the steps are separated by more than one tick, particularly in the lower steps.

When the LOB is split into two subgroups using the age of the order, the time interval used to split the LOB has an important role in the data in each pLOB. In the case of no new orders being placed at a certain price level for some time, the new LOB will have a declining quantity at that price level as orders move to the “old” pLOB until the quantity at the price is zero, and the step moves to a new price level. As a result, the price difference between steps is expected to be larger in the “new” pLOB than in the complete LOB and in the “old” pLOB. Table 5.3.1 presents the price difference between steps for the “new” pLOB consisting of orders < 40 minutes of age whereas Table 5.3.3 presents the price difference between steps in the “old” pLOB (that is the pLOB comprised of orders ≥ 40 minutes). Similarly, there is a difference between the “new” and “old” pLOB in the total quantity bid and offered at each step, with the “new” pLOB having a much more peaked distribution around the best bid-ask price and the “old” pLOB having a much more uniform distribution. Table 5.3.2 presents the data for the quantities in the “new” pLOB, and Table 5.3.4 presents similar data for the “old” pLOB.

Table 5.3.1: Step price differences in the "new" pLOB (March 2019 corn futures)

	Bid Minimum Price Difference	Bid Maximum Price Difference	Ask Minimum Price Difference	Ask Maximum Price Difference
Step 1-2	0.0025	0.0300	0.0025	0.0775
Step 2-3	0.0025	0.0325	0.0025	0.0175
Step 3-4	0.0025	0.0300	0.0025	0.0075
Step 4-5	0.0025	0.0700	0.0025	0.0050
Step 5-6	0.0025	0.0500	0.0025	0.0250
Step 6-7	0.0025	0.0500	0.0025	0.0600
Step 7-8	0.0025	0.0725	0.0025	0.0500
Step 8-9	0.0025	0.0400	0.0025	0.0500
Step 9-10	0.0025	0.0450	0.0025	0.6000
Step 10-11	0.0025	0.0725	0.0025	0.6000
Step 11-12	0.0025	0.1475	0.0025	0.6000
Step 12-13	0.0025	0.1475	0.0025	0.5950
Step 13-14	0.0025	0.1475	0.0025	0.5950
Step 14-15	0.0025	0.1475	0.0025	0.7000

Table 5.3.2 Step quantities in the "new" pLOB (March 2019 corn futures)

	Max. Quantity Bid	Min. Quantity Bid	Average Quantity Bid	Max. Quantity Ask	Min. Quantity Ask	Average Quantity Ask
Step 1	105,206	1	2,463.73	1	372,700	2,266.80
Step 2	63,841	1	1,487.25	1	74,399	1,667.90
Step 3	28,524	1	684.88	1	73,866	815.79
Step 4	19,196	1	418.48	1	48,024	342.64
Step 5	19,017	1	321.89	1	43,452	220.03
Step 6	19,017	1	198.05	1	42,977	167.06
Step 7	8,855	1	132.65	1	5,355	147.11
Step 8	8,678	1	100.36	1	5,389	96.46
Step 9	5,210	1	91.37	1	5,392	100.66
Step 10	5,205	1	80.13	1	5,392	72.39
Step 11	5,200	1	75.76	1	760	59.52
Step 12	5,195	1	54.50	1	692	44.55
Step 13	508	1	38.32	1	571	37.75
Step 14	508	1	37.22	1	817	39.62
Step 15	381	1	36.69	1	802	37.58

Table 5.3.3: Step price differences in the "old" pLOB (March 2019 corn futures)

	Bid Minimum Price Difference	Bid Maximum Price Difference	Ask Minimum Price Difference	Ask Maximum Price Difference
Step 1-2	0.0025	0.0350	0.0025	0.0200
Step 2-3	0.0025	0.0075	0.0025	0.0125
Step 3-4	0.0025	0.0050	0.0025	0.0050
Step 4-5	0.0025	0.0025	0.0025	0.0050
Step 5-6	0.0025	0.0025	0.0025	0.0050
Step 6-7	0.0025	0.0025	0.0025	0.0050
Step 7-8	0.0025	0.0025	0.0025	0.0025
Step 8-9	0.0025	0.0025	0.0025	0.0025
Step 9-10	0.0025	0.0025	0.0025	0.0025
Step 10-11	0.0025	0.0025	0.0025	0.0025
Step 11-12	0.0025	0.0025	0.0025	0.0025
Step 12-13	0.0025	0.0025	0.0025	0.0025
Step 13-14	0.0025	0.0025	0.0025	0.0025
Step 14-15	0.0025	0.0025	0.0025	0.0025

Table 5.3.4: Step quantities in the "old" pLOB (March 2019 corn futures)

	Max. Quantity Bid	Min. Quantity Bid	Average Quantity Bid	Max. Quantity Ask	Min. Quantity Ask	Average Quantity Ask
Step 1	32,929	1	981.51	72,386	1	1,011.66
Step 2	32,934	1	1,034.46	16,845	1	570.23
Step 3	32,795	1	763.11	16,939	1	411.22
Step 4	28,053	1	618.25	16,623	4	368.39
Step 5	28,053	16	770.00	17,154	35	526.75
Step 6	26,128	16	449.20	16,085	37	387.24
Step 7	26,128	58	456.13	10,219	40	243.72
Step 8	26,128	58	554.14	10,219	45	226.77
Step 9	25,882	58	483.01	10,219	56	251.71
Step 10	25,882	58	488.05	2,103	52	225.11
Step 11	6,024	61	335.90	10,219	52	218.24
Step 12	1,508	61	286.47	10,219	52	179.62
Step 13	1,415	61	274.12	10,219	52	202.46
Step 14	1,404	61	259.66	1,698	52	197.62
Step 15	1,425	61	246.42	1,699	51	179.39

A notable characteristic is that the complete LOB is "complete", in other words the first 15 steps each have a quantity however, in the short run, the "new" pLOB often consists of less than 15 steps (this is particularly noticeable for shorter ages).

The age used to split the complete LOB impacts the data available in each pLOB. As the age is shortened, the number of orders in the "new" pLOB declines which lead to a decline in the number of steps in the "new" pLOB. At the extreme, when an age of zero is used, the "new" pLOB is empty with zero steps and the "old" pLOB is equivalent to the complete LOB; similarly, when a maximum age of 5 days is used the "new" pLOB is equivalent to the complete LOB and the "old" pLOB is empty with zero steps. For this study, pLOBs were constructed using ages of 10 – 80 minutes on 10-minute intervals in addition to the analysis using the complete LOB.

Since the number of orders that make up the "new" pLOB declines as the age used to split the LOB declines, at small ages, the "new" pLOB may not include all 15 steps that were originally expected. Therefore, it may not be possible to differentiate between the aggregate prices between steps when splitting the LOB at young ages. For example, splitting the LOB using a 10-minute age interval with the old pLOB truncated to orders less than 1 day old, results in a pLOB that includes 10 steps, 96.5% of the snapshots (see Table 5.3.5). As a result, the aggregate price of steps 4 and 5 between snapshots is different for only 96.5% of the observations. Indeed, this will bias the results to indicate that the information content from "new" pLOB step 1-4 and "new" pLOB step 1-5 is more similar than it is, therefore; at short time intervals (less than 10 minutes), comparing the information from different steps may not be possible.

Table 5.3.5: Percentage of observations available during the week of November 11, 2018, for CH9 for the daytime session

Age split (minutes)	**10**	**20**	**30**	**40**	**50**	**60**	**70**
New pLOB Step 1	100	100	100	100	100	100	100
New pLOB Step 2	100	100	100	100	100	100	100
New pLOB Step 3	99.998	100	100	100	100	100	100
New pLOB Step 4	99.98	100	100	100	100	100	100
New pLOB Step 5	99.9	100	100	100	100	100	100
New pLOB Step 6	99.7	99.998	100	100	100	100	100
New pLOB Step 7	99.4	99.97	100	100	100	100	100
New pLOB Step 8	98.8	99.9	100	100	100	100	100
New pLOB Step 9	97.9	99.8	99.98	100	100	100	100
New pLOB Step 10	96.5	99.6	99.9	99.99	100	100	100
New pLOB Step 11	94.4	99.3	99.8	99.98	100	100	100
New pLOB Step 12	92.0	99.0	99.8	99.9	99.999	100	100
New pLOB Step 13	89.2	98.4	99.7	99.9	99.99	100	100
New pLOB Step 14	85.9	97.5	99.5	99.9	99.98	100	100
New pLOB Step 15	82.5	96.4	99.1	99.8	99.9	99.996	100

6. Model specification and information share measure

6.1 Aggregate LOB

Our goal is to calculate the information content of each pLOB where the pLOB is composed of orders on certain steps and meeting certain age criteria. This requires aggregating each pLOB into a single value. When limit orders are grouped by price and arranged sequentially the best price bid and ask group is assigned the "step 1" label and the second-best price on the bid and ask is assigned "step 2" label, and so on. Historically, studies have either focused on the best bid and ask (step 1) or some limited number of steps due to the data available in the MBP dataset. In previous studies by Cao, Hansch, and Wang (2009) and Arzandeh and Frank (2019), the quantity weighted average price was used to obtain a single value from a set of steps (with bid and ask sides included in the aggregation). We use the same aggregation method, which enables our results to be directly comparable to earlier studies. We compute the pLOB aggregate price for steps 1 to 15 as shown in equation (1).

$$\boxed{Aggregate\ price = \frac{\sum_{s=1}^{15} p_s q_s}{\sum_{s=1}^{15} q_s}} \tag{1}$$

where p_s refer to bid and ask prices at step s and q_s are their corresponding quantities.

6.2 Model variables

To estimate the information shares that results from orders of various ages we split the LOB into "new" orders and "old" orders. Additionally, our analysis includes two types of pLOB datasets, one where all orders are less than a certain age, and another type with two pLOBs datasets where one includes younger orders and the other includes older orders. One benefit of estimating a model using new orders only is to be able to study the impact of changing the age interval on the information share in a simple two-variable model (last price and aggregate pLOB).

We estimated six models using pLOBs constructed in different ways. Since we do not know when an order can be considered old, we test new/old order designation using 10 minutes, 20 minutes, …, and so on until 80 minutes as well as testing the complete LOB where age is not a factor in determining which orders to include. Some of the tests include up to 15 steps in the LOB, while others only include 10 steps. We also estimate the three- and four-variable models in Arzandeh and Frank (2019) as a base level confirmation of the ability to extract data from the MBO dataset, estimate the Modified Information Share and study the effect of age in the limit orders. Table 6.2.1 lists the models that are estimated in this study. These models will be referred to as their Model Name for the duration of the paper.

Table 6.2.1: Variables included in each estimated model.

Model Name	Variables
Model A	1) Last Trade Price 2) New $\frac{\sum_{s=1}^{10} p_s q_s}{\sum_{s=1}^{10} q_s}$
Model B	1) Last Trade Price 2) New $\frac{\sum_{s=1}^{15} p_s q_s}{\sum_{s=1}^{15} q_s}$
Model C	1) Last Trade Price 2) New $\frac{\sum_{s=1}^{10} p_s q_s}{\sum_{s=1}^{10} q_s}$ 3) Old $\frac{\sum_{s=1}^{10} p_s q_s}{\sum_{s=1}^{10} q_s}$
Model D	1) Last Trade Price 2) New $\frac{\sum_{s=1}^{15} p_s q_s}{\sum_{s=1}^{15} q_s}$ 3) Old $\frac{\sum_{s=1}^{15} p_s q_s}{\sum_{s=1}^{15} q_s}$
Model E	1) Last Trade Price 2) New $\frac{p_1 q_1}{q_1}$ 3) New $\frac{\sum_{s=2}^{10} p_s q_s}{\sum_{s=2}^{10} q_s}$
Model F	1) Last Trade Price 2) New $\frac{p_1 q_1}{q_1}$ 3) New $\frac{\sum_{s=2}^{3} p_s q_s}{\sum_{s=2}^{3} q_s}$ 4) New $\frac{\sum_{s=4}^{10} p_s q_s}{\sum_{s=4}^{10} q_s}$

In the market microstructure context, price discovery refers to the "discovery" of the true, but unobserved, price of a commodity. We use the observed series of prices of the commodity and estimate the contribution of each series to the true price. We include market orders by including the last trade price, and we include limit orders by including various aggregations of the LOB. One underlying assumption is that all series for a particular commodity (true unobserved and observed) are cointegrated. That is, in the short run the series may deviate from each other due to temporary shocks of information and strategies that traders may implement in response to those shocks, but in the long run, all prices pertaining to the commodity will move in the same direction because they are influenced by the same underlying process. Therefore, to allow for the cointegrating relationship between the observed prices we estimate a Vector Error Correction Model (VECM). To study the information contained in orders of different ages we use one or more variables for the observed prices in the pLOB, which can be composed of various ages of order and various combinations of steps.

6.3 Order age groupings

We hypothesize that the LOB contains the public and private information from traders. Lehecka, Wang, and Garcia (2014) showed that public information is consumed by the market in approximately 10 minutes. Intuitively we expect private information to be more persistent since individual traders may use methods to accumulate positions for periods much longer than 10 minutes. Therefore, we expect the private information to persist for more than 10 minutes, and we perform the analysis for age splits ranging from 10 minutes to 80 minutes on 10-minute intervals. For completeness, we also test a scenario that does not exclude order based on age, similar to that done by Arzandeh and Frank (2019) and Cao, Hansch, and Wang (2009). This set of pLOBs allows us to study how the information share of the last price changes with various combinations of pLOBs variables, discussed in Table 6.2.1.

A property of the dataset is that as the age interval of the pLOB expands, the total quantity in the pLOB increases. Table 6.3.1 shows the increase in pLOB quantity in corn and soybean for Model A (first 10 steps) and Model B (first 15 steps).

Table 6.3.1: Relationship between quantity and age

maximum age of orders (minutes)	Model A		Model B	
	Corn	Soybean	Corn	Soybean
10	2,298	523	2,601	656
20	3,105	669	3,596	855
30	3,683	755	4,341	982
40	4,159	813	4,969	1073
50	4,558	854	5,519	1140
60	4,909	888	6,005	1192
70	5,204	915	6,429	1236
80	5,457	936	6,802	1273
All orders	8,434	1116	12,169	1650

6.4 Lag order

An important component of the model specification is the lag length to incorporate in the model. Arzandeh and Frank (2019) use an 80-lag model which covers more than 10 minutes in their analysis. Cao, Hansch, and Wang (2009) use the AIC to determine a lag length of 5 after choosing a 5-minute snapshot interval, resulting in a model that incorporates a total of 25 minutes of data. For this study, lag length is explored using a set of vector autoregressions with lag order 1 through 80 and recording the lag order selection statistic of the Akaike's Information Criterion (AIC), and Likelihood Ratio tests (LR). Generally, lag order selection is an attempt to choose the fewer number of lags that eliminate autocorrelation in the residuals of the model. The AIC selects the model with the smallest sum squared regression (SSR) relative to the one-step ahead forecast. The (sequential) LR tests successive models (starting with the model with the most lags and declining lag order in each successive test) with a null hypothesis that all of the lag coefficients are zero and selects the first model that rejects the null hypothesis. This testing procedure is recommended by Lütkepohl (p.143, 2005). In general, the lowest number of lags is more desirable, however, the model structure is also guided by economic theory, which in our case there is a strong suggestion in previous studies that the model should encompass more than 10 minutes of microstructure data, and we should see the declining influence of the more distant lags. Although the lag order selection is potentially unique for each time series (there is a different time series for each day and each pLOB aggregate), several generalizations can be made. Generally, the LR tests resulted in much longer lag order selection, while the AIC resulted in much shorter lag orders. This was consistent between corn and soybean (see Appendix 9.5)

Table 6.4.1: Lag order selection

Lag Order Selection	AIC	LR
Mean Lag Order	8.1	72.6
Median Lag Order	6	76
St. dev. Lag Order	8.51	9.22

The expectation is to find that the LOB contains information that contributes to price discovery, similar to previous studies by Cao, Hansch, and Wang (2009) and Arzandeh and Frank (2019), therefore we expect our model to cover more than 10 minutes, and for consistency throughout the analysis we ran all models and all time series with the same number of lags. Preliminary testing with lags of lengths suggested by AIC (generally 10 or less lags) showed much higher occurrence of autocorrelation in the VEC residuals. Testing with 80 lags showed less autocorrelation in the residuals as well as having a consistent approach to each model and time series. As a result, we choose to run the analysis with 80 lags.

6.5 Integration Order

A set of cointegrated time series are individually integrated of the same order, but a linear combination will exhibit stationarity. Therefore, before estimating the model, stationarity tests are performed to check that the time series are integrated of order one. We used the Phillips-Peron (PP) (Phillips and Perron, 1988) Unit Root Test where the H_0 is that the time series contains a unit root, and the H_A is that the time series is stationary. We utilized the process developed in Dickey and Pantula (2002) and tested for stationarity of the second differences, the first differences, and the undifferenced time series. In rare cases, we did find second differencing required for individual time series but did not find any instances where all the time series in a single model run were integrated of order 2, or I(2). The results of these tests generally confirmed the presence of a unit root in the undifferenced time series, that was removed with first differences, allowing the conclusion that the data was I(1); however, there were instances where all of the time series in a model run were not I(1) and therefore called into question the cointegration of these time series. This was particularly evident in Model F where the test showed less than half the days showed cointegration. Detailed results for the stationarity tests are included in Appendix 9.3. Figure 6.5.1 and Figure 6.5.2 summarize the success rate when analysing each time series variable in a model for the same integration order.

We found that models with more variables and smaller age splits (less data) have greater propensity to fail tests for I(1), therefore calling into question the results of these models. Model A and B were nearly always integrated at order 1, while Model F was I(1) less than 50% of the datasets in all age groupings and less than 25% of the datasets if the pLOB were constructed with orders less than 30 minutes.

Figure 6.5.1: Percentage of days in which all variables in each model are I(1) for corn.

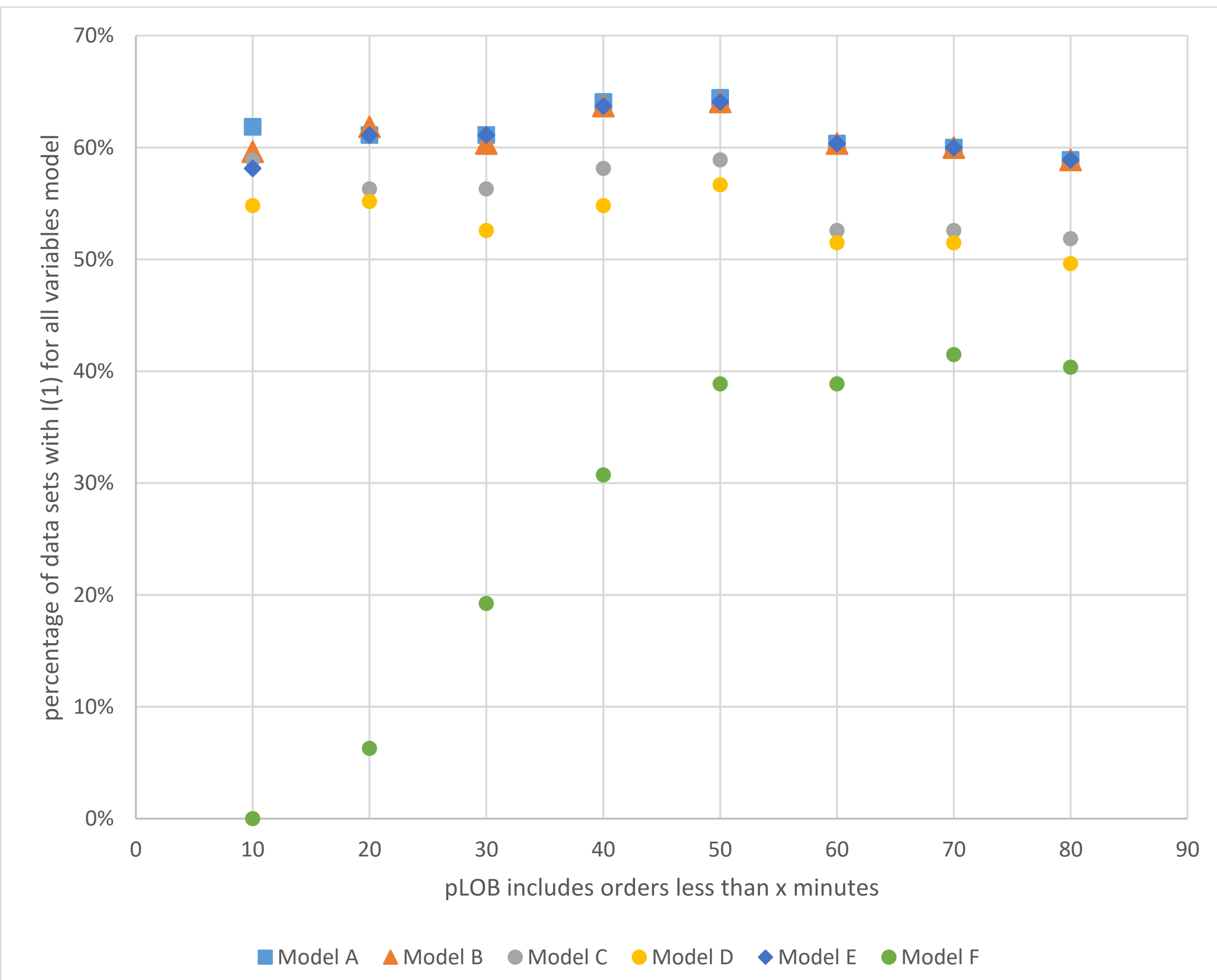
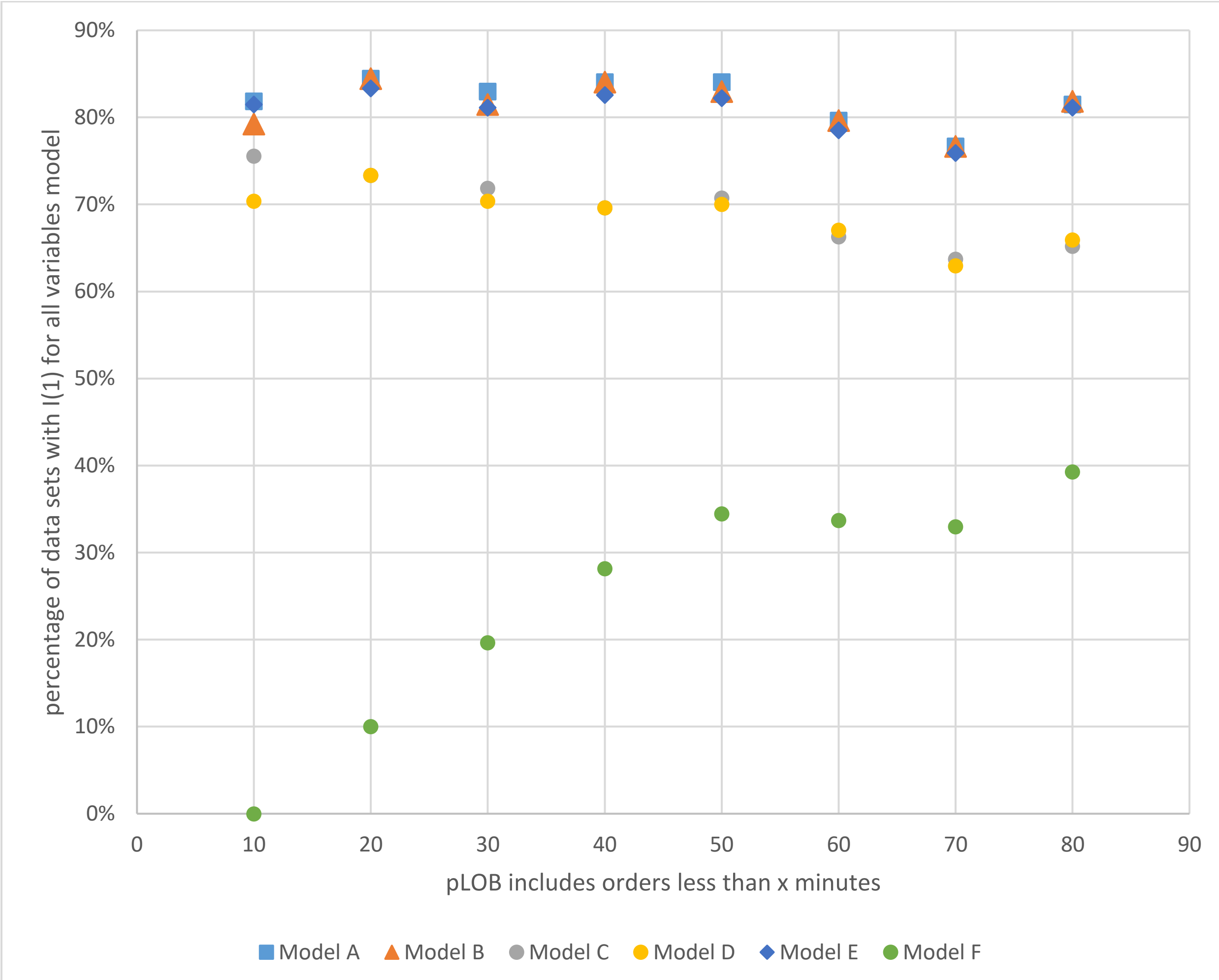


Figure 6.5.2: Percentage of days in which all variables in each model are I(1) for soybean



6.6 Cointegration Rank

We test for the presence of cointegrating relationships, using the Maximum Likelihood estimator of cointegrating equations developed by Johansen (1995). These tests show variable results, including the lack of cointegration in a few cases, however, generally, the tests indicated (K-1) cointegrating relationships.

We minimized the Hannan and Quinn Information Criterion to select the number of cointegrating equations. We expect Models A & B to have a rank of 1; Models C, D, and E to have a rank of 2 and Model F to have a rank of 3. Generally, the results are as expected, however, Models C, D, and E did have a significant number of occurrences where there were less than expected cointegrating equations. See Appendix 9.4 for additional detail.

Table 6.6.1 Estimated rank

	Corn				Soybean			
	Rank 0	Rank 1	Rank 2	Rank 3	Rank 0	Rank 1	Rank 2	Rank 3
Model A	0%	100%			0.3%	99.7%		
Model B	0%	100%			0.3%	99.7%		
Model C	0%	14.6%	85.4%		0.3%	18.3%	81.4%	
Model D	0%	14.7%	85.3%		0.3%	19.8%	79.9%	
Model E	0%	0.1%	99.9%		0.0%	0.3%	99.7%	
Model F	0%	0.0%	13.7%	86.3%	0.0%	0.0%	9.4%	90.6%

We note that Models C and D (which have separate variables for new and old pLOB) and Model F (which is the model with the most variables) have a significant occurrence of the rank being less than K-1. Our analysis used K-1 in all cases.

6.7 Vector Error Correction Model (VECM)

For an I(1) process and a vector $\boldsymbol{y}_t$ of endogenous variables of dimension $K\times 1$ the Vector Error Correction representation is,

$$\boxed{\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \ \dots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t} \quad \textbf{(2)}$$

where: $\boldsymbol{y}_t$ is a vector of endogenous variables containing the last transaction price (P_t) and the aggregate prices made up of various components of the pLOB as specified in Table 6.2.1 (this study estimates VECM models with 2, 3, and 4 variables), $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}_i$ are $K\times K$ matrices of coefficients, with $\boldsymbol{\Pi} = \boldsymbol{\alpha\beta'}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being matrices of rank r, $0 < r < K$, $I = 1, \dots, p\text{-}1$ is the number of lags**,** and $\boldsymbol{u}_t \sim (0, \boldsymbol{\Omega})$ is white noise.

An important consideration when using the VECM is to be able to assume that all the endogenous variables are jointly determined. In the case of our tests, we have two main areas of concern. The first is that each of the variables that makeup $\boldsymbol{y}_t$ are historical to varying degrees at each snapshot (that is each of the variables in $\boldsymbol{y}_t$ has a different time between their last update and the snapshot). The second is that when including orders older than an age interval as a pLOB there is some historical component to this variable. However, it should be noted that each of the variables in $\boldsymbol{y}_t$ is influenced by two things. The first is the impact from the trader adding, editing, or cancelling the orders. The second is the market trading orders which impact the quantities on the top steps and may impact the prices on all steps as well as the transaction prices. These two actions are the only thing that can influence the time series in the variables, and neither of these results in the predetermination of one of the variables. On a nanosecond scale, it is reasonable to assume that the variables that makeup $\boldsymbol{y}_t$ are not simultaneously updated, however, they are simultaneously determined in that contemporaneous action is needed to update the data that is grouped into the variables.

6.8 Testing Residuals of the VECM

Residuals of a correctly specified VECM are assumed to be independent, therefore we test the residuals for the presence of autocorrelation using the Lagrange Multiplier test for autocorrelation as described in Johansen (1995), which has a null hypothesis of no autocorrelation in the residuals. This test has a rejection rate that varies between 8% and 13% depending on the model specification. Reject of the null hypothesis is a good indication that the model is mis-specified. Some evidence of autocorrelation was found in the residuals. Instances where the autocorrelation exceeded four occurrences in the first 10 lags were excluded from the information share estimation.

Table 6.8.1: Lagrange Multiplier test for correlation in the VECM residuals

	Rejection rate in the first 10 lags of the residuals	
Model	Corn	Soybean
Model A	10.41%	10.23%
Model B	8.86%	10.14%
Model C	13.13%	13.46%
Model D	11.42%	12.56%
Model E	9.68%	9.93%
Model F	10.28%	8.60%

H_o: No autocorrelation in the residuals (after excluding occurrences with more than 4 rejections)

6.9 Modified Information Share (MIS)

Using the Granger representation theorem we can decompose $\boldsymbol{y}_t$ into I(1) and I(0) components (Lütkepohl, 2005 page 252) as follows,

$$\boxed{\mathbf{y}_t = \mathbf{y}_0^* + \mathbf{\Psi}\,\Sigma_{i=1}^{t}\,\mathbf{u}_i + \mathbf{\Psi}^*(L)\mathbf{u}_i} \tag{3}$$

where $\mathbf{y}_0^*$ are initial values, $\mathbf{\Psi}^*(L)\mathbf{u}_i$ is an I(0) process, $\mathbf{\Psi}\,\Sigma_{i=1}^{t}\mathbf{u}_i$ is an I(1) process, $\mathbf{\Psi} = \boldsymbol{\beta}_\perp\left[\boldsymbol{\alpha}'_\perp\left(\mathbf{I}_K - \Sigma_{i=1}^{p-1}\mathbf{\Gamma}_i\right)\boldsymbol{\beta}_\perp\right]^{-1}\boldsymbol{\alpha}'_\perp$ is the long-run impact matrix, and $\boldsymbol{\beta}_\perp$is an orthogonal complement of $\boldsymbol{\beta}$ (similarly for $\boldsymbol{\alpha}'_\perp$). The rows of $\mathbf{\Psi}$ are assumed to be identical because the different prices being considered correspond to the same commodity and therefore, in the long run, must be equal. That is, for $\boldsymbol{\psi}$ being an identical row in $\mathbf{\Psi}$, the long-run impact of innovations on each price is given by $\boldsymbol{\psi}\mathbf{u_t}$ and its variance is $\boldsymbol{\psi}\mathbf{\Omega}\boldsymbol{\psi}'$ (Hasbrouck 1995).

When $\mathbf{\Omega}$ is not diagonal, Lien and Shrestha (2009) propose the following factor structure for the innovations based on the innovation correlation matrix $\mathbf{\Phi}$:

$$\boxed{\boldsymbol{u}_t = \boldsymbol{F}\boldsymbol{u}_t^*} \tag{4}$$

where $\boldsymbol{\Omega} = \boldsymbol{F}\boldsymbol{F}'$, $\boldsymbol{F}^* = \left[\boldsymbol{G}\boldsymbol{\Lambda}^{-1/2}\,\boldsymbol{G}'\boldsymbol{V}^{-1}\right]^{-1}$, $\boldsymbol{G}$ is the matrix containing the eigenvectors of $\boldsymbol{\Phi}$ in its columns, $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements being the eigenvalues of $\boldsymbol{\Phi}$, $\mathbf{V}$ is a diagonal matrix with diagonal elements being the innovations standard deviations ($\sqrt{\boldsymbol{\Omega}_{ii}}$), and $\boldsymbol{u}_t^*$ is the transformed innovation with $E[\boldsymbol{u}_t^*] = 0$ and $E\left[\boldsymbol{u}_t^*\boldsymbol{u}_t^{*'}\right] = I$.

Using the long-run impact of innovations defined above ($\boldsymbol{\psi}\mathbf{u}_\mathbf{t}$) and Lien and Shrestha's (2009) factorization structure, their modified information share of price j is,

$$\boxed{MIS_j = \frac{\boldsymbol{\psi}\boldsymbol{F}_j^2}{\boldsymbol{\psi}\boldsymbol{\Omega}\boldsymbol{\psi}'}} \tag{5}$$

7. Results

Model A is comprised of two variables, the last price, and a single aggregate price for the LOB that is truncated at 10 steps. Additionally, the aggregate price for the LOB was composed only of orders that satisfied the age requirement. The average MIS for the pLOB in Model A and various order age limits, as well as for orders of all ages, are reported in Table 7.1. The pLOB MIS differs across age limits widely, ranging from about 19% (32%) for lower limits to 50% (48%) for larger limits in corn (soybean). When the age limit is set low, the pLOB is mostly comprised of young orders, in which case the pLOB appears to be less informative. In corn (soybean), the pLOB is more informative than the last transaction price when orders older than 70 (80) minutes are included. This may be due to two effects: i) Lower age limits capture a higher proportion of short-lived orders that do not necessarily carry information because they are either coming from impatient or liquidity traders who do not make trading decisions based on information but rather on order flow needs (Pascual and Veredas 2009), or they are being sent with the objective of "fishing" and "spoofing" and canceled in a short period of time (Cao, Hansch and Wang 2009); ii) The higher the age limit imposed in the LOB the higher the quantity of orders resting in the LOB, a variable that has been shown to be related to the information share (Hu et al. 2020). The results obtained for corn and soybean are, in general, consistent. However, for orders that are less than 60 minutes old the pLOB for soybean appears to be more informative than the pLOB for corn. This is consistent with the higher quantities observed in the soybean market relative to the corn market (Table 6.3.1). In corn, the MIS of the LOB seems to flatten after 70 minutes, whereas for soybean the MIS slightly increases up to the maximum age limit of 80 minutes.

Model B is comprised of the same two variables as Model A, except that the pLOB is truncated at 15 steps. The MIS obtained for model B are in line with those obtained for Model A, that is, the pLOB constructed with orders restricted to lower age limits contain less information than the pLOB with additional older orders. In corn, the additional steps seem to be informative for all age limits, as the MIS for the pLOB is about 2% to 4% higher than its 10-step counterpart. In contrast, for soybean, the additional steps used in Model B seem to be noisier as the MIS for the pLOB is lower for 15 steps than for 10 steps.

Table 7.1: Modified Information Shares for Model A

	Corn		Soybean	
Orders age (min)	Last Price	Step 1-10	Last Price	Step 1-10
10	80.72	19.28	67.80	32.20
20	74.14	25.86	62.70	37.30
30	66.19	33.81	56.52	43.48
40	60.21	39.79	55.35	44.65
50	55.41	44.59	53.99	46.01
60	52.16	47.84	53.37	46.63
70	50.32	49.68	52.42	47.58
80	49.81	50.19	52.12	47.88
All *	50.12	49.88	48.61	51.39

*All orders regardless of age

Table 7.2: Modified Information Shares for Model B

	Corn		Soybean	
Orders age (min)	Last Price	Step 1-15	Last Price	Step 1-15
10	78.94	21.06	72.74	27.26
20	70.01	29.99	66.27	33.73
30	61.77	38.23	61.70	38.30
40	55.45	44.55	57.09	42.91
50	51.55	48.45	55.94	44.06
60	49.52	50.48	54.53	45.47
70	47.63	52.37	53.29	46.71
80	47.73	52.27	53.22	46.78
All *	54.83	45.17	48.75	51.25

*All orders regardless of age

Model C is comprised of three variables; the last transaction price, the aggregate price of the "new" pLOB, and the aggregate price of the "old" pLOB, with the LOB truncated to the first 10 steps. The average MIS for Model C for the various age splits used are shown in Table 7.3. For corn and soybean, the information content of the transaction price is lower, and that of the LOB (considering both new and old pLOBs) is higher, relative to models A and B. This is probably due to the higher quantity of orders that are included in the LOB in Model C. The results for Model C also show that the "new" pLOB contains more information than the "old" pLOB, except for the "younger" category of orders that are less than 10 minutes old in corn. As explained for models A and B, this could be the result of a higher proportion of noisier orders associated with impatient or liquidity traders, or "fake" orders used for fishing and spoofing, plus being the grouping associated with a low quantity. For soybean, the MIS of the new pLOB is always higher than the MIS of the old LOB, even for the younger category associated with the lower quantity. The MIS of the new pLOB is about 9% and 20% higher than that of the old pLOB for corn and soybean, respectively, suggesting that orders entering the LOB in the soybean market are more informative relative to the corn market. Overall, the results show that accounting for the age of the orders is important to assess the contribution of the different prices in the price discovery process.

Model D is similar to model C with the LOB expanded to steps 1-15. These results reinforced that "new" orders carry more information than "old" orders, as well as providing more robustness to the previous findings. Similar to Model B, the additional steps seem to be informative for all age limits, while this is less evident in soybean.

Table 7.3: Modified Information Shares for Model C

	Corn			Soybean		
pLOB based on age (min)	Last Price	New pLOB Step 1-10	Old pLOB Step 1-10	Last Price	New pLOB Step 1-10	Old pLOB Step 1-10
10	65.97	15.99	18.04	53.16	27.13	19.71
20	63.63	24.70	11.68	55.77	33.78	10.45
30	55.93	27.58	16.49	51.93	35.53	12.54
40	48.75	30.36	20.89	47.79	37.06	15.15
50	44.14	32.58	23.27	45.63	38.46	15.91
60	41.46	32.79	25.75	44.30	37.46	18.24
70	39.27	34.02	26.71	42.11	38.16	19.73
80	38.41	33.75	27.84	39.38	37.04	23.58

Table 7.4: Modified Information Shares for Model D

	Corn			Soybean		
pLOB based on age (min)	Last Price	New pLOB Step 1-15	Old pLOB Step 1-15	Last Price	New pLOB Step 1-15	Old pLOB Step 1-15
10	63.71	17.09	19.20	59.78	21.76	18.46
20	61.92	25.88	12.19	57.80	30.96	11.24
30	52.38	30.93	16.69	51.90	35.18	12.92
40	45.68	33.47	20.85	49.73	36.67	13.59
50	40.34	36.05	23.60	47.65	36.99	15.35
60	36.93	36.80	26.27	44.93	37.70	17.37
70	33.77	37.04	29.19	43.50	35.21	21.29
80	34.51	34.90	30.59	41.33	34.38	24.29

Model E was estimated to validate the new MBO data preparation and methods, using the same variables defined by Arzandeh and Frank (2019). Model E was also estimated restricting the age of the orders in the LOB. The MIS obtained when all orders are included (no age restriction) are remarkably similar to the results obtained by Arzandeh and Frank (2019). Interestingly, here too the information content of the last price declines when orders of increasing age are added. Similarly, the MIS of the step 1 orders declined when adding older order. The results for model F are also similar as those obtained by Arzandeh and Frank (2019) and consistent with the findings in Model E.

Table 7.5: Modified Information Shares for Model E

	Corn			Soybean		
Orders age (min)	Last Price	Step 1	Step 2-10	Last Price	Step 1	Step 2-10
10	45.66	42.43	11.91	37.41	39.35	23.24
20	41.41	44.02	14.57	35.42	38.09	26.49
30	36.81	41.80	21.39	31.42	36.10	32.48
40	34.49	38.22	27.29	30.49	35.20	34.31
50	32.14	34.79	33.07	28.70	35.23	36.07
60	28.89	34.21	36.90	27.61	25.35	37.04
70	26.29	33.71	40.00	26.51	35.38	38.12
80	24.78	34.04	41.18	26.79	34.95	38.27
All *	24.03	35.58	40.39	26.35	35.12	38.53
**	24.91	32.60	42.49	25.48	33.86	40.66

*All orders regardless of age

** Results from Arzandeh and Frank (2019)

Table 7.6: Average Modified Information Shares for Model F

	Corn				Soybean			
Orders age (min)	Last Price	Step 1	Step 2-3	Step 4-10	Last Price	Step 1	Step 2-3	Step 4-10
< 10	36.96	36.08	10.67	16.29	30.91	32.16	23.56	13.37
< 20	30.98	33.46	18.31	17.26	27.40	28.96	28.67	14.96
< 30	25.02	29.06	25.57	20.35	24.3	28.17	29.84	17.69
< 40	21.76	27.23	28.68	22.34	22.74	28.35	29.95	18.97
< 50	20.06	26.45	30.34	23.14	22.28	24.41	30.17	19.13
< 60	18.88	26.17	29.26	25.68	21.66	28.67	29.67	20.00
< 70	18.89	25.22	28.76	27.13	21.69	29.42	31.22	17.67
< 80	18.59	26.18	30.20	25.03	21.68	28.91	30.89	18.52
All *	17.72	28.92	34.70	18.66	21.44	30.22	32.84	15.51
**	18.17	25.09	29.61	27.13	18.23	27.28	26.28	28.21

*All orders regardless of age
** Results from Arzandeh and Frank (2019)

In all models, the MIS of the last price generally declines as the age of the orders increases. In some cases, this occurs throughout the tests up to 80 minutes, which confirms that changes to the older orders convey information for a period of approximately 70-80 minutes.

In summary, in models A and B, we found that excluding orders on the basis of age, generally reduces the information share of the LOB. This occurred in the estimations using 10 steps and 15 steps. The benefit of having less restrictive age criteria to exclude orders declines to a minimum at 70 – 80 minutes of age. In models C and D, the new and old orders both appeared in the system and confirmed that changes to the "new" pLOB were more informative than changes to the "old" pLOB in all age trails except the 10-minute age split done in Model D. In models E and F, we show the results when using all orders (regardless of age) to be comparable to the Arzandeh and Frank (2019) analysis. Appendix 9.7 contains the detailed tables and figures of the results.

8. Conclusions

We parsed the new MBO data file from the CME Group for corn and soybean futures for the period December 02, 2018, to December 27, 2019, identified the individual orders, and created two age groups of limit orders. We studied the relationship between the information content of the orders and the age of the orders using the Modified Information Share that was developed by Lien and Shrestha (2009). Our analysis focused on the last transaction price, and aggregations of the LOB.

Our results suggest that the information embedded in limit orders is relatively persistent, often exceeding 70 minutes. The models that included the last price as well as both the new and old limit orders (models B and C), expanding the age range used to build the "new" pLOB resulted in additional information share for the "new" pLOB. And, in models that included the last price and various aggregates of the "new" pLOB steps, the information share embedded in the last price declined (conversely, the information share in the LOB increased) when expanding the age criteria for the "new" pLOB to 70 and 80 minutes, as well as the result from Model F which found that the information share of the last price was minimized by adding all order to the LOB regardless of age.

We find that new orders carry more information than old orders, regardless of the age criteria used to create the two age groups. However, we also note that old orders maintain a significant share of information. Interestingly, the "old" pLOB carries additional information share if older age criteria is used to split the orders into two groups. This could be due to the older age criteria isolating the oldest orders, which presumably contain information that has not "expired." Old orders that are still active and have not been cancelled could be an indication of "good" information. This could also be a result of subgroups of orders having information that is out of phase, and therefore nearly cancels each other out, resulting in higher information share for the last transaction price. Grouping the orders that comprise each variable so that the individual groups contain orders that complement each other may be an important consideration in future research.

9. Appendix

9.1 Parsing

Packets contain many different pieces of data that are marked with a “tag”. The following tags were encountered in the packets with order and trade data.

Useful Tags from CME FIX files

TAG	Used?	Description
9	Ignored	Message Length
10	Ignored	Message CheckSum (always end of message)
34	Ignored	Message Sequence Number
35	Used	Message Type
37	Used	Order ID for MBO entry
48	Ignored	Security ID (redundant with Symbol)
49	Ignored	Sender ID (always “CME”)
55	Used	Symbol
52	Ignored	Sending Time
60	Used	Transaction Processing Time
75	Used	Trade Date
83	Ignored	Reporting sequence
268	Used	Number of MD entries
269	Used	MBP/MBO Entry Type
270	Used	Price of the MBP/MBO entry
271	Ignored	Quantity of the MBP entry
279	Ignored	MBP update Action
346	Ignored	Number of Order in MBP entry
1023	Ignored	LOB Step level in MBP entry
1128	Ignored	Application Version ID
5799	Ignored	Event Indicator
9633	Used	Reference ID to corresponding tag268 sequence
37705	Used	Number of MBO entries
37707	Ignored	Order priority for MBO entry
37706	Used	Display Qty for MBO entry
37708	Used	MBO Update Action

The FIX packets are made up of two or three parts, depending on whether the packet contains MBP and MBO data, or just MBO data. The first part of the packet identifies the type of information contained in the packet as well as the trade date, and the sending time. Part two and part three are repeating sets of data with specifics for the symbol, price, quantity, update type, and entry type. We identified three interesting patterns for parsing, one trade pattern and two order patterns. These three patterns are shown in the following tables.

Example of trade update packet

Part 1	1128=9, 9=510, 35=X, 49=CME, 75=20181112, 34=208120, 52=20181112053509032992649, 60=20181112053509030287355, 5799=00000001
	268=3
Part 2	279=0, 269=2, 48=275617, 55=ZCH9, 83=17976, 270=380.25, 271=15, **346=5**, 5797=2, 37711=213082
	279=0, 269=2, 48=660347, 55=ZCU9, 83=15440, 270=397.25, 271=3, **346=1**, 5797=0, 37711=213084
	279=0, 269=2, 48=102262, 55=ZCZ9, 83=15774, 270=402.0, 271=1, **346=1**, 5797=0, 37711=213086
	37705=7
Part 3	37=702973770070, 32=15
	37=702973769516, 32=1
	37=702973769520, 32=1
	37=702973769525, 32=1
	37=702973769045, 32=8
	37=702973769563, 32=3
	37=702973769589, 32=1
End of Message	10=168

*Note that tag 268 identifies that there will be 3 entries in part 2, and tag 346 identifies how many entries in part 3 are related to each entry in part 2. Tag 37705 identifies the total number of entries in part 3.

Resulting order updates

Date Tag75	Time Tag60	Symbol Tag55	Update Tag279	Entry Tag269	Price Tag270	Qty Tag32	Order ID Tag37
20181112	20181112 053509 030287355	ZCH9	new	trade	380.25	15	702973770070
		ZCH9	new	trade	380.25	1	702973769516
		ZCH9	new	trade	380.25	1	702973769520
		ZCH9	new	trade	380.25	1	702973769525
		ZCH9	new	trade	380.25	8	702973769045
		ZCU9	new	trade	397.25	3	702973769563
		ZCZ9	new	trade	402.00	1	702973769589

Example of order update with MBO data

Part 1	1128=9, 9=350, 35=X, 49=CME, 75=20181112, 34=18437, 52=20181112010003222606959, 60=20181112010003221889459, 5799=00000100
	268=3
Part 2	279=2, 269=1, 48=275617, 55=ZCH9, 270=380.0, 37706=1, 37707=7099235259, 37=702973703373
	279=1, 269=1, 48=275617, 55=ZCH9, 270=380.0, 37706=3, 37707=7099235425, 37=702973703481
	279=1, 269=1, 48=275617, 55=ZCH9, 270=380.0, 37706=2, 37707=7099236082, 37=702973703962
End of Message	10=097

*Note that tag 268 identifies that there will be 3 entries in part 2.

Resulting order updates

Date Tag75	Time Tag60	Symbol Tag55	Update Tag279	Entry Tag269	Price Tag270	Qty Tag32	Order ID Tag37
20181112	20181112 010003 221889459	ZCH9	Delete	Ask	380.00	1	702973703373
		ZCH9	Update	Ask	380.00	3	702973703481
		ZCH9	update	Ask	380.00	2	702973703962

Example of order update with a combination of MBP and MBO data

Part 1	1128=9, 9=432, 35=X, 49=CME, 75=20181112, 34=17359, 52=20181112010002650012080, 60=20181112010002649061679, 5799=00000100
	268=3
Part 2	279=1, 269=0, 48=173904, 55=ZCZ8, 83=959, 270=368.5, 271=43, 346=16, 1023=1
	279=2, 269=1, 48=173904, 55=ZCZ8, 83=960, 270=368.75, 271=1, 346=1, 1023=1
	279=0, 269=1, 48=173904, 55=ZCZ8, 83=961, 270=371.25, 271=96, 346=11, 1023=10
	37705=2
Part 3	37=702973703437, 37707=7099235355, 37706=1, **9633=1**, 37708=2
	37=702973703478, 37707=7099235411, 37706=1, **9633=2**, 37708=2
End of Message	10=198

*Note that tag 268 identifies that there are 3 entries in part 2 and tag 37705 identifies that there are 2 entries in part 3. Tag 9633 identifies which part 2 entry is related to each part 3 entry.

Resulting Order Data

Date Tag75	Time Tag60	Symbol Tag55	Update Tag37708	Entry Tag269	Price Tag270	Qty Tag32	Order ID Tag37
20181112	20181112 010002 64906167	ZCZ9	Delete	Bid	368.50	1	702973703437
		ZCZ9	Delete	Ask	368.75	1	702973703478

9.2 Observations: Orders each week and Lifespan of Orders

Week of	Mean Order Lifespan (minutes)	Median Order Lifespan (seconds)	Count of Orders During Week
20181202	218.8	7.3	346,559
20181209	241.7	6.2	355,080
20181216	290.0	7.1	318,077
20181223	444.0	5.1	195,938
20181230	362.4	7.0	241,131
20190106	313.5	9.8	319,747
20190113	301.1	21.5	343,846
20190120	418.9	13.2	249,190
20190127	341.6	14.5	319,898
20190203	365.9	14.3	307,165
20190210	354.9	10.0	320,119
20190217	383.5	8.3	264,232
20190224	122.7	6.4	443,803
20190303	188.9	10.3	364,351
20190310	181.4	11.0	392,366
20190317	218.6	9.6	353,504
20190324	221.5	5.0	364,982
20190331	240.1	9.3	306,296
20190407	282.7	9.6	263,072
20190414	365.4	5.0	196,583
20190421	166.6	5.7	357,826
20190428	182.7	3.6	420,743
20190505	169.4	3.0	500,223
20190512	122.1	2.6	667,270
20190519	80.2	2.1	726,820
20190526	39.1	2.3	742,875
20190602	41.9	2.6	744,906
20190609	44.5	2.4	605,191
20190616	33.3	3.2	538,288
20190623	90.0	1.9	436,690
20190630	123.7	3.8	312,078
20190707	103.3	3.6	433,985
20190714	105.2	4.8	462,214
20190721	147.4	5.1	376,289
20190728	131.5	6.5	390,195
20190804	130.1	6.9	396,438

20190811	119.7	7.5	456,113
20190818	168.6	9.5	334,385
20190825	169.1	11.2	347,605
20190901	210.8	14.0	271,233
20190908	165.1	12.6	348,758
20190915	205.4	15.3	296,995
20190922	210.9	8.4	302,944
20190929	144.1	5.8	427,968
20191006	162.9	9.3	374,692
20191013	179.7	9.4	338,814
20191020	216.0	9.0	301,414
20191027	196.6	8.3	345,031
20191103	161.4	6.9	409,954
20191110	227.8	10.7	285,623
20191117	214.5	9.7	279,613
20191124	124.7	5.1	260,953
20191201	112.1	8.7	393,487
20191208	127.5	9.2	382,406
20191215	182.1	11.4	320,982
20191222	365.2	10.4	169,898

9.3 Stationarity Test Results

Stationarity Test Results using Phillips and Perron (1988) to test that all variables in the model are I(1). This table shows detailed results from the corn models using pLOB that are less than 40 minutes old. Similar datasets were examined for soybean and the other age intervals. Models with variables that all require first differencing to become stationary are marked with a 1, while models marked with a 0 did not satisfy this requirement.

	pLOB constructed from orders less than 40 minutes old					
	Model A	Model B	Model C	Model D	Model E	Model F
20181203	0	0	0	0	0	0
20181204	1	1	1	1	1	1
20181205	1	1	1	1	1	1
20181206	0	0	0	0	0	0
20181207	0	0	0	0	0	0
20181210	1	1	1	1	1	1
20181211	0	0	0	0	0	0
20181212	1	1	1	1	1	0
20181213	1	1	1	1	1	1
20181214	1	1	1	1	1	1
20181217	0	0	0	0	0	0
20181218	0	0	0	0	0	0
20181219	1	1	1	0	1	1
20181220	1	1	0	0	1	1
20181221	1	1	1	1	1	1
20181224	0	0	0	0	0	0
20181226	0	0	0	0	0	0
20181227	0	0	0	0	0	0
20181228	1	1	1	1	1	1
20181231	1	1	0	1	1	1
20181233	1	1	0	0	1	0
20181234	1	1	1	1	1	1
20181235	0	0	0	0	0	0
20190107	0	0	0	0	0	0
20190108	1	1	1	1	1	1
20190109	0	0	0	0	0	0
20190110	0	0	0	0	0	0
20190111	0	0	0	0	0	0
20190114	1	1	1	1	1	1
20190115	0	0	0	0	0	1
20190116	0	0	0	0	0	0
20190117	1	1	1	1	1	1

20190118	0	0	0	0	0	0
20190122	1	1	1	1	1	0
20190123	0	0	0	0	0	0
20190124	0	0	0	0	0	0
20190125	1	1	1	1	1	1
20190128	0	0	0	0	0	0
20190129	0	0	0	0	0	0
20190130	0	0	0	0	0	0
20190131	0	0	0	0	0	0
20190132	0	0	0	0	0	0
20190204	0	0	0	0	0	0
20190205	1	1	1	1	1	1
20190206	0	0	0	0	0	0
20190207	1	1	0	0	1	0
20190208	1	1	1	1	1	0
20190211	0	0	0	0	0	0
20190212	1	1	1	1	1	1
20190213	1	1	1	1	1	0
20190214	1	1	1	1	1	1
20190215	0	0	0	0	0	0
20190219	1	1	1	1	1	1
20190220	1	1	1	1	1	1
20190221	1	1	1	1	1	0
20190222	0	0	0	0	0	0
20190225	1	1	1	0	1	0
20190226	1	1	0	0	1	0
20190227	1	1	1	1	1	1
20190228	1	1	1	1	1	1
20190229	1	1	1	1	1	1
20190304	1	1	1	1	1	1
20190305	1	1	1	1	1	1
20190306	0	0	0	0	0	0
20190307	1	1	1	1	1	0
20190308	1	1	1	1	1	1
20190311	0	0	0	0	0	0
20190312	1	1	1	1	1	1
20190313	1	1	1	1	1	1
20190314	0	0	0	0	0	0
20190315	1	1	1	1	1	1
20190318	1	1	0	0	1	1
20190319	1	1	1	0	1	1

20190320	1	1	1	1	1	1
20190321	1	1	1	1	1	0
20190322	1	1	1	1	1	1
20190325	1	1	1	1	1	0
20190326	1	1	1	1	1	1
20190327	0	0	0	0	0	0
20190328	0	0	0	0	0	0
20190329	1	1	1	1	1	1
20190332	0	0	0	0	0	0
20190333	0	0	0	0	0	0
20190334	1	1	1	1	1	1
20190335	0	0	0	0	0	0
20190336	1	1	1	1	1	1
20190408	1	1	1	1	1	1
20190409	1	1	1	1	1	0
20190410	0	0	0	0	0	0
20190411	1	1	1	1	1	0
20190412	1	1	1	1	1	0
20190415	0	0	0	0	0	0
20190416	0	0	0	0	0	0
20190417	0	0	0	0	0	0
20190418	0	0	0	0	0	0
20190422	1	1	1	1	1	0
20190423	0	0	0	0	0	0
20190424	1	1	1	1	1	1
20190425	1	1	1	1	1	1
20190426	0	0	0	0	0	0
20190429	0	0	0	0	0	0
20190430	1	1	1	1	1	0
20190431	1	1	1	1	1	0
20190432	1	1	1	1	1	1
20190433	1	1	1	1	1	0
20190506	0	0	0	0	0	0
20190507	1	1	1	0	1	1
20190508	1	1	1	1	1	0
20190509	1	1	1	1	1	1
20190510	1	1	1	1	1	0
20190513	1	1	1	1	1	0
20190514	0	0	0	0	0	0
20190515	1	1	1	1	1	0
20190516	1	1	1	1	1	1

20190517	1	1	1	1	1	0
20190520	1	1	1	1	1	1
20190521	1	1	1	1	1	0
20190522	0	0	0	0	0	0
20190523	1	1	1	1	1	0
20190524	1	1	1	1	1	1
20190528	1	1	1	1	1	0
20190529	1	1	1	0	1	0
20190530	1	1	1	1	1	0
20190531	1	1	1	0	1	0
20190603	1	1	1	1	1	0
20190604	1	1	1	1	1	0
20190605	1	1	1	1	1	1
20190606	1	1	0	0	1	0
20190607	1	1	1	1	1	1
20190610	0	0	0	0	0	0
20190611	1	1	1	1	1	0
20190612	1	1	1	1	1	0
20190613	1	1	1	1	1	0
20190614	1	1	1	1	1	0
20190617	1	1	0	0	1	0
20190618	1	1	1	1	1	0
20190619	1	1	1	1	1	1
20190620	1	1	0	0	1	0
20190621	1	1	1	1	1	0
20190624	1	1	1	1	1	0
20190625	1	1	1	1	1	0
20190626	1	1	1	1	1	0
20190627	1	1	1	1	1	0
20190628	1	1	1	1	1	0
20190631	0	0	0	0	0	0
20190632	0	0	0	0	0	0
20190633	1	1	1	1	1	1
20190635	0	0	0	0	0	0
20190708	1	1	1	1	1	0
20190709	0	0	0	0	0	0
20190710	1	1	1	1	1	1
20190711	1	1	1	1	1	0
20190712	1	1	1	1	1	0
20190715	1	1	1	1	1	0
20190716	1	1	1	1	1	1

20190717	1	1	1	1	1	1
20190718	1	1	1	1	1	0
20190719	0	0	0	0	0	0
20190722	1	1	1	1	1	0
20190723	1	1	1	1	1	0
20190724	1	1	1	1	1	1
20190725	1	1	1	1	1	1
20190726	0	0	0	0	0	0
20190729	0	0	0	0	0	0
20190730	0	0	0	0	0	0
20190731	1	1	1	1	1	1
20190732	1	1	1	1	1	0
20190733	1	1	1	1	1	0
20190805	1	1	0	0	1	1
20190806	1	1	1	1	1	1
20190807	1	1	1	1	1	0
20190808	1	1	1	0	1	0
20190809	0	0	0	0	0	0
20190812	1	1	1	1	1	0
20190813	1	1	1	1	1	0
20190814	1	1	1	1	1	0
20190815	1	1	1	1	1	1
20190816	1	1	1	0	1	1
20190819	0	0	0	0	0	0
20190820	1	1	0	0	1	1
20190821	1	1	1	1	1	1
20190822	0	0	0	0	0	0
20190823	0	0	0	0	0	0
20190826	0	0	0	0	0	0
20190827	0	0	0	0	0	0
20190828	1	1	1	1	1	1
20190829	0	0	0	0	0	0
20190830	1	1	1	1	1	0
20190903	1	1	1	1	1	1
20190904	1	1	1	0	1	0
20190905	0	0	0	0	0	0
20190906	1	1	0	0	1	1
20190909	1	1	1	1	1	0
20190910	1	1	1	1	1	0
20190911	1	1	1	1	1	0
20190912	1	1	1	1	1	0

20190913	0	0	0	0	0	0
20190916	1	1	1	1	1	1
20190917	1	1	1	1	1	0
20190918	1	1	1	1	1	0
20190919	0	0	0	0	0	0
20190920	1	1	1	1	1	0
20190923	0	0	0	0	0	0
20190924	1	1	1	1	1	1
20190925	0	0	0	0	0	0
20190926	1	1	1	1	1	0
20190927	0	0	0	0	0	0
20190930	1	1	1	1	1	0
20190931	0	0	0	0	0	0
20190932	0	0	0	0	0	0
20190933	0	0	0	0	0	0
20190934	0	0	0	0	0	0
20191007	0	0	0	0	0	0
20191008	1	0	0	0	1	1
20191009	1	1	1	1	1	0
20191010	1	1	1	1	1	0
20191011	1	1	0	0	1	1
20191014	0	0	0	0	0	0
20191015	1	1	1	1	1	1
20191016	1	1	1	1	1	1
20191017	1	1	1	1	1	1
20191018	0	0	0	0	0	0
20191021	1	1	1	1	1	1
20191022	0	0	0	0	0	0
20191023	1	1	1	1	1	0
20191024	0	0	0	0	0	0
20191025	1	1	1	1	1	0
20191028	1	1	1	1	1	0
20191029	1	1	1	1	1	1
20191030	1	1	1	1	1	0
20191031	1	1	1	1	1	1
20191032	1	1	1	1	1	1
20191104	1	1	1	1	1	0
20191105	0	0	0	0	0	0
20191106	1	1	1	1	1	1
20191107	1	1	1	1	1	0
20191108	1	1	1	1	1	0

20191111	1	1	1	1	1	0
20191112	1	1	1	1	1	0
20191113	0	0	0	0	0	0
20191114	0	0	0	0	0	0
20191115	1	1	1	1	1	1
20191118	1	1	1	1	1	0
20191119	1	1	1	0	1	0
20191120	0	0	0	0	0	0
20191121	0	0	0	0	0	0
20191122	0	0	0	0	0	0
20191125	0	0	0	0	0	0
20191126	0	0	0	0	0	0
20191127	1	1	1	1	1	1
20191129	1	1	1	1	1	1
20191202	1	1	1	1	1	0
20191203	1	1	1	1	1	0
20191204	1	1	0	0	0	0
20191205	1	1	1	1	1	1
20191206	0	0	0	0	0	0
20191209	0	0	0	0	0	0
20191210	1	1	1	1	1	0
20191211	0	0	0	0	0	0
20191212	1	1	0	0	1	0
20191213	0	0	0	0	0	0
20191216	1	1	1	1	1	0
20191217	0	0	0	0	0	0
20191218	1	1	1	1	1	0
20191219	0	0	0	0	0	0
20191220	1	1	1	1	1	1
20191223	0	0	0	0	0	0
20191224	0	0	0	0	0	0
20191226	0	0	0	0	0	0
20191227	0	0	0	0	0	0

9.4 Cointegration Rank Estimates

Cointegration Rank was estimated using Johansen maximum likelihood (ML). Results for corn and soybean are presented here for all 6 models using 50 minutes of age. We expect Models A & B to have a rank of 1, Models C, D & E to have a rank of 2 and Model F to have a rank of 3.

Model A & B	Corn		Soybean	
Age (minutes)	Rank 0	Rank 1	Rank 0	Rank 1
10	0.0%	100.0%	0.0%	100.0%
20	0.0%	100.0%	0.0%	100.0%
30	0.0%	100.0%	0.4%	99.6%
40	0.0%	100.0%	0.4%	99.6%
50	0.0%	100.0%	0.4%	99.6%
60	0.0%	100.0%	0.4%	99.6%
70	0.0%	100.0%	0.4%	99.6%
80	0.0%	100.0%	0.4%	99.6%

Model C	Corn			Soybean		
Age (minutes)	Rank 0	Rank 1	Rank 2	Rank 0	Rank 1	Rank 2
10	0.0%	0.4%	99.6%	0.0%	0.4%	96.0%
20	0.0%	2.2%	97.8%	0.0%	2.2%	91.0%
30	0.0%	5.6%	94.0%	0.4%	4.9%	85.5%
40	0.0%	11.2%	88.4%	0.4%	11.2%	77.2%
50	0.0%	17.6%	82.1%	0.4%	21.6%	65.9%
60	0.0%	21.0%	78.7%	0.4%	29.5%	57.5%
70	0.0%	28.5%	71.3%	0.4%	34.7%	51.6%
80	0.0%	30.3%	69.4%	0.4%	41.8%	44.7%

Model D	Corn			Soybean		
Age (minutes)	Rank 0	Rank 1	Rank 2	Rank 0	Rank 1	Rank 2
10	0.0%	0.0%	100.0%	0.0%	0.0%	96.4%
20	0.0%	3.0%	97.0%	0.0%	2.6%	90.6%
30	0.0%	7.5%	92.2%	0.4%	6.7%	83.8%
40	0.0%	12.4%	87.3%	0.4%	12.7%	75.9%
50	0.0%	17.6%	82.1%	0.4%	23.1%	64.7%
60	0.0%	21.7%	78.0%	0.4%	31.3%	56.0%
70	0.0%	25.1%	74.6%	0.4%	38.4%	48.7%
80	0.0%	30.7%	68.8%	0.7%	43.7%	43.1%

Model E	Corn			Soybean		
Age (minutes)	Rank 0	Rank 1	Rank 2	Rank 0	Rank 1	Rank 2
10	0.0%	0.0%	100.0%	0.0%	0.0%	96.4%
20	0.0%	0.0%	100.0%	0.0%	0.0%	93.1%
30	0.0%	0.0%	100.0%	0.0%	0.4%	89.6%
40	0.0%	0.0%	100.0%	0.0%	0.4%	86.7%
50	0.0%	0.0%	100.0%	0.0%	0.4%	84.0%
60	0.0%	0.4%	99.6%	0.0%	0.4%	81.4%
70	0.0%	0.4%	99.6%	0.0%	0.4%	79.0%
80	0.0%	0.4%	99.6%	0.0%	0.4%	76.7%

Model F	Corn				Soybean			
Age (minutes)	Rank 0	Rank 1	Rank 2	Rank 3	Rank 0	Rank 1	Rank 2	Rank 3
10	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	96.4%
20	0.0%	0.0%	13.1%	86.9%	0.0%	0.0%	8.2%	85.4%
30	0.0%	0.0%	16.9%	83.1%	0.0%	0.0%	11.9%	79.2%
40	0.0%	0.0%	13.5%	86.5%	0.0%	0.0%	12.7%	76.0%
50	0.0%	0.0%	13.5%	86.5%	0.0%	0.0%	12.3%	73.9%
60	0.0%	0.0%	19.1%	80.9%	0.0%	0.0%	11.2%	72.6%
70	0.0%	0.0%	17.2%	82.8%	0.0%	0.0%	9.7%	71.6%
80	0.0%	0.0%	16.1%	83.9%	0.0%	0.0%	9.0%	70.1%

9.5 Lag Order Selection

This is a typical result of lag order selection methods. A preferred option is the AIC which was used in Cao, Hansch, and Wang (2009) and Arzandeh and Frank (2019). However, we note that high lag orders (similar to the LR method) better fits these data given the high degree of bid-ask bounce as well as results in much better behaviours in the VECM residuals with regards to autocorrelation. The order selection results are given in the table below.

Model	Criteria Used	Avg of Daily Order Selection	StDev of Daily Order Selection	Avg of Daily Order Selection	StDev of Daily Order Selection
		Corn		Soybean	
Model A	AIC	9.8	7.7	9.1	6.8
Model A	LR	69.6	11.4	69.9	11.0
Model B	AIC	10.2	8.8	9.6	7.7
Model B	LR	68.9	12.1	69.7	11.7
Model C	AIC	9.1	9.8	8.9	10.3
Model C	LR	73.7	7.7	74.3	7.0
Model D	AIC	9.5	10.6	9.0	10.7
Model D	LR	73.3	8.0	74.2	7.1
Model E	AIC	6.0	5.3	4.8	3.0
Model E	LR	73.0	7.8	73.0	8.4
Model F	AIC	6.5	8.8	5.4	7.0
Model F	LR	76.0	5.3	75.9	5.4

9.6 Autocorrelation in the VECM residuals

Residuals were tested for autocorrelation using the Lagrange Multiplier test. VECM were excluded if the LM test found more than 4 occurrences of autocorrelation in the residuals.

Count of Autocorrelation Occurrences Identified in first 10 lags of residuals									
Date (YYYYMMDD)	Max Age of Orders (minutes)								All Orders
	10	20	30	40	50	60	70	80	
20181203	0	0	2	1	0	0	0	0	0
20181204	0	0	0	1	1	0	0	0	1
20181205	0	0	0	0	0	0	0	0	0
20181206	1	1	1	0	1	1	2	2	2
20181207	0	0	0	0	0	0	0	1	0
20181210	0	0	0	0	0	0	0	0	0
20181211	0	1	1	1	1	1	1	1	3
20181212	1	0	1	1	0	0	0	0	0
20181213	0	0	0	0	0	0	1	0	1
20181214	1	0	0	0	0	0	0	0	1
20181217	1	1	0	1	2	0	0	0	0
20181218	3	1	3	3	4	4	5	5	2
20181219	1	0	0	0	0	0	0	0	0
20181220	0	0	0	0	0	0	0	0	0
20181221	1	0	0	0	0	0	1	1	0
20181224	0	1	0	0	1	1	1	1	1
20181226	2	3	1	3	1	1	1	1	1
20181227	1	2	4	4	4	3	4	4	3
20181228	0	0	0	0	0	0	1	1	1
20181231	3	3	4	4	4	4	5	5	7
20181233	0	2	1	2	1	0	0	0	0
20181234	0	1	0	0	1	0	0	0	0
20181235	1	2	1	1	0	0	1	0	0
20190107	1	1	3	3	4	2	3	4	2
20190108	0	0	0	1	1	1	1	1	1
20190109	1	3	3	2	2	2	3	1	1
20190110	0	0	1	0	1	1	1	1	1
20190111	0	0	0	0	0	0	0	0	1
20190114	3	5	4	6	7	7	6	7	8
20190115	0	0	0	0	1	1	0	0	0
20190116	0	0	0	0	2	1	1	1	1
20190117	0	1	1	0	0	0	0	0	0

20190118	0	0	0	0	0	0	1	1	2
20190122	0	0	0	0	0	0	0	0	2
20190123	2	1	0	2	1	1	1	2	1
20190124	2	0	1	3	2	3	3	5	1
20190125	0	1	2	2	1	1	0	1	2
20190128	3	2	1	3	2	1	2	1	2
20190129	0	0	0	0	0	0	0	0	1
20190130	2	0	0	1	0	0	0	1	1
20190131	1	0	1	3	1	3	2	3	3
20190132	0	1	5	6	3	3	3	3	4
20190204	0	0	1	1	1	1	1	2	1
20190205	0	0	0	0	0	0	0	0	0
20190206	0	0	0	0	0	0	0	0	6
20190207	0	4	5	4	4	4	4	4	3
20190208	0	0	1	0	0	0	0	0	0
20190211	0	0	0	0	0	0	0	0	0
20190212	0	0	2	0	1	0	0	0	0
20190213	1	0	0	0	0	0	1	0	0
20190214	0	0	0	0	0	0	0	0	0
20190215	1	1	1	1	1	1	1	1	1
20190219	1	0	2	2	0	1	0	2	1
20190220	1	2	4	3	4	3	3	3	3
20190221	0	0	1	1	0	0	0	1	1
20190222	0	0	0	0	0	0	0	0	0
20190225	0	1	3	1	1	1	2	2	1
20190226	0	0	1	2	3	2	3	3	4
20190227	0	2	1	2	2	2	2	2	1
20190228	0	0	1	1	2	0	1	2	1
20190229	0	0	0	2	2	1	1	1	0
20190304	0	1	2	1	2	1	2	0	0
20190305	1	0	0	1	0	2	1	1	0
20190306	0	1	0	1	2	0	1	1	0
20190307	2	2	2	2	5	4	3	3	4
20190308	0	1	0	1	1	0	0	1	1
20190311	0	0	3	2	5	0	1	3	1
20190312	0	3	4	5	2	3	3	3	2
20190313	0	0	0	0	1	1	0	0	0
20190314	0	0	0	1	0	0	0	1	1
20190315	2	6	4	5	5	6	5	6	3
20190318	0	1	2	2	2	2	3	3	0
20190319	0	2	2	1	1	2	2	2	0

20190320	1	1	0	2	0	1	0	0	1
20190321	1	1	2	1	1	0	1	0	0
20190322	0	0	0	0	0	0	1	0	0
20190325	0	1	0	1	1	1	0	0	0
20190326	0	0	0	0	0	0	0	0	0
20190327	0	0	1	0	0	0	0	0	1
20190328	3	1	0	0	2	0	1	1	1
20190329	0	0	0	0	0	0	0	0	0
20190408	0	2	3	3	0	0	0	0	2
20190409	0	0	1	1	1	1	1	1	0
20190410	1	0	0	0	0	0	0	1	0
20190411	1	1	1	1	2	0	1	1	1
20190412	0	1	2	1	4	2	2	1	1
20190415	3	3	3	2	2	3	2	3	4
20190416	0	0	0	0	0	0	0	0	0
20190417	0	1	0	0	1	1	1	0	0
20190418	0	0	0	0	0	0	2	1	0
20190422	0	1	0	1	0	0	0	0	0
20190423	0	2	1	0	0	1	0	0	3
20190424	0	0	0	0	0	0	0	0	0
20190425	0	0	1	0	0	0	0	0	1
20190426	4	5	5	4	4	4	3	3	2
20190429	2	0	0	0	0	0	0	0	0
20190430	1	1	1	1	1	2	1	2	1
20190431	1	1	0	1	0	0	0	0	1
20190432	1	0	2	1	1	1	1	2	2
20190433	0	0	0	0	0	0	0	0	0
20190506	6	6	7	7	7	7	7	7	7
20190507	0	0	0	1	1	1	0	0	1
20190508	0	0	0	0	0	0	0	0	0
20190509	0	0	1	3	0	0	0	0	1
20190510	3	2	1	1	1	2	2	2	0
20190513	3	3	2	0	2	0	0	0	1
20190514	1	6	5	5	7	4	6	6	2
20190515	0	0	0	0	0	0	0	0	0
20190516	4	5	5	5	5	5	5	5	5
20190517	1	5	3	4	4	4	5	5	1
20190520	5	7	2	3	2	2	2	1	2
20190521	0	0	0	0	0	0	0	0	0
20190522	1	6	3	4	5	4	5	3	1
20190523	1	0	0	0	0	0	0	0	0

20190524	2	3	2	2	1	1	2	1	0
20190528	2	2	2	1	1	1	1	1	0
20190529	1	0	0	1	0	0	1	1	2
20190530	0	0	0	0	0	0	0	0	0
20190531	4	2	3	3	3	3	3	3	3
20190603	1	4	3	2	1	0	0	0	0
20190604	2	0	1	0	0	0	0	0	1
20190605	1	0	0	0	0	0	0	0	0
20190606	0	2	0	0	0	0	0	0	0
20190607	0	1	1	2	0	1	2	2	0
20190610	1	2	3	2	2	3	2	2	1
20190611	0	0	0	0	0	0	1	1	0
20190612	2	1	1	2	1	2	2	1	1
20190613	0	0	0	0	0	0	0	1	2
20190614	3	1	0	2	1	1	1	1	2
20190617	0	1	0	0	0	0	0	0	0
20190618	2	0	0	0	0	0	0	0	0
20190619	3	3	3	3	3	3	3	2	3
20190620	1	0	3	1	1	1	1	2	1
20190621	0	0	1	0	0	0	0	0	2
20190624	0	2	2	1	1	1	2	1	0
20190625	1	0	0	0	0	0	0	0	0
20190626	0	3	0	3	2	3	2	2	1
20190627	0	1	0	1	1	1	1	0	0
20190628	1	0	0	0	0	0	1	0	0
20190631	0	0	0	0	1	1	1	1	1
20190632	4	4	3	3	3	2	2	2	1
20190633	3	4	3	3	1	1	1	0	0
20190635	0	0	3	3	4	2	3	2	2
20190708	3	1	4	5	4	4	3	3	3
20190709	4	5	7	6	5	5	5	6	3
20190710	5	1	0	2	2	2	2	2	1
20190711	1	1	2	2	2	2	2	2	3
20190712	6	8	9	9	7	7	7	7	6
20190715	0	0	0	0	2	0	0	0	1
20190716	2	2	2	3	4	2	3	4	2
20190717	0	0	1	1	0	0	0	0	0
20190718	0	0	0	0	0	0	0	0	0
20190719	1	2	2	2	3	2	2	2	2
20190722	0	1	1	1	1	1	1	1	2
20190723	1	1	2	0	1	1	1	1	1

20190724	0	0	0	0	0	0	0	0	0
20190725	1	3	2	2	2	3	3	2	3
20190726	2	3	3	5	5	7	5	5	1
20190729	1	2	1	2	3	1	1	1	3
20190730	0	1	0	1	1	0	1	0	1
20190731	0	0	0	0	0	1	0	0	0
20190732	2	1	2	3	3	0	0	0	1
20190733	1	2	2	2	2	2	2	2	0
20190805	0	1	0	0	1	0	0	0	0
20190806	0	4	3	5	5	3	3	2	2
20190807	1	3	4	6	5	5	4	4	2
20190808	0	2	0	1	1	0	0	0	2
20190809	1	0	1	1	1	2	2	2	0
20190812	4	5	6	6	6	6	5	5	4
20190813	2	2	2	4	5	4	4	3	6
20190814	3	5	5	3	3	3	2	2	0
20190815	1	2	2	2	3	0	0	1	1
20190816	0	4	6	5	4	3	3	2	1
20190819	1	1	2	0	0	1	0	2	3
20190820	1	2	1	2	2	2	2	2	0
20190821	1	1	1	1	1	1	1	1	1
20190822	0	2	4	4	3	2	2	2	3
20190823	0	0	0	0	0	0	0	0	2
20190826	1	1	0	0	0	0	0	0	0
20190827	0	0	0	0	0	0	0	0	0
20190828	0	2	0	0	1	0	1	0	0
20190829	1	0	0	0	0	0	0	0	0
20190830	2	0	0	0	0	1	1	1	4
20190903	0	0	0	0	1	0	0	1	0
20190904	0	0	0	0	0	0	0	0	0
20190905	1	6	6	4	6	5	5	4	2
20190906	1	0	0	0	0	0	0	0	0
20190909	0	1	1	2	2	3	3	2	1
20190910	0	0	0	0	0	0	0	0	0
20190911	0	0	0	1	1	0	1	1	0
20190912	0	0	0	0	0	0	0	0	0
20190913	2	0	2	1	1	1	1	0	0
20190916	0	2	2	2	1	1	2	2	0
20190917	1	1	2	3	2	1	3	3	0
20190918	1	0	0	0	0	0	0	0	0
20190919	2	2	4	3	3	2	3	4	2

20190920	0	3	1	2	2	1	1	1	1
20190923	1	0	1	1	3	4	8	7	5
20190924	2	1	1	0	0	0	0	0	1
20190925	0	2	1	1	1	1	1	1	1
20190926	0	0	0	0	0	0	0	0	0
20190927	2	7	7	7	7	6	4	4	1
20190930	1	0	1	1	1	1	1	1	1
20190931	0	0	0	1	1	1	0	1	0
20190932	0	4	2	2	3	2	1	1	1
20190933	0	1	0	0	1	1	0	0	0
20190934	0	1	2	1	1	1	2	2	0
20191007	2	1	0	1	2	1	2	2	1
20191008	5	3	3	3	3	3	3	2	4
20191009	0	0	0	0	0	0	0	0	0
20191010	1	1	1	0	1	1	1	1	0
20191011	0	1	0	0	0	0	0	0	0
20191014	0	2	0	0	0	0	0	0	0
20191015	0	0	1	0	1	1	1	0	0
20191016	0	0	0	0	0	0	0	0	0
20191017	0	0	0	0	0	0	0	0	1
20191018	1	6	5	5	6	6	4	4	5
20191021	1	4	3	3	2	2	2	2	3
20191022	3	3	2	3	2	3	2	2	1
20191023	2	1	2	1	1	1	1	1	1
20191024	0	1	1	1	0	0	1	0	1
20191025	3	1	1	0	0	0	0	0	0
20191028	0	4	4	4	3	3	3	3	4
20191029	0	2	2	2	1	2	2	2	2
20191030	0	0	0	0	0	0	0	0	0
20191031	0	0	0	0	0	0	0	1	1
20191032	5	6	6	6	6	7	6	6	4
20191104	0	0	0	0	0	0	0	0	0
20191105	1	1	1	2	2	1	1	0	1
20191106	1	0	0	0	0	0	0	0	1
20191107	0	0	0	0	0	0	0	0	0
20191108	0	0	0	0	0	0	0	0	0
20191111	3	3	5	2	3	2	2	2	3
20191112	0	2	0	0	0	0	0	0	3
20191113	0	2	2	2	2	1	0	0	0
20191114	0	1	2	2	2	2	2	2	2
20191115	0	0	0	0	0	0	0	0	0

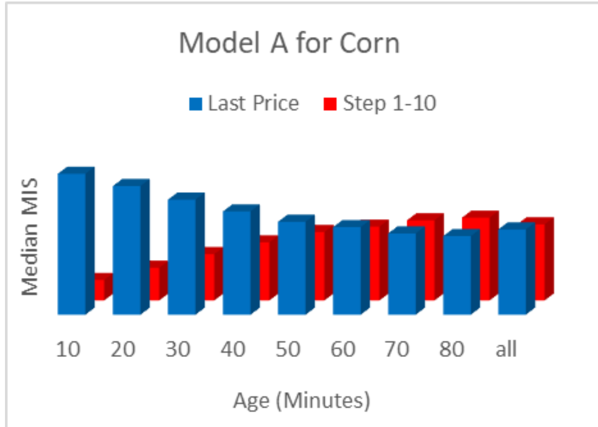
20191118	0	3	2	1	1	1	0	1	2
20191119	0	0	0	0	0	0	0	0	0
20191120	0	0	0	1	1	1	1	1	1
20191121	0	0	3	1	2	1	3	2	6
20191122	0	1	1	0	0	1	1	2	5
20191125	0	1	0	0	0	0	0	0	0
20191126	1	0	2	1	2	1	1	1	0
20191127	0	0	0	0	0	0	0	0	1
20191129	1	0	0	0	0	0	0	0	0
20191202	0	0	0	0	0	0	0	0	0
20191203	1	2	3	1	1	2	0	0	0
20191204	0	1	2	2	3	2	2	2	8
20191205	0	0	0	0	0	0	0	0	3
20191206	1	1	1	1	0	1	0	0	0
20191209	0	0	0	1	3	0	1	1	0
20191210	0	0	0	0	0	0	0	0	0
20191211	0	0	0	1	1	1	1	1	5
20191212	6	4	4	4	4	5	4	4	4
20191213	2	0	0	0	0	1	1	1	3
20191216	0	0	0	0	0	0	0	0	1
20191217	0	1	0	0	0	0	0	0	3
20191218	0	2	2	2	0	1	0	0	1
20191219	2	8	8	8	9	10	9	9	1
20191220	1	3	1	2	2	1	1	1	0
20191223	1	2	2	2	1	1	1	1	0
20191224	1	0	0	0	0	0	0	0	1
20191226	0	1	0	1	1	2	1	0	0
20191227	0	0	0	0	1	1	0	0	0

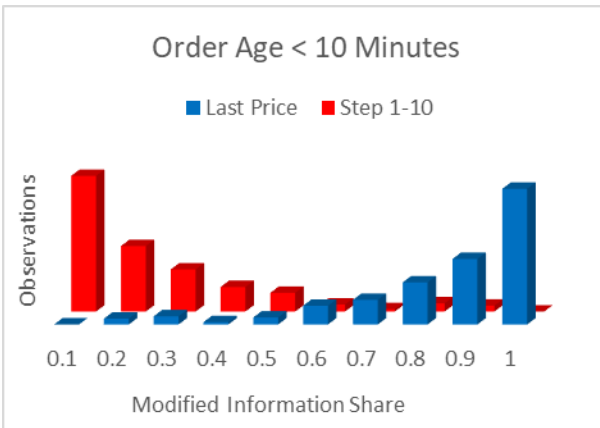
9.7 Modified Information Shares

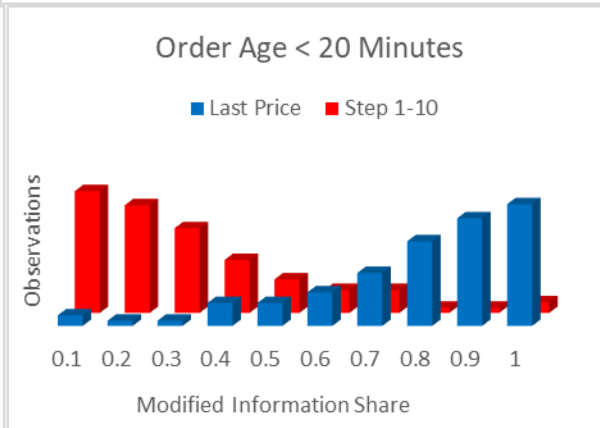
The MIS was calculated for each day and then reported as an average for the period after removing model runs that resulted in more than 4 instances in autocorrelation in the first 10 lags. The distribution of the MIS is bound by 0% and 100%, therefore is susceptible to significant skew that may be particularly notable when the average tends toward one of the bounds. The following charts show the distribution of the MIS for each of the models specified.

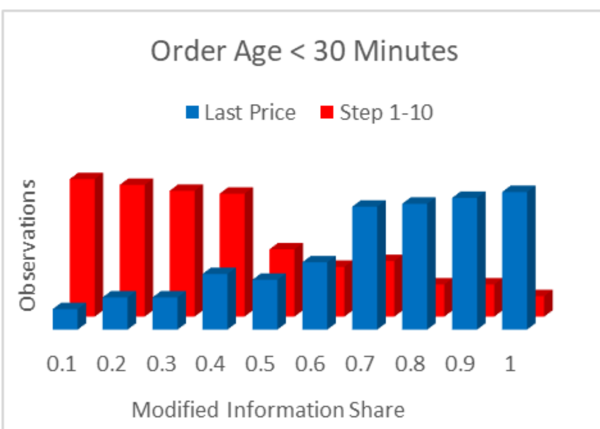
9.7.1 Histogram for MIS for Corn in Model A

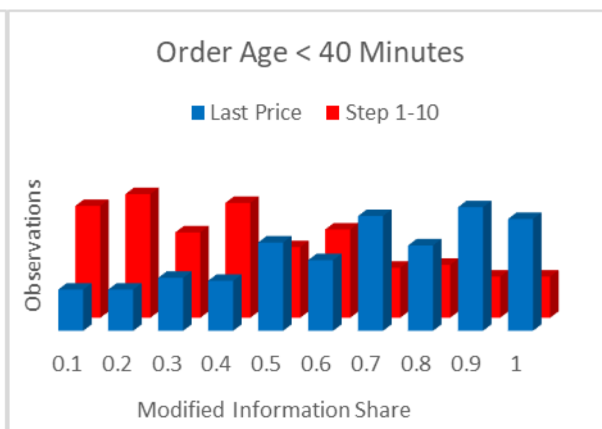
Modified Information Share (Median)		
Max Age	Last Price	Steps 1-10
10	0.87	0.13
20	0.80	0.20
30	0.71	0.29
40	0.64	0.36
50	0.58	0.42
60	0.54	0.46
70	0.50	0.50
80	0.49	0.51
all	0.53	0.47

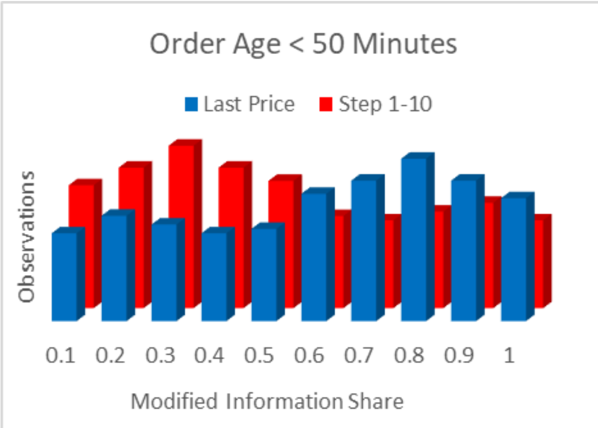


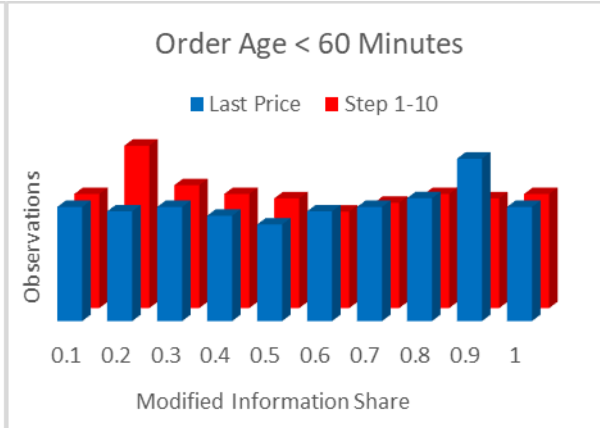


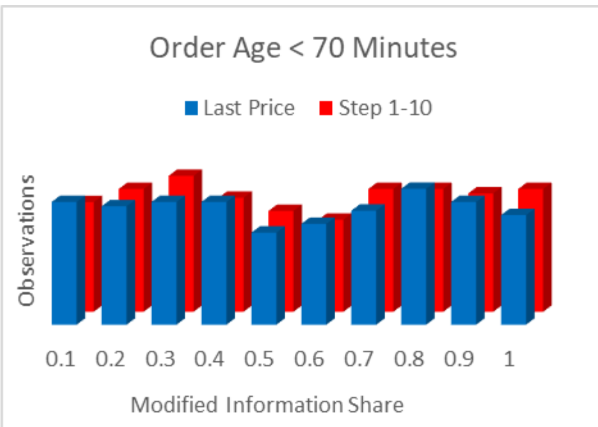


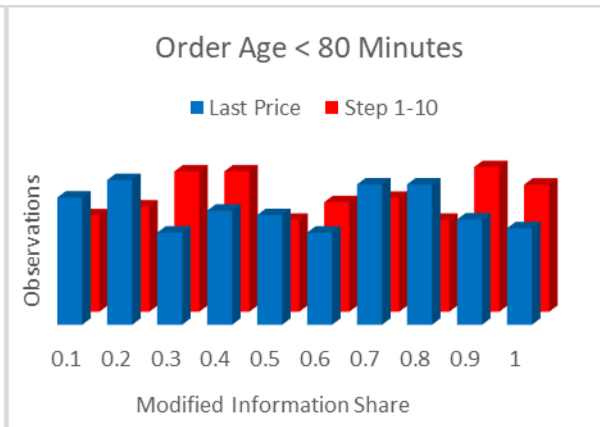





Order Age < 50 Minutes
Last Price
Step 1-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share


Order Age < 60 Minutes
Last Price
Step 1-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

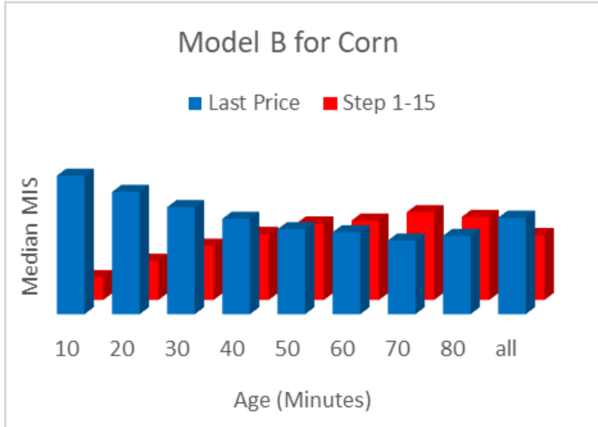

Order Age < 70 Minutes
Last Price
Step 1-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

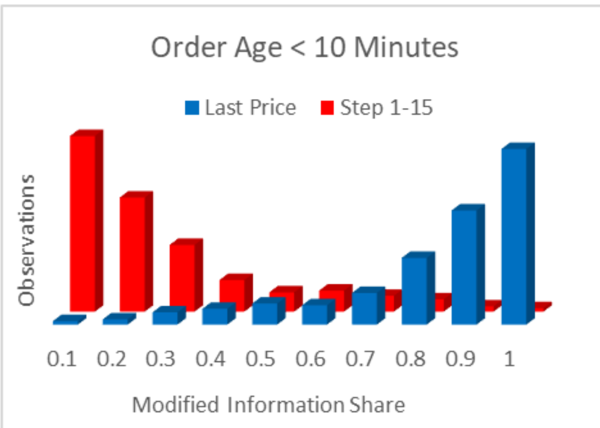

Order Age < 80 Minutes
Last Price
Step 1-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

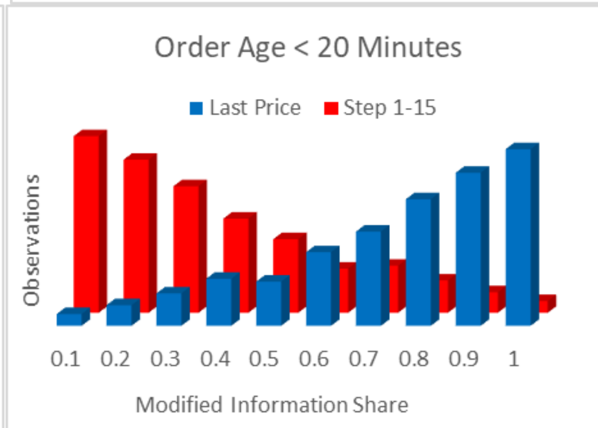

All Orders
Last Price
Step 1-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

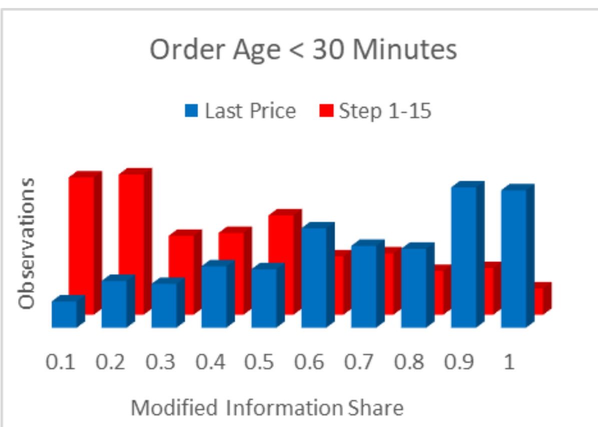
9.7.2 Histogram for MIS for Corn in Model B

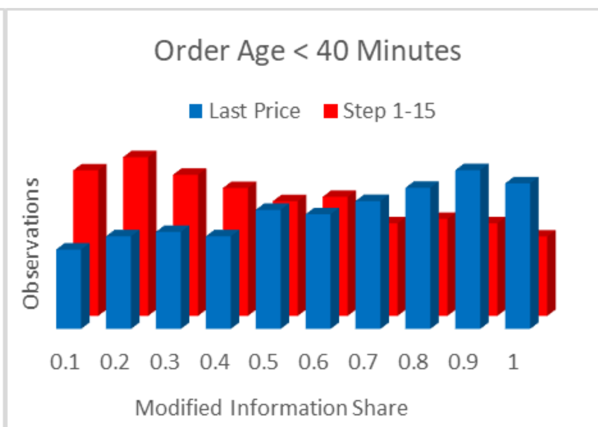
Modified Information Share (Median)		
Max Age	Last Price	Steps 1-15
10	0.86	0.14
20	0.76	0.24
30	0.66	0.34
40	0.59	0.41
50	0.53	0.47
60	0.51	0.49
70	0.46	0.54
80	0.49	0.51
all	0.60	0.40

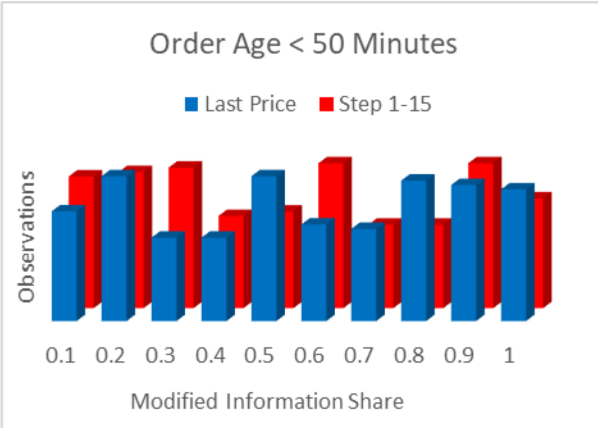


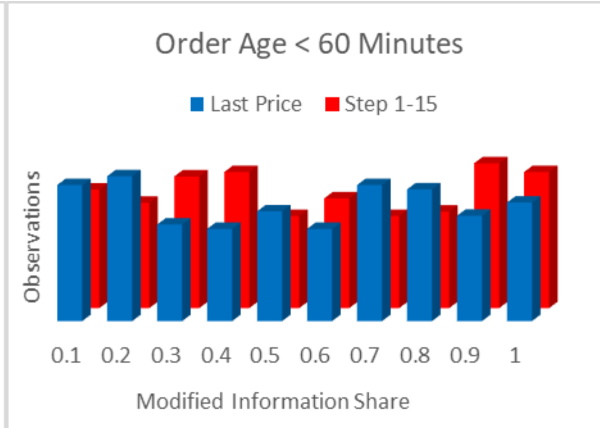


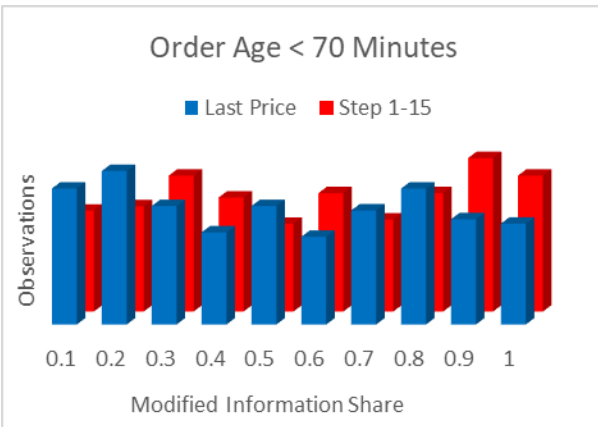


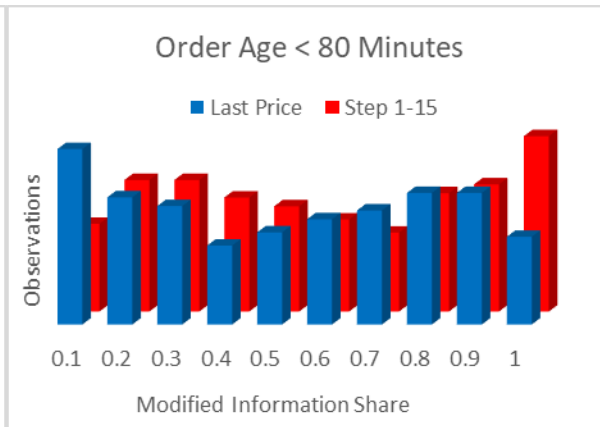


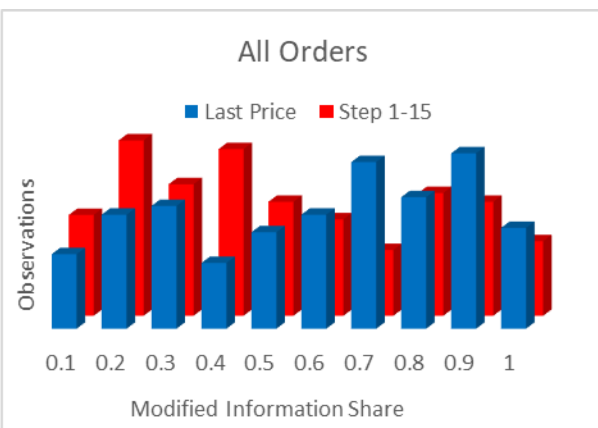



Order Age < 50 Minutes
Last Price
Step 1-15
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share


Order Age < 60 Minutes
Last Price
Step 1-15
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

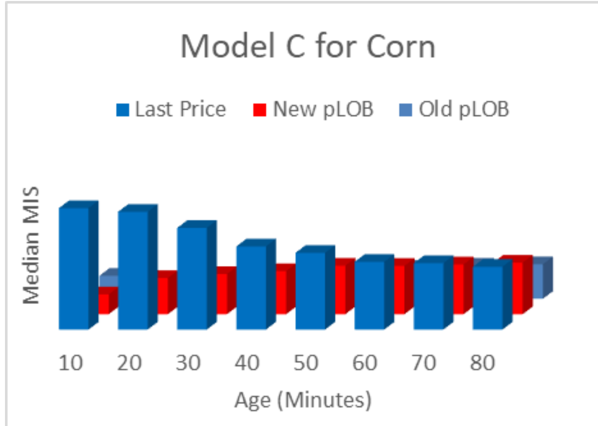

Order Age < 70 Minutes
Last Price
Step 1-15
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

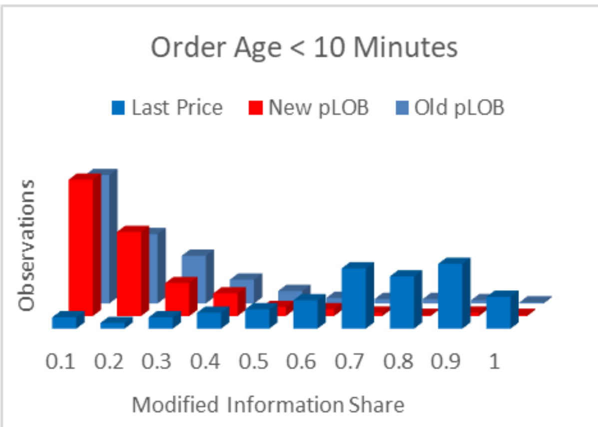

Order Age < 80 Minutes
Last Price
Step 1-15
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

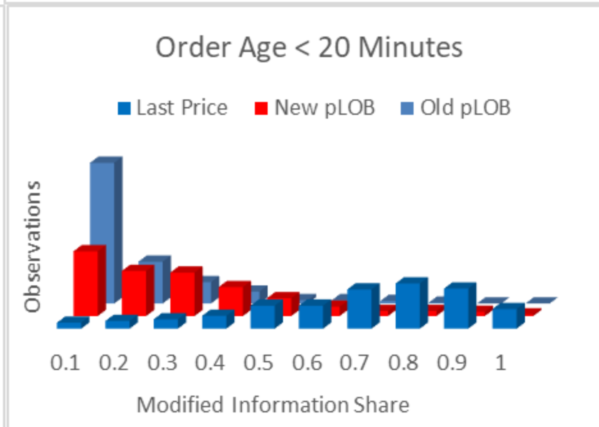

All Orders
Last Price
Step 1-15
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

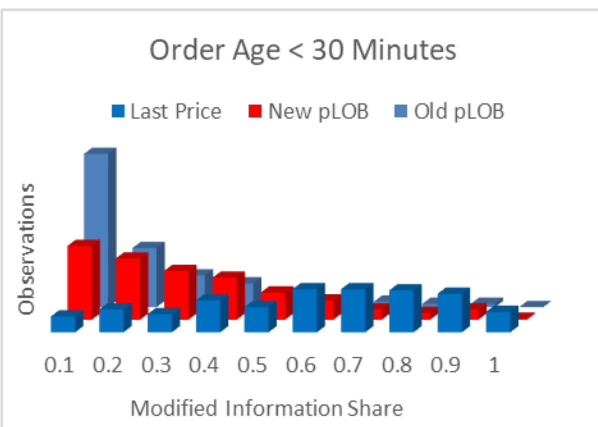
9.7.3 Histogram for MIS for Corn in Model C

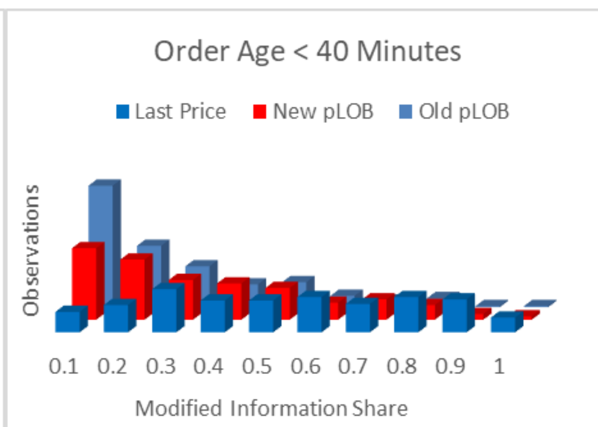
Modified Information Share (Median)			
Max Age	Last Price	New pLOB	Old pLOB
10	0.70	0.11	0.13
20	0.68	0.21	0.07
30	0.58	0.23	0.09
40	0.48	0.25	0.14
50	0.44	0.28	0.17
60	0.39	0.28	0.18
70	0.38	0.29	0.19
80	0.36	0.30	0.20

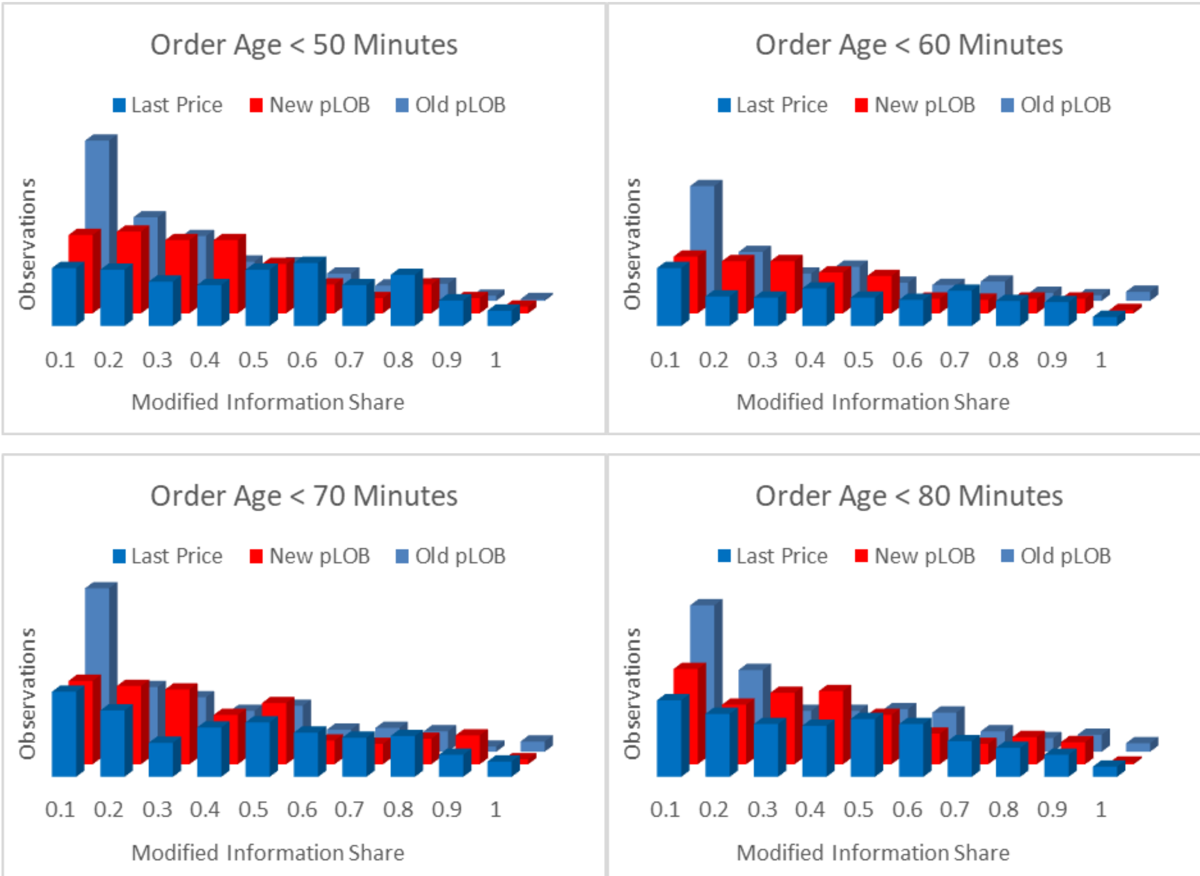






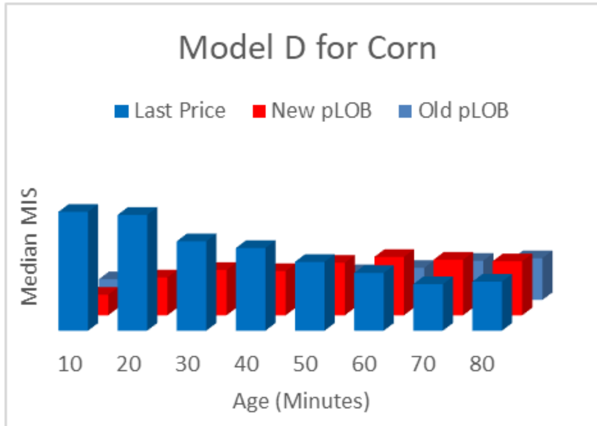


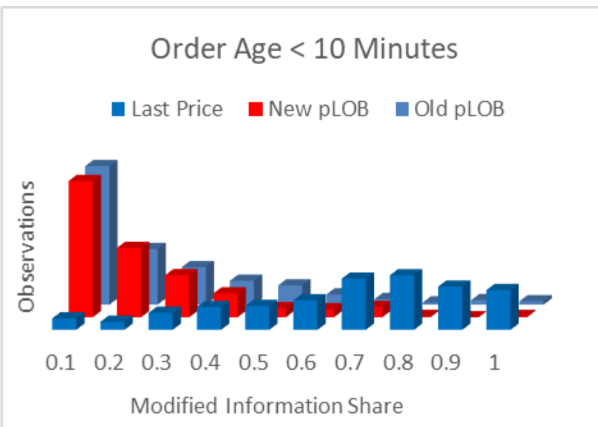


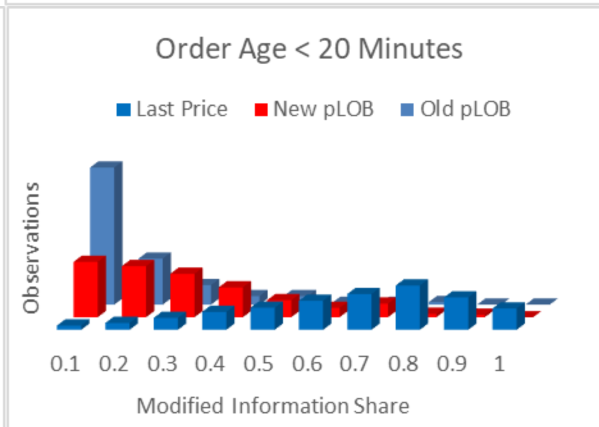

Order Age < 50 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 60 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 70 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 80 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

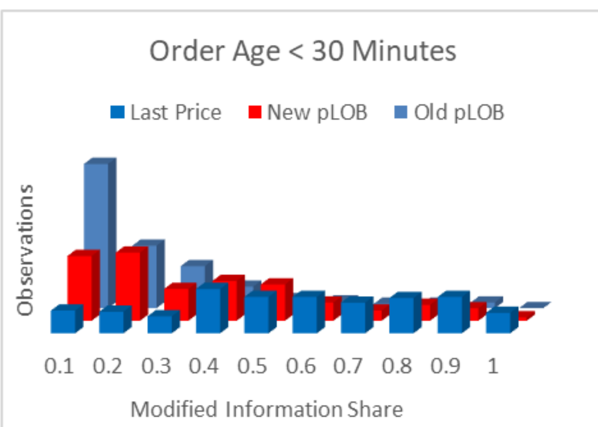
9.7.4 Histogram for MIS for Corn in Model D

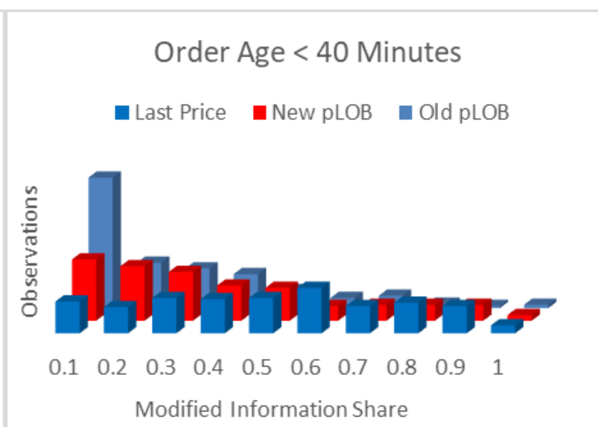
Modified Information Share (Median)			
Max Age	Last Price	New pLOB	Old pLOB
10	0.68	0.12	0.12
20	0.67	0.22	0.07
30	0.51	0.26	0.11
40	0.48	0.26	0.13
50	0.40	0.30	0.16
60	0.33	0.34	0.18
70	0.27	0.32	0.23
80	0.28	0.31	0.24

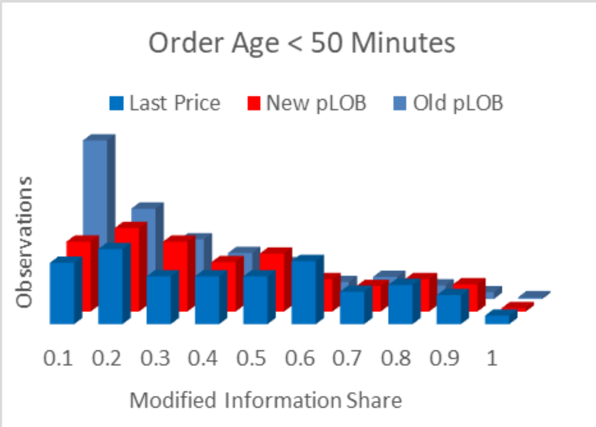


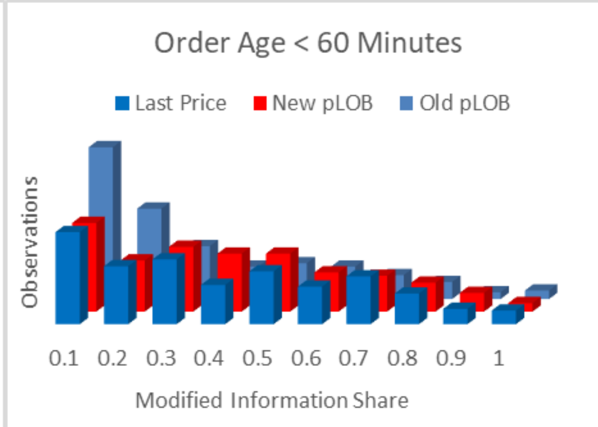


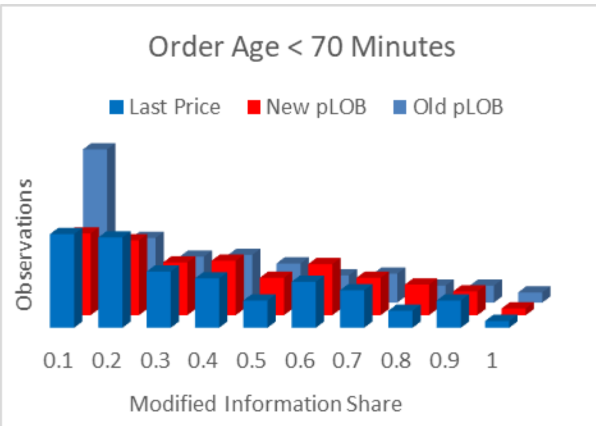


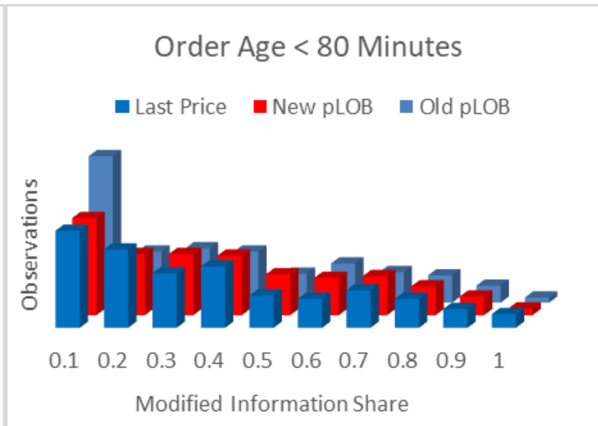





Order Age < 50 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Modified Information Share

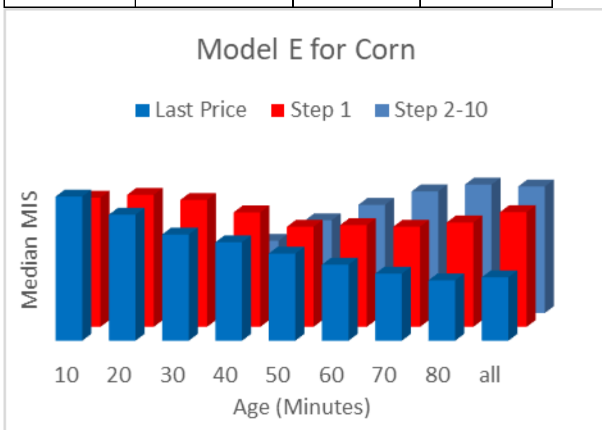

Order Age < 60 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Modified Information Share

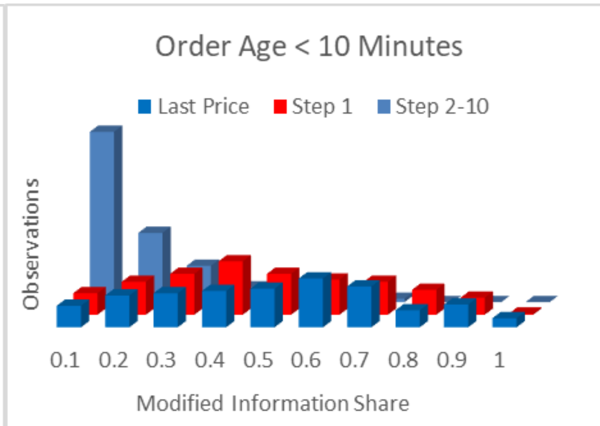

Order Age < 70 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Modified Information Share

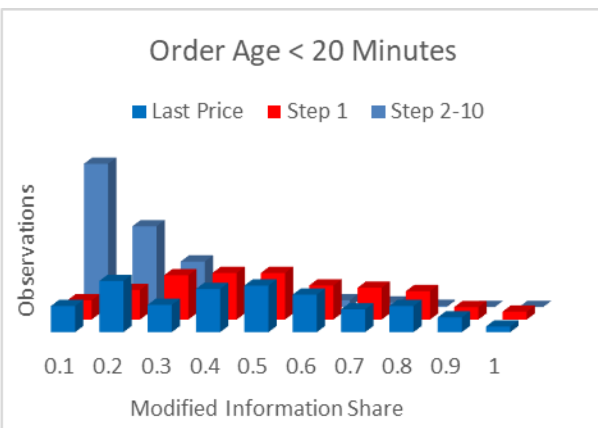

Order Age < 80 Minutes
Last Price
New pLOB
Old pLOB
Observations
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Modified Information Share

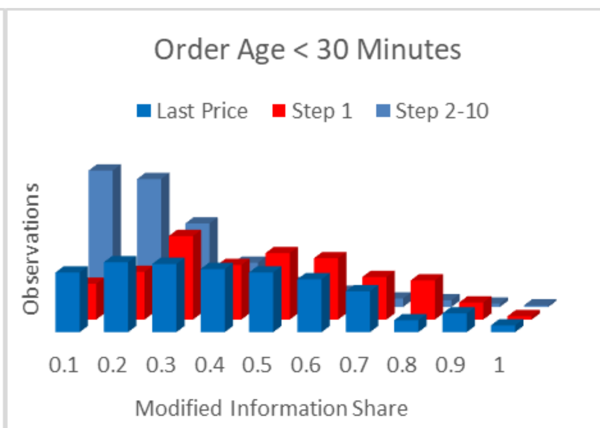
9.7.5 Histogram for MIS for Corn in Model E

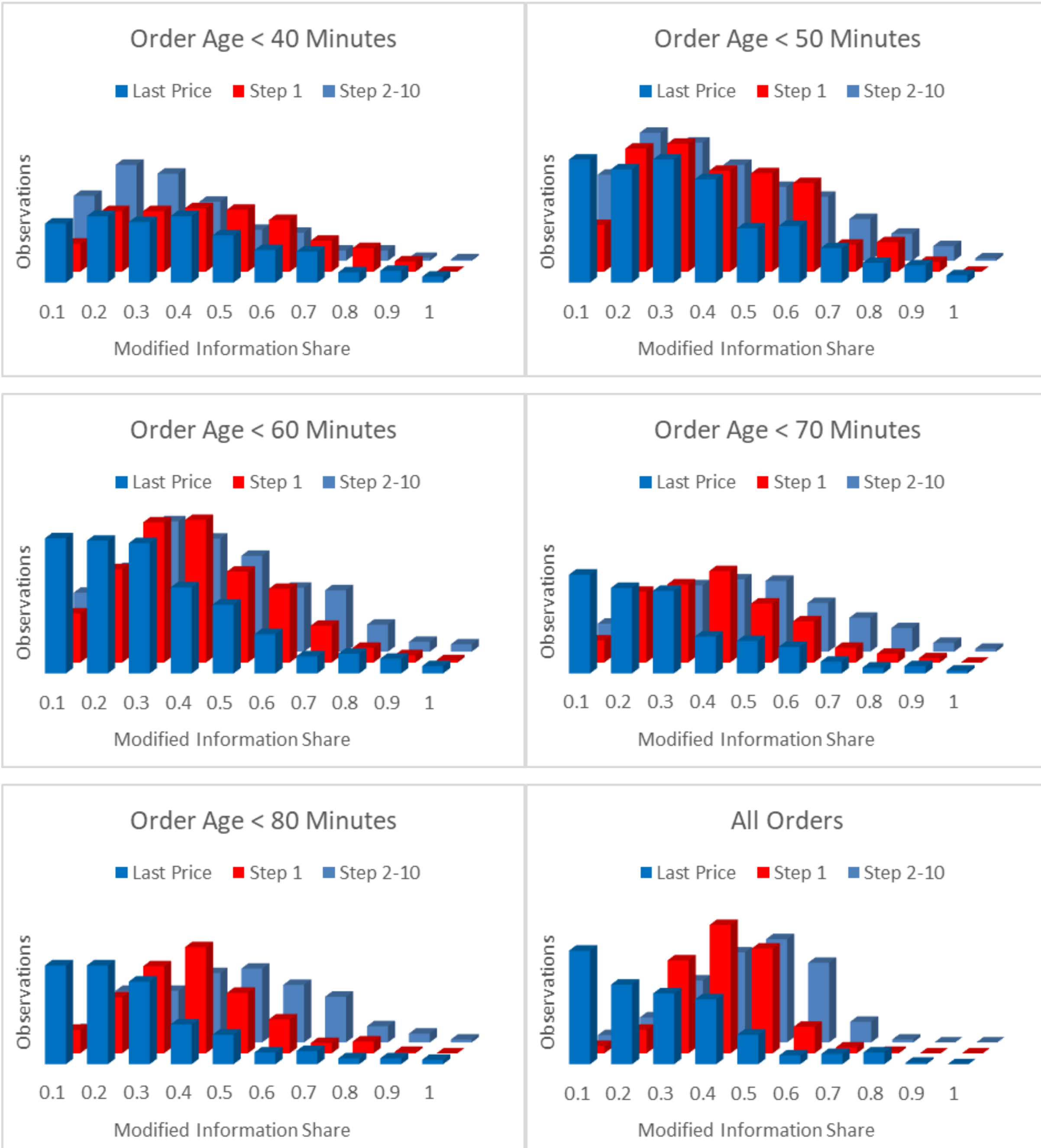
Modified Information Share (Median)			
Max Age	Last Price	Step 1	Steps 2-10
10	0.46	0.41	0.08
20	0.40	0.42	0.11
30	0.34	0.41	0.16
40	0.32	0.37	0.23
50	0.28	0.32	0.30
60	0.24	0.33	0.35
70	0.21	0.32	0.39
80	0.19	0.33	0.41
all	0.20	0.37	0.41





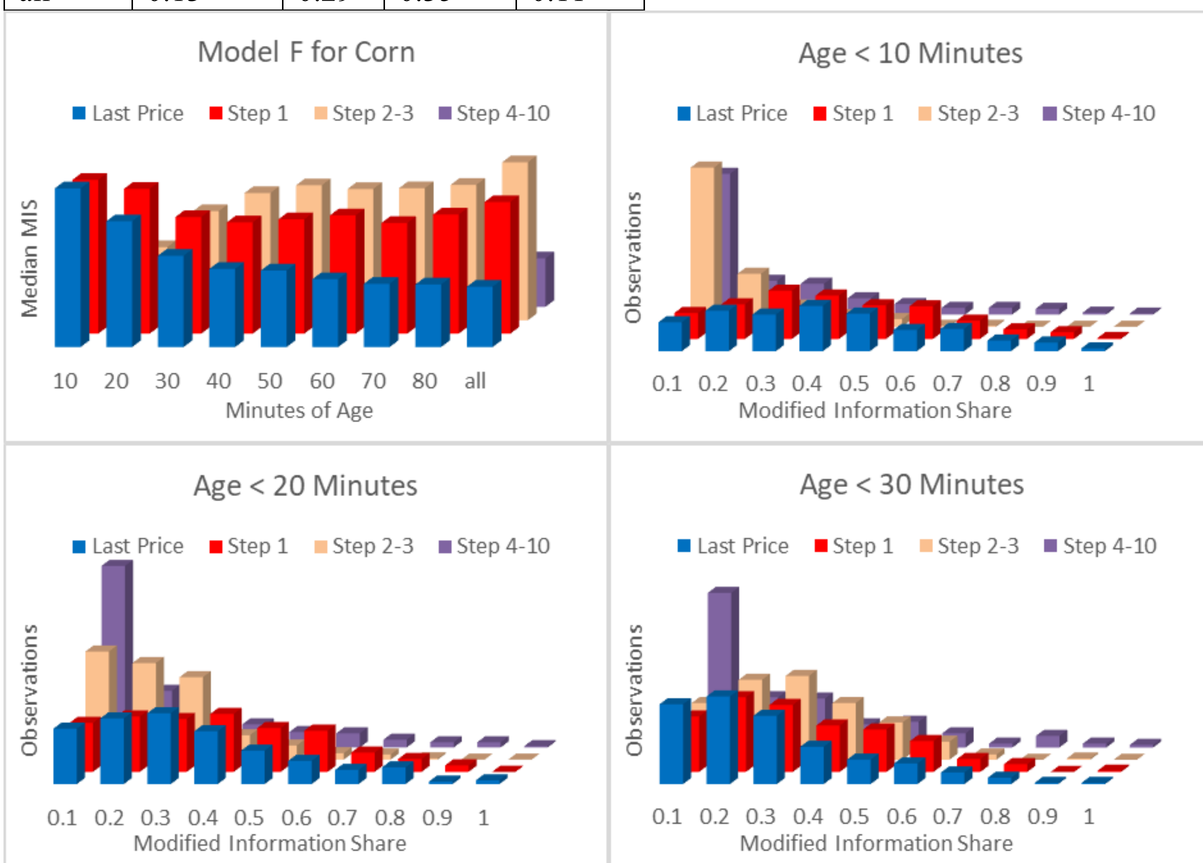


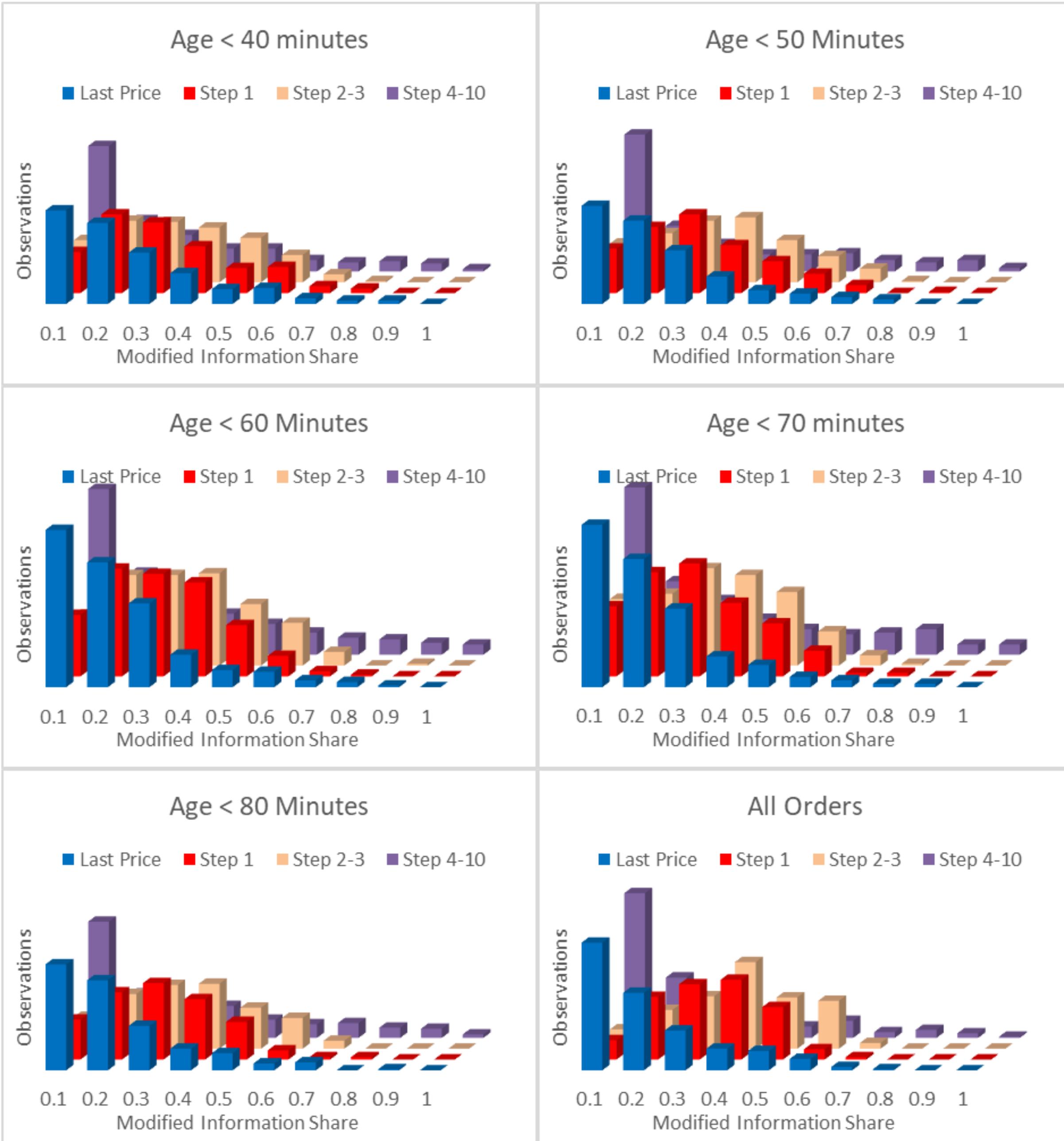



Order Age < 40 Minutes
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 50 Minutes
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 60 Minutes
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 70 Minutes
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Order Age < 80 Minutes
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
All Orders
Last Price
Step 1
Step 2-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

9.7.6 Histogram for MIS for Corn in Model F

Modified Information Share (Median)				
Max Age	Last Price	Step 1	Steps 2-3	Steps 4-10
10	0.35	0.34	0.06	0.08
20	0.28	0.32	0.16	0.09
30	0.20	0.26	0.24	0.12
40	0.17	0.25	0.28	0.14
50	0.17	0.25	0.30	0.13
60	0.15	0.26	0.29	0.16
70	0.14	0.24	0.29	0.17
80	0.14	0.26	0.30	0.17
all	0.13	0.29	0.35	0.11




Age < 40 minutes
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Age < 50 Minutes
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Age < 60 Minutes
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Age < 70 minutes
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
Age < 80 Minutes
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share
All Orders
Last Price
Step 1
Step 2-3
Step 4-10
Observations
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Modified Information Share

10. References

Arzandeh, Mehdi; Frank, Julieta. "Price Discovery in Agricultural Futures Markets." *American Journal of Agricultural Economics.* 2019, 1482–1498.

Cao, Charles; Hansch, Oliver; Wang, Xiaoxin. "The Information Content of an Open Limit-Order Book." *Journal of Futures Markets.* 2009, 16–41.

CME Group Website, visited September 2021, https://www.cmegroup.com

CME Rulebook, Rule 532, https://www.cmegroup.com/content/dam/cmegroup/rulebook/CBOT/I/5.pdf

Dickey, David; Pantula, Sastry. "Determining the Order of Differencing in Autoregressive Processes." *Journal of Business & Economic Statistics.* 2002, 18–24.

Engle, Robert; Russel, Jeffery. "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica.* 1998, 1127–1162.

Hasbrouck, Joel. "Price Discovery in High Resolution." *Journal of Financial Econometrics*. 2021, 395–430.

Hasbrouck, Joel. "Empirical Market Microstructure." Oxford University Press. 2007

Hasbrouck, Joel. "One Security Many Markets: Determining the Contributions to Price Discovery." *Journal of Finance*. 1995, 1175–1199.

Holder, Mark; Langrehr, Frederick. "A Case Study: The Chicago Board of Trade and Electronic Trading: A Case on Responding to Technological Change." *Journal of Financial Education.* 2001, 88–98.

Hu, Zhepeng; Mallory, Mindy; Serra, Teresa; Garcia, Philip. "Measuring Price Discovery Between Nearby and Deferred Contracts in Storable and Nonstorable Commodity Futures Markets." *Agricultural Economics*. 2020, 825–840.

Johansen, Søren. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. *Oxford University Press.* 1995.

Lehecka, Georg V; Wang, Xiaoyang; Garcia, Philip. “Gone in Ten Minutes: Intraday Evidence of Announcement Effects in the Electronic Corn Futures Market.” *Applied Economic Perspectives and Policy*. 2014, 504–526.

Lien, Donald; Shrestha, Keshab. “A New Information Share Measure.” *Journal of Futures Markets.* 2009, 377–395.

Lütkepohl, Helmut. “New Introduction to Multiple Time Series Analysis.” Springer. 2005.

Pascual, Roberto; Veredas, David. “What Pieces of Limit Order Book Information Matter in Explaining Order Choice by Patient and Impatient Traders?” *Quantitative Finance*. 2009, 527–545.

Phillips, Peter; Perron, Pierre. “Testing for a Unit Root in Time Series Regression.” *Biometrika*. 1988, 335-346