Price Discovery in Agricultural Futures Markets: Should We Look beyond the Best Bid-Ask Spread?

Mehdi Arzandeh and Julieta Frank

Price discovery is the incorporation of information to prices through the actions of traders. Previous studies in financial markets have found evidence that informed traders may submit limit orders instead of market orders as part of their trading strategies. If so, the steps of limit order book (LOB) beyond the best bid and best ask spread (BAS) may contain valuable information and contribute to price discovery of the underlying asset. This is the first attempt to examine the informativeness of the LOB beyond the BAS for agricultural commodities. We reconstruct the LOB using market depth data and use three information share approaches to test to what extent the steps of LOB beyond the BAS contribute to price discovery. This is done for five major agricultural commodities, namely live cattle, lean hogs, corn, wheat, and soybeans, as well as the E-mini Standard and Poor's 500 Index (S&P 500) futures contracts. The results show that the steps of the LOB beyond the BAS contribute by over 27% to price discovery of futures contracts. Across agricultural commodities, the steps of the LOB beyond the BAS have more information for grains than meats. Moreover, beyond the BAS, the steps closer to the top of the book contain more information for livestock and E-mini S&P 500. For grains, the steps farther from the BAS are as informative as the steps closer to the BAS. These findings suggest that informed traders in futures electronic markets actively use limit orders with price steps beyond the BAS.

Key words: Commodity markets, electronic trading, futures markets, information share, limit order book, market microstructure, price discovery.

JEL codes: C32, G12, G13, G14, Q11, Q13.

Agricultural commodity futures were traditionally traded in the open outcry pit, but over the past decade there has been a major shift to trading on the electronic platform. Grain and livestock futures contracts trading electronically made up less than 5% of overall trade in 2006 and grew to over 80% and 90%, respectively, in 2011 (Irwin and Sanders 2012). Today the Chicago Mercantile Exchange (CME) Group, one of the leading exchanges for the trade of agricultural commodities, has migrated its agricultural futures trading to the electronic platform. The electronic system differs significantly from the traditional open outcry system. One major difference is the presence of the limit order book (LOB) in the electronic system, which contains actual bid and ask prices and their corresponding volumes at different steps (Gould et al. 2013).

Trades in the electronic platform are conducted through a computerized system where all traders submit their orders with the number of contracts they want to trade and their intended prices. Traders can buy or sell contracts at existing market prices. If the price for which a trader intends to sell (buy) a contract is less than or equal (greater than or equal) to the price for which another trader intends to buy (sell) the contract, the trade will take place. This is also known as a market order. If, however, a trader's bid price is lower than the lowest ask price for the contract (i.e., the best ask), the order will remain active in the exchange electronic system on

Mehdi Arzandeh is an assistant professor in the Department of Business Administration, Lakehead University. Julieta Frank is an associate professor in the Department of Agribusiness and Agricultural Economics, University of Manitoba. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), the University of Manitoba Graduate Enhancement of Tri-Council Stipends (GETS) program, and WestGrid and Compute Canada. Correspondence may be sent to: Julieta.Frank@umanitoba.ca.

Amer. J. Agr. Econ. 101(5): 1482–1498; doi: 10.1093/ajae/aaz001
Published online March 26, 2019

the bid side until it is matched or deleted (or the contract expired). The bid side, thus, can be thought of as the demand side for the underlying contract. Similarly, if a trader's ask price is higher than the highest bid price (i.e., the best bid), it remains active on the ask side until it is matched or deleted (or expired). The ask side can be considered the supply side for the contract. The orders resting in the system are called limit orders and the system storing these orders is the LOB. At any point in time, the LOB contains all the resting orders on the demand and supply sides at different price steps. In the LOB, the best bid and best ask are the highest bid and the lowest ask prices, respectively, at that point in time which are referred to as "top of the book." The difference between the lowest ask and the highest bid is called "the spread" or bid-ask spread (BAS). The other bids and asks are resting in descending and ascending order beyond the best bid and best ask, respectively, in the LOB.

The information contained in the LOB has been the subject of much controversy. If informed traders use limit orders, their information is presumably reflected in the book. If, however, informed traders use market orders, the orders in the book may not contain any of their private information. Several studies on the type of orders used by informed traders and the extent to which the LOB carries information have been conducted, but the results are mixed (some examples are Harris and Panchapagesan 2005; Madhavan, Potter, and Weaver 2005; Kaniel and Liu 2006). In addition, only few of those studies analyze futures markets and none of them examine agricultural commodities.

One of the most important functions of futures markets is price discovery, which is the process of incorporating market participants' new information into market prices. Many studies have examined the contribution of related price series (such as securities trading in different markets or spot and futures prices of a commodity) to an underlying efficient price, defined as the price of the commodity in the absence of microstructure frictions. Hasbrouck's (1995) Information Share (IS) and Gonzalo and Granger's (1995) Permanent-Transitory (PT) measures have been widely used to assess the contribution of the related series to price discovery (some examples are Chu, Hsieh, and Tse 1999; Booth et al. 2002; de Jong 2002; Huang 2002; Harris, McInish, and Wood 2002). Cao, Hansch, and Wang (2009) study the information share of the steps of the LOB beyond the BAS on the underlying price. By looking at 100 active Australian stocks, these authors find that a share of about 22% of the price discovery can be attributed to the steps of the LOB beyond the best bid and best ask, whereas the remaining 78% is contributed by the best bid and ask and the last transaction price. However, the informational content of the LOB in agricultural markets may differ considerably for a variety of reasons. Markets for futures contracts are different from stock markets because many market participants trade in futures markets for the purpose of hedging and risk management. This implies that trading algorithms that are practiced in the two markets can be different. Agricultural commodity trading in futures markets can be different from trading other contracts due to differences in market characteristics such as tick size, availability of the commodity, frequency of trading, proportion of informed traders, etc. Even though much research on price discovery has been done for agricultural futures markets in the traditional outcry system, research on the electronic market at the microstructure level remains novel.

The objective of the article is to assess the informational content of the LOB beyond the BAS in agricultural futures markets. We reconstruct the full LOB and compute both the BAS at the best quotes and summarize the bids and asks at subsequent steps of the LOB beyond the best quotes. The informational content of the order book is then assessed by estimating the contribution of each of these series to price discovery. The study, specifically, is performed using nearby contracts for five major agricultural commodities, namely live cattle, lean hogs, corn, wheat, and soybeans. In order to compare the results with other actively traded futures contracts for which more research has been performed, the E-mini Standard and Poor's 500 Index (S&P 500) is also examined. Agricultural commodities are generally less traded and their market characteristics may be different from those of other products. Livestock traders have access to four steps beyond the best quotes and grain and E-mini S&P 500 traders have access to nine steps beyond the best quotes in real time. Therefore, a better understanding of the contribution of the LOB to price movements may play a fundamental role in developing their trading algorithms and strategies.

One difficulty in assessing the information content of the LOB is that Hasbrouck's IS and Gonzalo and Granger's PT measures use different approaches to estimating the contribution of price series to the common price, and there is no consensus in the literature favoring either estimate. While the PT measure is unique, it ignores the correlations between different price series (Baillie et al. 2002). On the other hand, while IS accounts for this correlation, it is not unique as it is sensitive to the ordering of price series in the model. Lien and Shrestha (2009) propose an alternative measure, the modified information share (MIS), which is unique and also independent of the ordering of the price series in the model. Here we estimate IS, PT, and MIS to assess the information content of the LOB.

Background

The evidence on the extent to which price steps beyond the BAS carry information about the efficient price is mixed. Glosten (1994), Rock (1996), and Seppi (1997) argue that informed traders favor and actively submit market orders, suggesting that the LOB beyond the best bid and ask contains little information.[1] However, Bloomfield, O'Hara, and Saar (2005) use an experimental market setting and find that in an electronic market, informed traders submit more limit orders than market orders. This suggests that key trader information is contained in the LOB. In the context of stock trading on the New York Stock Exchange (NYSE), Harris and Panchapagesan (2005) show that the imbalances in the limit buy and sell orders in the book have information regarding the short run price movements and that NYSE specialists benefit from it by buying for their own accounts when the book is heavy on the buy side, and sell when it is heavy on the sell side, especially for more active stocks. Kaniel and Liu (2006) show that informed traders prefer limit orders, and that limit orders convey more information than market orders. Baruch (2005) provides a theoretical model showing that an LOB improves liquidity and information efficiency of prices. Boehmer, Saar, and Yu (2005) find that transaction prices are closer to the efficient price after the NYSE's adoption of the LOB system. In contrast, Madhavan, Potter, and Weaver (2005) find larger spreads and higher volatility after the Toronto Stock Exchange disseminated the top four price steps of the LOB in April 1990. More recently, Biais, Foucault, and Moinas (2015) and Martinez and Roşu (2011) develop theoretical models to compare algorithmic trading with human trading in the informativeness of prices. Both studies suggest that algorithmic trading is advantageous compared to human trading due to their quicker response to new information, and that algorithmic trading uses market orders to exploit information. Hautsch and Huang (2012), on the other hand, estimate impulse response functions for thirty stocks traded on Euronext Amsterdam and find that limit orders, especially for orders posted on up to two steps beyond the market price, have a significant effect on quote adjustments. Moreover, Eisler, Bouchaud, and Kockelkoren (2012) find further support on the effect of limit orders on market prices, and Cont, Kukanov, and Stoikov (2014) argue that the order flow imbalances between supply and demand at the best bid and ask spread are the main driving force behind market price changes.

Research on agricultural markets at the microstructure level has relied on the transaction prices to study the best bid and ask. Examples of the earlier studies are Brorsen (1989), Bryant and Haigh (2004), and Hasbrouck (2004). Among the more recent studies, Frank and Garcia (2011), Shah and Brorsen (2011), and Martinez et al. (2011) used trade data to estimate the BAS in measuring the cost of liquidity and comparing open outcry to electronic trading for different agricultural commodity markets. Wang, Garcia, and Irwin (2014) reconstructed the best bid and ask steps of a limit order book to study the liquidity costs in corn futures markets. Using CME Group RLC market depth data, Aidov (2013) and Aidov and Daigler (2015) reconstruct the five-step LOB for futures contracts of the 10-Year U.S. Treasury note, corn, light sweet crude oil (WTI), Euro/U.S. dollar, yen/U.S. dollar, and gold futures to study the characteristics of market depth in electronic futures market.[2] These authors find that the steps beyond the BAS contain a large amount of depth for all

[1] Rock, K. 1996. The Specialist's Order Book and Price Anomalies. Unpublished Working Paper, Harvard University, Boston, MA.

[2] RLC market depth data was discontinued in 2009.

the futures contracts studied. Aidov (2013) concludes that market participants in the U.S. electronic futures market actively manage depth along the LOB. More recent studies on market microstructure in agriculture use the information share measures in order to determine the leading markets in the price discovery of agricultural commodities. Arnade and Hoffman (2015) find that the futures markets dominate the cash markets in the price discovery of soybean and soybean meal. Janzen and Adjemian (2017) study which one of the four futures markets of Chicago, Kansas City, Minneapolis, and Paris, France, lead the other markets in price discovery of wheat. Our study extends the previous literature on agricultural commodity electronic trading in some important ways. First, we reconstruct the LOB for four major agricultural commodities besides corn (live cattle, lean hogs, wheat, and soybeans) and for the E-mini S&P 500 futures contracts. Second, this is the first study that examines the contribution of the LOB beyond the BAS to price discovery in agricultural commodity markets.

Information Share Measures

Hasbrouck's IS and Gonzalo and Granger's PT are the two most well-known information share measures used in the majority of the literature (some examples are Anand and Subrahmanyam 2008; Chen and Gau 2010; Frijns, Gilbert, and Tourani-Rad 2010; Korczak and Phylaktis 2010; Anand et al. 2011; Fricke and Menkhoff 2011; Liu and An 2011; Chen and Chung 2012; Chen and Choi 2012; Rittler 2012; Chen, Choi, and Hong 2013). However, some weaknesses have been identified in both measures and therefore efforts have been made to generate new measures. The PT measure does not account for the contemporaneous correlation of disturbances across price series. The IS measure accounts for this correlation by using a Cholesky factorization; however, it is non-unique and sensitive to the ordering of price series in the model. Lien and Shrestha (2009) avoid the order-dependency problem by using an eigenvector factorization of the correlation matrix of residuals (instead of the Cholesky factorization) in their modified IS (MIS) new measure. These authors later extend their MIS measure for the cases where the price discovery contribution of different but related financial securities are analyzed, such as price discovery in markets for different securities issued by the same firm, and propose the Generalized Information Share, GIS (Lien and Shrestha 2014). Yan and Zivot (2010) and Putnins (2013) showed that the information shares calculated using IS and PT do not account for the different levels of noise in the price series. These authors argue that this may result in misleading measures of information share and develop a new information share metric, the Informational Leadership (IL), by combining IS and PT. The IL is, however, applicable to a two-price series setting. Another measure of price discovery was developed by Grammig and Peter (2013) to address the IS problem of non-uniqueness, particularly for longer sampling intervals. These authors assume a multivariate mixture distribution to develop the tail-dependent information shares (TLS). Like MIS, TLS follows from Hasbrouck's (1995) contribution of a price series variance to the variance of the efficient price as the measure of the series information share by means of reduced vector error correction model (VECM) long-run impact coefficients. However, the variance decomposition under TLS is performed using VECM, which is extended by the mixture parameters and estimated by a two-step process. This, unlike IS, results in an order-neutral measure, which it is claimed to be superior to IS and PT when correlations of price innovations in the tails differ from those in the center of the distributions. Lien and Wang (2016) compare the IS measure with the two unique, more recent, information shares of MIS and TLS. These authors find that TLS performs poorly for the simulated data even when the underlying assumptions of the approach are met. Moreover, their results show that MIS at most marginally improves the information share computed by the IS measure. Therefore, they support the use of the IS measure as a method of computing the information shares of different price series.

Structural Microstructure Model

The LOB contains a list of prices and the corresponding number of contracts, or volume, at which traders are willing to buy or sell a commodity. The market for each commodity is therefore characterized by a set of prices, each representing market expectations of the "true" or efficient price of the contract.

Market expectations are conditional on previous prices and on the set of prices (or quotes) for the commodity at time t (Hasbrouck 2002). A change in the efficient price is a reflection of new information in the market, where new information is conveyed through trades (Hasbrouck 1991) and orders (Hasbrouck 1996; Harris and Panchapagesan 2005). However, the efficient price is not observed because of the presence of microstructure effects such as the bid-ask bounce and price discreteness, but it is impounded in the observed prices (i.e., in the transaction price and in the bids and asks along the LOB).

The efficient price evolves over time as new information is being incorporated. Since new information cannot be predicted, the efficient price is said to follow a random walk. Following Hasbrouck's (2002) multiple price microstructure model, we can decompose the observed prices into the random walk component (m_t) and a term containing various microstructure effects ($\mathbf{s}_t$):

$$(1) \quad m_t = m_{t-1} + u_t$$

$$(2) \quad \mathbf{p}_t = \mathbf{1}m_t + \mathbf{s}_t$$

where u_t are *iid* $(0, \sigma_u^2)$ innovations originating from the arrival of new information, $\mathbf{p}_t = (p_{1t}, p_{2t},\ldots, p_{Kt})'$ are the observed prices, $\mathbf{1}_{Kx1} = (1,\ldots, 1)'$, and $\mathbf{s}_t = (s_{1t}, s_{2t},\ldots, s_{Kt})'$ represents zero-mean covariance stationary processes originating from microstructure effects. Note that $\mathbf{s}_t$ represents a temporary component of the current observed prices and therefore should not affect the price of the commodity in the long run. In equation (2), only $\mathbf{p}_t$ is known and therefore m_t and $\mathbf{s}_t$ cannot be identified. However, using the properties of the observed prices, it is possible to identify the variance of the efficient price, σ_u^2, and to measure the contribution of each observed price innovation to that of the efficient price.

Reduced-form Model and Measures of Price Discovery

Given the random walk component described in equation (1), the observed prices in $\mathbf{p}_t$ are assumed to be integrated. Further, the observed prices are linked in the long run as the expectations for each price refers to the future price of the same underlying commodity. These prices can be each non-stationary, but they move as a group. Prices may diverge in the short run due to microstructure trading frictions but in the long run they must converge due to arbitrage and/or equilibrium relationships (Hasbrouck 2002). Integrated prices that are linked in the long run are said to be cointegrated. The VECM captures the short term deviations of prices from the equilibrium relation and the cointegrating relationship among the set of prices. It also provides a convenient framework to estimate the contribution of each price to the innovation of the implicit efficient price by means of innovation variance decomposition.

The VECM has been used in previous studies to represent the dynamic relationship of prices and depth in the LOB of a single market. Hautsch and Huang (2012) estimate a VECM to assess the price effect of posting a limit order in the LOB for 30 stocks trading on Euronext Amsterdam. The vector of endogenous variables include the best bid and ask, and the depth for each side (buy and sell) and step (first to fifth) of the LOB. Lo and Hall (2015) estimate a VECM to study the liquidity replenishment mechanism of the LOB for 30 stocks trading on the Australian Securities Exchange (ASX). In the model, these authors incorporate the best bid and ask, and their corresponding depth, as well as the depth from the second to the fifth steps of the LOB. Cao, Hansch, and Wang (2009) estimate a VECM to examine the information content of the LOB in ASX. In these authors' model, the endogenous variables are the transaction price, the midpoint between the best bid and ask, and the quantity-weighted prices beyond the best bid and ask observed in the LOB. Both Euronext and ASX are pure order-driven markets where prices and quantities desired to be bought or sold are posted in the LOB, with no dealers or other designated liquidity providers. We focus on the information contained in the set of prices available for each commodity as our objective is to identify the behavior of the implicit efficient price in the observed prices. Our approach is therefore similar to Cao, Hansch, and Wang (2009).

Let $\mathbf{p}_t$ be the vector of K endogenous prices of a particular commodity in the same market. A general error correction model in matrix notation for these price series can be written as

$$(3) \quad \Delta \mathbf{p}_t = \boldsymbol{\alpha}\boldsymbol{\beta}^{'}\mathbf{p}_{t-1} + \sum_{l=1}^{q} \boldsymbol{A}_l \Delta \mathbf{p}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\alpha}$ denotes the $K \times r$ loading matrix or the matrix of coefficients that reflect how quickly price series return to their long run equilibrium, $\boldsymbol{A}$ is the $K \times K$ matrix of coefficients of own and other prices previous changes that represent the short term adjustment process, and $\boldsymbol{\varepsilon}_\mathrm{t}$ is a white noise error term with variance-covariance matrix $\boldsymbol{\Omega}$.

The first term contains the cointegrating relations and it represents the long run equilibrium between the set of prices for the underlying commodity. In equation (3), $\boldsymbol{\beta}$ represents the coefficients of this cointegration process, or the (non-unique) $K \times r$ cointegrating matrix. Hasbrouck (1995) suggests the following form: $\boldsymbol{\beta}' = [\boldsymbol{\iota} : -\boldsymbol{I}]$, where $\boldsymbol{\iota}$ is a $(K-1) \times 1$ vector of 1's and $\boldsymbol{I}$ is the $(K-1)$ identity matrix. It is assumed that the price series are integrated of order one (I(1)) and that the system consists of a single common stochastic trend (Stock and Watson 1988), that is, the system has $r = K - 1$ cointegrating vectors.

The VECM in equation (3) can be characterized by the vector moving average (VMA) representation $\Delta \mathbf{p}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\Psi}(L) = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \ldots$ and $\boldsymbol{\Psi}_{\boldsymbol{i}}$ are matrices of coefficients (Hasbrouck 1995). Alternatively, this equation can be written as follows, which is known as a Beveridge-Nelson decomposition (Beveridge and Nelson 1981):

$$(4) \quad \mathbf{p}_t = \mathbf{p}_0 + \boldsymbol{\Psi}(1)\sum_{i=1}^{t} \varepsilon_i + \boldsymbol{\Psi}^{*}(\boldsymbol{L})\boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\Psi}(1)$ is the $K \times K$ impact matrix in the lag operator, L, or the sum of the moving average coefficients. Therefore, $\boldsymbol{\Psi}(1)\varepsilon_\mathrm{t}$ is the long run impact of an innovation, ε_t, in a price on each of the prices that are due to new information. The long run impact on all prices is the same and thus $\boldsymbol{\Psi}(1)$ has identical rows (a reflection of the specific structure of the cointegrating vector). We denote the common row in $\boldsymbol{\Psi}(1)$ by $\psi = (\psi_1,\ \psi_2, \ldots, \psi_K)$. The $K \times K$ matrix $\boldsymbol{\Psi}^{*}(\boldsymbol{L})$, which is also in the lag operator, L, is the part of the price change that is resulted from transitory shocks of bid-ask bounces, inventory adjustments, or other market imperfections.

The assumption of $r = K - 1$ implies that the impact matrix $(\boldsymbol{\Psi}(1))$ has a rank of one. Therefore, from the Engle-Granger representation theorem (Engle and Granger, 1987), it follows that $\boldsymbol{\beta}^{'}\boldsymbol{\Psi}(1) = 0$ and $\boldsymbol{\Psi}(1)\boldsymbol{\alpha} = 0$, which results in a common row in $\boldsymbol{\Psi}(1)$ or that the long run impact of ε_t on each price is identical. Following de Jong (2002), equation (4) can be rewritten as

$$(5) \quad \mathbf{p}_t = \mathbf{p}_0 + \boldsymbol{\beta}_{\perp}\boldsymbol{\alpha}_{\perp}^{'}\sum_{i=1}^{t} \varepsilon_i + \boldsymbol{\Psi}^{*}(\mathrm{L})\boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\beta}_{\perp}$ and $\boldsymbol{\alpha}_{\perp}$ are orthogonal to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively, that is, $\boldsymbol{\beta}^{'}\boldsymbol{\beta}_{\perp} = 0$ and $\boldsymbol{\alpha}^{'}\boldsymbol{\alpha}_{\perp} = 0$ are satisfied. Equation (5) is closely related to how Stock and Watson (1988) represent the common trend. That is, prices have a nonstationary common factor with a permanent effect $(\boldsymbol{f}_t)$ and a stationary transient component $(\boldsymbol{G}_t)$ given by

$$(6) \quad \mathbf{p}_t = \boldsymbol{f}_t + \boldsymbol{G}_t.$$

The common trend representation in Stock and Watson (1988) represented in equation (6) and the Beveridge-Nelson decomposition of equation (4) are the basis for the information share measures that follow.

Gonzalo and Granger Permanent-Transitory Effect (PT)

Gonzalo and Granger (1995) suggest that each of the prices in a system potentially contributes to the common trend or the efficient price. Therefore, the common factor is defined as a combination of prices given by $\boldsymbol{f}_t = \boldsymbol{\Gamma}\mathbf{p}_t$, where $\boldsymbol{\Gamma}$ is a $1 \times K$ vector of coefficients for the common factor with elements $(\gamma_1,\ \gamma_2, \ldots, \gamma_K)$; they show that $\boldsymbol{\Gamma}$ is orthogonal to the matrix of the error correction coefficients $\boldsymbol{\alpha}$ and the common trend representation, and therefore can be written as $\boldsymbol{f}_t = \boldsymbol{\alpha}_{\perp}^{'}\boldsymbol{p}_\mathrm{t}$. Finally, the PT measure of the contribution of the jth price to the efficient price is related to γ_j in $\boldsymbol{\Gamma}$ or $\alpha_{\perp j}$ in $\boldsymbol{\alpha}_{\perp}$. That is, the PT information share only depends on γ or $\boldsymbol{\alpha}_{\perp}$. Harris, McInish, and Wood (2002) normalize the vector coefficients of the common trend such that the sum of the price information shares equals one. Based on Harris, McInish, and Wood (2002), PT can be computed as $PT_j = \frac{\alpha_{\perp j}}{\sum_{j=1}^{K} \alpha_{\perp j}}$.

Hasbrouck Information Share (IS)

Hasbrouck (1995) also uses the common factor representation by Stock and Watson

(1988) to develop an information share in order to measure the contribution of different market prices to the efficient price. The difference between IS and the previous approach is that in IS the variance of the common factor is decomposed and each price contributes to the efficient price based on how its variance contributes to the variance of the efficient price. The variance of the common factor innovations is given by $var(\psi \varepsilon_t) = \psi \mathbf{\Omega} \psi'$, where ψ is a common row vector in the $\mathbf{\Psi}(1)$ matrix in equation (4). We compute the parameters in $\mathbf{\Psi}(1)$ directly using Johansen's factorization and the estimated coefficients from the VECM in equation (3) by $\mathbf{\Psi}(1) = \boldsymbol{\beta}_{\perp} \mathbf{\Pi} \boldsymbol{\alpha}'_{\perp}$, where $\mathbf{\Pi} = \left(\boldsymbol{\alpha}'_{\perp} (\boldsymbol{I} - \sum_{l=1}^{q} \boldsymbol{A}_l) \boldsymbol{\beta}_{\perp} \right)^{-1}$ and $\boldsymbol{I}$ is the K identity matrix (Baillie et al. 2002). The IS of the jth price, then, can be calculated by $IS_j = \frac{(\psi_j \sigma_j)^2}{\psi \mathbf{\Omega} \psi'}$, where σ_j is the jth price's standard deviation in the variance-covariance matrix $\mathbf{\Omega}$.

Hasbrouck (1995) suggests that if the variance-covariance matrix of residuals ($\mathbf{\Omega}$) in the VECM representation (equation 3) is not diagonal, that is, the price innovations are significantly correlated across the price series, IS and PT can result in misleading information shares. Hasbrouck (1995) uses the Cholesky factorization of the residual's covariance matrix to eliminate the contemporaneous correlation. Based on the Cholesky factorization, $\mathbf{\Omega} = \boldsymbol{M}\boldsymbol{M}'$, where $\boldsymbol{M}$ is a lower triangular matrix. The IS measure can then be written as

$$(7) \quad IS_j = \frac{[\psi \boldsymbol{M}]_j^2}{\psi \mathbf{\Omega} \psi'}.$$

Even though the IS calculated using equation (7) solves the correlation problem, it creates another issue, that is, the measure being sensitive to the ordering of prices in the system. This occurs because when correlation exists, that is, the nondiagonal elements of M are nonzero, the IS measure imposes more weights on the prices that appear earlier in the system. To overcome this problem, Hasbrouck (1995) proposes calculating upper and lower bounds for each price by placing them first and last in the system. In the multivariate cases, all permutations of the variables must be computed to find the upper and lower bounds (Hasbrouck 2002).[3]

Baillie et al. (2002), de Jong (2002), and Yan and Zivot (2010) show that the PT measure can be computed by

$$(8) \quad PT_j = \frac{\psi_j}{\sum_{j=1}^{k} \psi_j}.$$

Therefore, we use the coefficients of the long run impact matrix, $\mathbf{\Psi}(1)$, to measure IS and PT in equations (7) and (8), respectively.

Modified Information Share (MIS)

The modified information share (MIS) was developed by Lien and Shrestha (2009) to address the problems with the previous approaches, that is, a potential nondiagonal variance-covariance matrix of residuals in PT and the order sensitivity in IS. In MIS, on the one hand, a type of factorization is used to eliminate potential contemporaneous correlations. On the other hand, instead of the variance-covariance matrix $\mathbf{\Omega}$, which is sensitive to the ordering of price series, MIS uses the innovation correlation matrix, which is invariant to different orders of price series in the system. The factorization procedure is as follows. Consider the innovation correlation matrix denoted by $\mathbf{\Phi}$ of the VECM in equation (3). Consider also a diagonal matrix $\mathbf{\Lambda}$, where its elements are the eigenvalues of the correlation matrix ($\mathbf{\Phi}$), and the corresponding eigenvectors are given by the columns of matrix $\boldsymbol{G}$. Finally, suppose matrix $\mathbf{V}$ is diagonal and its elements are the standard deviation of the innovations or the square root of diagonal elements of $\mathbf{\Omega}$. Lien and Shrestha (2009) show that the innovations can be transferred to $\varepsilon_t = \boldsymbol{F} z_t$, where z_t is the transferred innovation with zero mean, $E[z_t] = 0$, and identity covariance matrix, $E[z_t z_t'] = \boldsymbol{I}$. Moreover, $\boldsymbol{F} = \left[\boldsymbol{G} \Lambda^{-1/2} \boldsymbol{G}' \boldsymbol{V}^{-1} \right]^{-1}$ and $\mathbf{\Omega} = \boldsymbol{F}\boldsymbol{F}'$. This factorization results in the MIS measure given by

$$(9) \quad MIS_j = \frac{[\psi \boldsymbol{F}]_j^2}{[\psi \boldsymbol{F}] \mathbf{\Omega} [\psi \boldsymbol{F}]'}.$$

MIS has the desirable feature that it calculates an information share which is order

[3] Baillie et al. (2002) show how PT and IS can result in similar information shares in the absence of contemporaneous correlation in the residuals and under the assumption that there exists only one single common factor in the system.

invariant and it accounts for contemporaneous correlations in innovations. This is particularly important when correlation exists in the residuals. When correlation is nearly nonexistent, MIS approaches IS and PT, and when innovations are completely correlated, MIS approaches $1/K$ or an equal share for all the K price series.

We use a Lien and Shrestha (2009) MIS metric to determine the contribution of the BAS and the LOB beyond the BAS to price discovery of the six markets under study. We also measure and report the two common information share metrics, that is, a Hasbrouck (1995) IS and a Gonzalo and Granger (1995) PT. An extended survey on the developments in information share research can be found in Narayan and Smyth (2015).

Data

We use the *market depth* data files from the CME Group, which provide every incremental book update required to reconstruct the LOB with nanosecond precision.[4] Data are available to reconstruct a five-step-deep book for live cattle and lean hogs and a ten-step-deep book for corn, wheat, soybeans, and E-mini S&P 500. We estimate the information contained in the LOB for the period November 23, 2015 to March 31, 2016. Live cattle and lean hogs futures contracts trade in one session, from 8:30 am to 1:05 pm CT. Corn, wheat, and soybeans futures contracts trade in two sessions, 8:30 am to 1:20 pm CT (morning session) and 7:00 pm to 7:45 am CT (evening session). We use data from the morning session from Monday to Friday only due to the low volume traded in the evening session and on Sunday. For the E-mini S&P 500 futures contract we use the most active daily trading hours, from 7:30 am to 3:15 pm CT, from Monday to Friday.[5] There is no trading on the CME Group on Saturdays and the two federal holidays of Jan. 18 and Feb. 15 in our sample. We also remove the data for Sundays and the days for which there are extended trading halts. The latter is mostly the case for a few futures contracts with partial pre-holiday (a day prior) and post-holiday (a day after) trading with extended trading breaks.

Roll Dates

Traders in the CME Group futures roll their futures positions from one futures contract month to the next before the futures contracts are very close to termination and thus becoming very illiquid. Traditionally, traders in the CME Group roll forward expiring futures contracts eight calendar days before the contract expiry (i.e., the "roll date"). The eight calendar day roll period seems to be a good roll date for the E-mini S&P 500. However, the period proves to be too short for the agricultural commodity futures contracts. We define a roll date for the current contract as the date when its aggregate volume traded for two consecutive days falls below that of the second nearest contract.[6] Based on this rule, traders roll their positions about 42 and 22 calendar days before the expiration of the futures contracts for live cattle and lean hogs, respectively. For corn wheat and soybeans, the roll date is 21, 24, and 22 calendar days before the expiration, respectively.

The Limit Order Book

The *market depth* data are formatted using the Financial Information eXchange (FIX/BINARY) protocol, which comprises a series of messages containing information such as bids and asks with their corresponding quantity and step in the LOB, trade prices and quantities, order sending time, and changes in the LOB such as order deletions and bids, asks, and quantities updates that would define a new book. Each message is processed to reconstruct the LOB as follows. If a message contains information on a new market order, then there is an immediate match and a trade takes place.[7] If the trade results in a partial matching of the best bid or ask, the LOB

[4] Access to the data for replicating the analysis in this article is available from the authors upon request.

[5] These hours correspond to 8:30am to 4:15pm Eastern Time. There is a 15-minute trading halt from Monday to Friday at 3:15 pm to 3:30 pm Central Time.

[6] We also considered alternative rules where we examined the daily number of trade price changes and the daily average duration of price changes to determine the roll dates. These rules almost always result in the same roll dates as the case of the aggregate volume rule. In the isolated exceptions, the roll date is one day before the roll date defined by the aggregate volume rule.

[7] Futures trading in the CME Group follows a price-time priority system, that is, orders matching the best bid or ask prices are executed first. If two orders in the LOB have the same bid or ask prices, priority is given to the order that arrived first.

remains the same except for the change in the number of contracts at the top of the book. On the other hand, if the trade results in a full matching of the best bid or ask, all price steps beyond the best bid or ask move one step towards the top of the book and the spread widens. If a message contains information on a new limit order with a better price than the best bid or ask, that is, inside the spread, the top of the book changes and the new price becomes the best bid or ask price. If the spread or the difference between any two steps on either the buy or sell side of the book is greater than one tick (the minimum change in price allowed), traders can gain priority by submitting an order inside the spread or between two existing steps. In this case, the spread narrows and the remaining prices on the same side move one step further down along the LOB. An order can be deleted, which also updates the LOB. If it is a partial deletion, the prices in the LOB remain the same and only the corresponding quantities are altered.[8] However, if the entire quantity on a price step is deleted, the succeeding price steps move one step upward in the LOB.[9]

The CME Group supports implied functionality, which is the ability to combine spread and outright markets in one order book with the objective of increasing liquidity.[10] An accurate picture of the LOB for futures contracts for a market with implied functionality at any point in time, therefore, is the one which comes from the consolidated limit order book (CLOB) that accounts for both the outright book and the implied limit order book (ILOB). The ILOB is reconstructed using data from the *market depth* files in the same way as described above for the outright book. Data are available to reconstruct a two-step-deep implied book for all six futures markets. The outright and the implied books are then merged into a CLOB as follows. If the price steps in the ILOB are the same as those in the LOB, the implied quantities are added to the LOB's corresponding price steps to get the CLOB. If prices are different, however, price steps coming from LOB and ILOB are compared and sorted for each bid (descending) and ask (ascending) side to form the CLOB. Even though we reconstruct and employ the consolidated CLOB, we refer to it as LOB hereafter for simplicity.

Price Duration

Limit and market orders that continuously update the LOB inherently arrive in an irregular, timely manner. However, regularly spaced data is needed for our underlying econometric models. Previous studies suggest taking snapshots of the LOB at regular times. For example, Hasbrouck (1995) and Cao, Hansch, and Wang (2009) both use one-second snapshot data for 30 Dow stocks and 100 of the most active Australian stocks, respectively. The time duration between snapshots is important because if it is too long, important information might be overlooked and if it is too short, we might create a data set with a lot of observations that are repeated with no new information and cause other problems such as heteroskedasticity (Engle and Russell 1998). The literature is not clear on how to select an optimal duration for time intervals. We use the average duration of transaction price changes. Following Engle and Russell (1998), we denote every trade price change as a price event and define a duration variable, d_i, given by $d_i = t_i - t_{i-1}$, where t_i is the time of the ith transaction. We construct regularly spaced time series of the LOB for each product on the basis of how frequently their transaction price changes during the period of study. Summary statistics of the daily average durations are presented in supplementary online appendix table A1. As it can be seen, the price duration of one-second snapshots used in the finance literature (e.g., Hasbrouck 1995 and Cao, Hansch, and Wang 2009), is a good approximation for a product such as the E-mini S&P 500, which is highly frequently traded. However, for agricultural commodities, the trading frequency and price fluctuation is considerably lower and therefore we select a longer snapshot duration to avoid a high number of repeated observations. The price duration also varies across agricultural commodities. Therefore, we choose the durations based on price

[8] Traders operating in the CME Group have the option of submitting iceberg or hidden-size orders, which are limit orders that specify a "visible" portion of the order size. Once that quantity is filled the remaining portion of the order size is revealed. This might result in an underestimation of the information contained in the LOB when the proportion of iceberg orders is high.

[9] When a hidden order is placed at the BAS and the displayed quantity has been filled, another portion of the order will be displayed to the marketplace and, therefore, the spread remains unchanged. In addition, traders can gain priority by submitting limit orders inside the BAS. In very active markets such as the E-mini S&P500, the BAS would normally remain at the tick, or the minimum allowable price change.

[10] An implied order is a futures order generated based on the spread of orders in the outright market, the spread market, or other implied orders.

events. The average durations are 7.40, 11.97, 8.63, 11.94, 7.60, and 1.12 seconds for live cattle, lean hogs, corn, wheat, soybeans, and E-mini S&P 500, respectively.

Summary Statistics of the LOB

The number of observations for each product reflects the frequency of transaction price updates. Table 1 and table A2 in the supplementary online appendix show that during the study's time period, the LOB for E-mini S&P 500 futures contracts have considerably more price events. Across agricultural commodities, soybeans has the shortest average price duration. The number of price events is more or less similar for corn and live cattle, and for wheat and lean hogs. Among all products, the average volume traded for corn is the highest. On average, the BAS is about 0.046 cents (1.8 ticks) for live cattle, 0.040 cents (1.6 ticks) for lean hogs, 0.27 cents (1.1 ticks) for grains, and 25.6 cents (1 tick) for E-mini S&P 500.[11] Table A2 in the supplementary online appendix shows that along the LOB and up to the fourth step, corn has a considerably higher depth than the rest of the products on average—even higher than E-mini S&P 500. After the fourth step, E-mini S&P 500 has a higher depth than corn. Overall, the first two steps beyond the BAS seem to have a significantly higher depth than the remaining steps for agricultural commodities. Together with more or less equal price differences for bids and asks along the book, this implies that the two steps closer to the top of the book are relatively "denser" for agricultural commodities. For E-mini S&P 500, surprisingly, further steps appear to have a slightly higher depth than the steps close to the top of the book. This means that traders of agricultural futures contracts submit more limit orders at the steps closer to the top of the book, whereas traders of the E-mini S&P 500 futures contracts prefer the steps further from the top of the book. Therefore, in addition to studying differences of the LOB between agricultural commodities and E-mini S&P 500 at the aggregate level, differences of the LOB at the step level can shed light on how trading in agricultural commodity markets differs from other markets.

[11] The BAS values are calculated as the lowest ask price, less the highest bid price for each data point. The average BAS reported are the arithmetic mean of the BAS values.

Price Variables in the LOB

An LOB consists of different price steps and the associated number of futures contracts, at any point in time. The relationship between the bid price steps and the number of contracts, related to each bid price step and aggregated across all orders, can be thought of as a market demand step function. Similarly, a market supply step function derives from the relationship between the ask price steps and the related aggregate contracts. The following weighted price reflects the price and quantity aspects of an LOB at a given point in time:

$$WP^{n_1-n_2} = \frac{\sum_{s=n_1}^{n_2}\left(Q_s^b P_s^b + Q_s^a P_s^a\right)}{\sum_{s=n_1}^{n_2}\left(Q_s^b + Q_s^a\right)}, \quad n_1 \le n_2 \tag{10}$$

where $WP^{n_1-n_2}$ is the weighted price of step n_1 to step n_2. This summarizes all the information contained in the LOB from step n_1 to n_2. Moreover, Q and P are quantity and price of the demand side (denoted b) or the supply side (denoted a), respectively. When $n_1 = n_2 = 1$, the weighted price becomes $WP^1 = \frac{Q_1^b P_1^b + Q_1^a P_1^a}{Q_1^b + Q_1^a}$. Cao, Hansch, and Wang (2009) use *MID*, which is the arithmetic mean of the best bid and best ask to capture the information of the spread. *MID* only changes when the best bid or the best ask change, whereas WP^1 also changes as a result of a change in the quantities at the best bid or ask. Vo (2007) studies the quantity at the best bid and ask prices and its relationship with the BAS for Toronto Stock Exchange stocks, while Frino, Lepone, and Wearin (2008) examine the relationship for three interest rates futures contracts on the Sydney Futures Exchange. The results of both studies show a negative relationship between the two variables, which implies that market participants manage both price and quantity as a part of their trading strategies. Thus, we use WP^1 to capture the information contained in the spread.[12]

[12] The variables representing a summary of the LOB can be also constructed in a way that they reflect the price-quantity characteristics of the bid side and ask side separately. This allows us to examine the possible asymmetries in the informativeness of the bid and ask sides. To assess this, we constructed a summary variable only for the bid side of the LOB as $WP_b^{n_1-n_2} = \frac{\sum_{s=n_1}^{n_2} Q_s^b P_s^b}{\sum_{s=n_1}^{n_2} Q_s^b}$, $n_1 \le n_2$. Analogously, the variable for the ask side was constructed using

Table 1. Summary Statistics of the Snapshot LOB

	Live cattle	Lean hogs	Corn	Wheat	Soybeans	E-mini S&P 500
Observations[a]	176,142	116,082	175,463	126,789	201,377	2,263,729
Price[b]	132.498	66.283	365.90	471.69	879.08	197,931.20
Volume[c]	1.74	1.62	4.93	2.92	3.40	4.42
Tick	0.025	0.025	0.25	0.25	0.25	25
Mean Height buy	0.026	0.025	0.24	0.24	0.24	24.04
Mean Height sell	0.025	0.024	0.24	0.24	0.24	24.01
Mean Quant. buy	9.08	9.06	433.43	68.39	97.33	366.32
Mean Quant. sell	8.54	10.13	404.40	63.36	100.03	371.87

Note: Mean Height and Mean Quant. are, respectively, the mean heights and the mean number of contracts for buy and sell sides in different steps of the LOB across all reconstructed LOBs in the sample period. [a]Total number of snapshots for the nearby contracts. [b]Average transaction price for the nearby contracts. Prices are in cents per pound for livestock, cents per bushel for grains, and cents per contract for E-mini S&P 500. Bid and ask units are the same as those for the transaction prices. [c]Average volume per transaction for the nearby contracts. Volume and Quantities values are expressed in number of contracts.

Figure 1, table 1, and supplementary online appendix table A2 provide the summary statistics of the shape of the average LOB for each of the six markets over the studied time period. For the agricultural commodities, the number of contracts resting on the second and third steps of the book on both buy and sell sides, are considerably higher than those on the steps beyond the third step. To study this further we develop two variables, one for the first two steps and one for the remaining steps of the LOB. However, this is not the case for the E-mini S&P 500 for which the contracts appear to spread more or less equally over the second to tenth steps of the book. Moreover, in contrast to what Cao, Hansch, and Wang (2009) observe in the Australian stocks that the heights (see figure 1 legend) tend to be shorter for the steps close to the top of the book than the steps that are further away, the heights in our dataset are almost equal across all steps for all the products. This is also illustrated in figure 1.

Results

We estimated two VEC models. The first one contains three variables, where $\boldsymbol{p}_t = (Price_t, WP^1_t, WP^{2-5}_t)'$ for live cattle and lean hogs, and $\boldsymbol{p}_t = (Price_t, WP^1_t, WP^{2-10}_t)'$ for the other markets. The second model contains four variables and is based on the observed shape of the average LOB. In all cases, $Price_t$ stands for transaction price. We break down the steps of the LOB beyond the BAS into two variables, one with steps closer to the top of the book and one farther, to examine whether the information of the steps beyond the BAS is uniformly distributed. For E-mini S&P 500, the contracts spread more or less equally across all steps. Therefore, we construct two variables for the agricultural commodities (one for the second and third steps and one for the remaining steps) and two variables for the E-mini S&P 500 (one for the second to fifth steps and one for the remaining steps). That is, in the four-variable model $\boldsymbol{p}_t = (Price_t, WP^1_t, WP^{2-3}_t, WP^{4-5}_t)'$ for live cattle and lean hogs, $\boldsymbol{p}_t = (Price_t, WP^1_t, WP^{2-3}_t, WP^{4-10}_t)'$ for corn, wheat, and soybeans, and $\boldsymbol{p}_t = (Price_t, WP^1_t, WP^{2-5}_t, WP^{6-10}_t)'$ for the E-mini S&P 500.

The VEC model in equation (3) is estimated for each day and the average information share computed using equations (7), (8), and (9) over all days is reported for each product in table 2 and table 3.[13] Moreover, all models are estimated using 80 lags for each price series.[14] The Lagrange Multiplier (LM) test and autocorrelation functions showed no

$WP^{n_1-n_2}_a = \frac{\sum_{s=n_1}^{n_2} Q^a_s P^a_s}{\sum_{s=n_1}^{n_2} Q^a_s}$, $n_1 \leq n_2$. The results showed no considerable asymmetry between the information shares of the two sides of the book and therefore are not presented, but are available from the authors upon request.

[13] Unit root at levels and first differences for each price series is tested using the Augmented Dickey–Fuller unit root test and all price series are found to be I(1) for most days. The cointegration rank is tested using three statistics, namely Johansen's trace statistic at the 99% confidence level, minimizing the Schwarz Bayesian information criterion (BIC), and minimizing the Hannan and Quinn information criterion (HQIC; Gonzalo and Pitarakis 1998; Aznar and Salvador 2002). The rank is found to be 2 for the three-price series model and 3 for the four-price series model for almost all days.

[14] The number of lags is usually determined by minimizing a criterion such as AIC or BIC. These tests were performed using a maximum number of 80 lags. Using 80 lags results in minimum AIC and BIC, and the ACF of the residuals show no significant autocorrelation. Thus, 80 lags were added to the model. Also, 80 lags cover over ten minutes of time for the agricultural commodities datasets used here given that our time intervals are greater than 7 seconds for all agricultural commodities. Moreover, Lehecka, Wang, and Garcia (2014) show that announcement effects of major USDA reports on corn futures prices and volume disappear in ten minutes, using intraday Chicago Board of Trade data for July 2009 to May 2012.

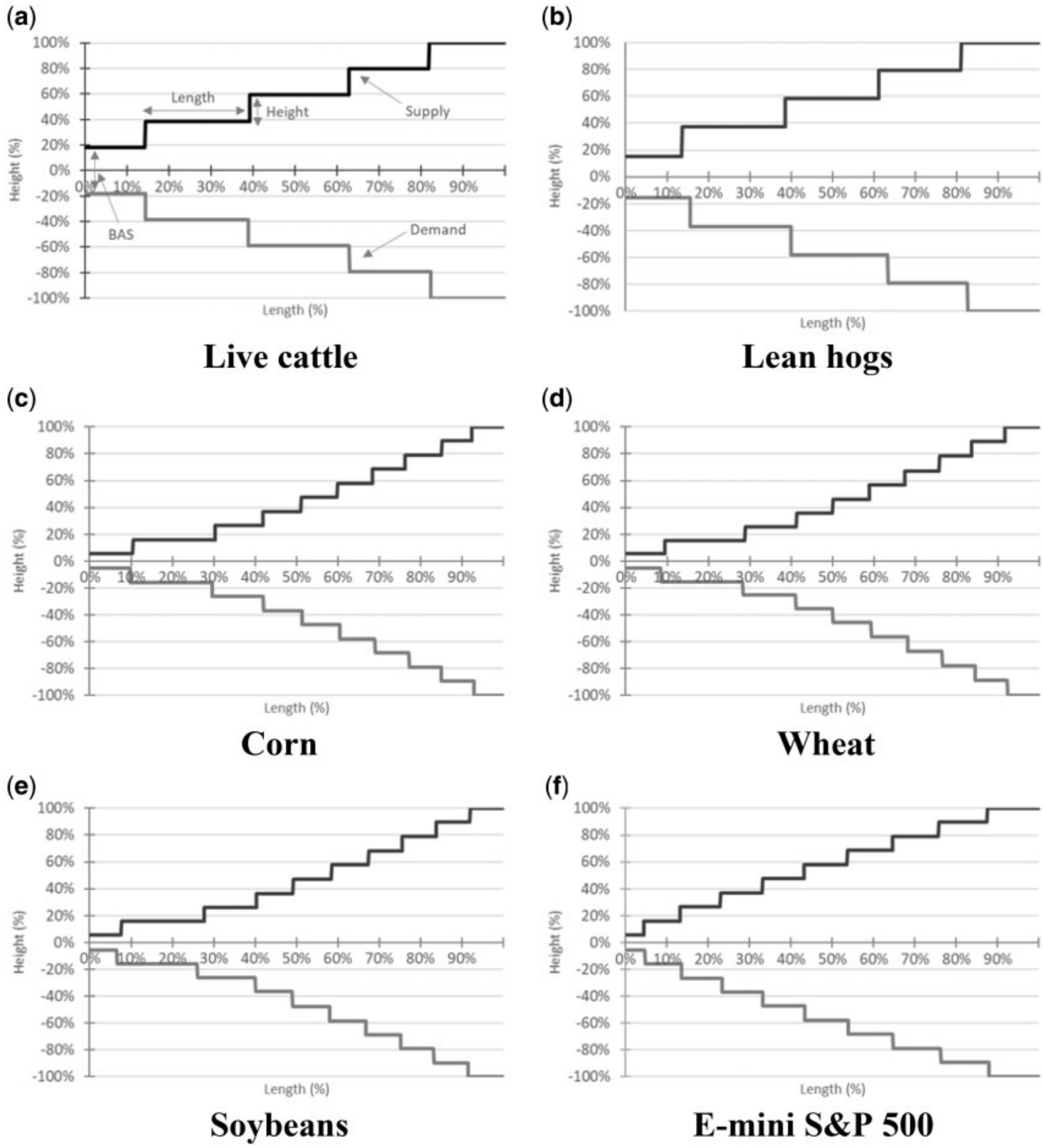

Figure 1. Average LOB

Note: The height of a step ***i*** in the step functions is the difference between price ***i*** and price ***i*** – 1. For instance, the height of step 4 on the demand side is the fourth-best bid, less the third-best bid. The length of a step ***i*** is the summation of the contracts across all orders for price ***i*** on each demand or supply side. The mean of the best bid and ask, denoted *MID*, is used to compute the first step heights for both supply and demand sides (e.g., Height 1 buy is the best bid less *MID*). The heights and lengths of the demand and supply step functions are then normalized using the summation of all heights and all lengths, respectively.

significant autocorrelation in the residuals of the VEC model.

Table 2 shows the mean of the information share measures (i.e., IS, PT, and MIS) for the model with three price series. In general, PT and MIS measures are in line with each other and show that, for grains and E-mini S&P 500, the highest share is contained in the LOB beyond the BAS. The MIS for the steps beyond the BAS is higher for grains relative to livestock, and is the highest for E-mini S&P 500. The information share contained in the steps beyond the BAS based on PT are 32.76% and 35.21% for lean hogs and live cattle, respectively. The PT measure indicates a high information share for the steps beyond the BAS for wheat (42.26%) and soybeans (43.72%). The

Table 2. Information Share Measures for the Three-Variable Model (%)

		IS_H	IS_L	IS_M	PT	MIS
Live cattle	*Price*	77.67	11.69	40.31	26.83	28.54
	WP^1	63.18	8.66	32.40	37.96	36.53
	$WP^{2\text{-}5}$	55.14	5.36	27.29	35.21	34.92
Lean hogs	*Price*	79.59	13.80	42.78	32.20	31.49
	WP^1	58.55	8.08	30.53	35.04	35.26
	$WP^{2\text{-}5}$	52.79	5.47	26.69	32.76	33.25
Corn	*Price*	53.84	16.87	33.31	20.87	24.91
	WP^1	62.87	16.09	37.20	26.23	32.60
	$WP^{2\text{-}10}$	50.38	12.21	29.49	52.90	42.49
Wheat	*Price*	68.69	12.75	37.42	28.52	27.49
	WP^1	59.75	8.98	31.58	29.22	32.42
	$WP^{2\text{-}10}$	58.41	9.08	31.01	42.26	40.10
Soybeans	*Price*	66.92	11.44	35.75	26.12	25.48
	WP^1	64.50	10.22	34.09	30.16	33.86
	$WP^{2\text{-}10}$	57.60	8.49	30.16	43.72	40.66
E-mini S&P 500	*Price*	77.30	10.16	38.09	23.56	22.59
	WP^1	74.06	6.26	34.98	17.00	32.22
	$WP^{2\text{-}10}$	58.00	3.81	26.92	59.44	45.19

Note: The measures in the table were calculated for each day and averaged over all days. IS_H and IS_L represent the upper and lower bounds of IS measure, respectively. IS_M is calculated by averaging the IS_H and IS_L for each day and then averaging over all days.

Table 3. Information Share Measures for the Four-Variable Model (%)

		IS_H	IS_L	IS_M	PT	MIS
Live cattle	*Price*	79.65	8.66	37.53	23.92	24.20
	WP^1	57.12	5.45	26.60	30.15	28.90
	$WP^{2\text{-}3}$	42.91	3.75	19.83	28.79	28.26
	$WP^{4\text{-}5}$	34.45	3.30	16.04	17.14	18.64
Lean hogs	*Price*	80.80	9.83	39.10	28.43	26.36
	WP^1	53.01	5.05	25.04	28.46	28.40
	$WP^{2\text{-}3}$	40.30	3.29	18.80	26.05	26.48
	$WP^{4\text{-}5}$	35.70	3.83	17.06	17.06	18.77
Corn	*Price*	51.37	11.01	28.08	15.69	18.17
	WP^1	61.68	9.95	32.25	19.06	25.09
	$WP^{2\text{-}3}$	36.88	8.00	20.21	34.34	29.61
	$WP^{4\text{-}10}$	35.59	7.65	19.47	30.91	27.13
Wheat	*Price*	70.01	8.52	34.12	22.80	21.52
	WP^1	59.22	6.62	28.61	24.43	26.56
	$WP^{2\text{-}3}$	33.06	3.40	15.84	27.16	25.77
	$WP^{4\text{-}10}$	43.84	5.46	21.42	25.38	26.14
Soybeans	*Price*	65.81	6.44	31.21	18.42	18.23
	WP^1	63.68	7.32	30.67	25.04	27.28
	$WP^{2\text{-}3}$	33.31	4.00	16.12	28.20	26.28
	$WP^{4\text{-}10}$	44.94	6.02	22.01	28.34	28.21
E-mini S&P 500	*Price*	73.23	4.46	31.61	14.48	13.51
	WP^1	82.00	8.19	36.69	21.05	28.57
	$WP^{2\text{-}5}$	44.26	2.04	18.84	45.85	34.46
	$WP^{6\text{-}10}$	30.72	0.90	12.86	18.62	23.46

Note: The measures in the table were calculated for each day and averaged over all days. IS_M is calculated by averaging the IS_H and IS_L for each day and then averaging over all days.

information share for the steps further from the top of the book is even higher for corn (52.90%) and E-mini S&P 500 (59.44%). The IS measure, on the other hand, attributes a higher share to transaction prices. Table 2 also shows that the higher-lower bound for IS is considerably wide for all products, which emphasizes the importance of a measure for price discovery that is order insensitive. Nevertheless, all three measures indicate a substantial contribution of the steps beyond the BAS to price discovery for all futures contracts, higher than that of Cao, Hansch, and Wang's (2009) study for the Australian stocks (22%). Considering the midpoint IS measure (IS_M), the information share of the steps beyond the BAS for grains is about 30% (29.49% for corn, 30.16% for soybeans, and 31.01% for wheat). The IS_M suggests a slightly lower share for the steps beyond the BAS for live cattle (27.29%), lean hogs (26.69%), and E-mini S&P 500 (26.92%).

The analyzed futures markets appear to be more homogenous with regard to the information contained in the BAS (represented by WP^1 in tables 2 and 3). With the exception of PT for E-mini S&P 500, the BAS contributes about 30% to the price discovery, and contains more information for the meats group compared to the other products based on the PT and MIS measures. On the other hand, the BAS contains more information for corn, E-mini S&P 500, and soybeans relative to live cattle, wheat, and lean hogs, considering the IS_M measure. In addition, price makes a moderate contribution to the efficient price based on the PT and MIS measures. However, the IS_M measure attributes a strong share to price, especially for live cattle and lean hogs (table 2).[15] The three measures lead to

[15] As a robustness check we calculated the information share measures at different frequencies (1-, 2-, 3-, 5-, 15- and 60-second snapshots for the agricultural commodities, and 0.8-, and 2-second snapshots for E-mini S&P500). Changes in duration around the durations selected in this study do not change the estimated information shares appreciably. As a second robustness check, we omitted the variable Price from the model's endogenous variables and sampled our data using the average duration of WP^1 (498 milliseconds for live cattle, 561 milliseconds for lean hogs, 124 milliseconds for corn, 75 milliseconds for soybeans, 143 milliseconds for wheat, and 33 milliseconds for E-mini S&P 500). According to the MIS measure, WP^1 has a higher contribution relative to the other two price series ($WP^{2\text{-}3}$ and $WP^{4\text{-}5}$) in livestock markets. For grains, $WP^{2\text{-}3}$ has the highest information share of the three variables and $WP^{4\text{-}10}$ carries about as much information as WP^1. For E-mini S&P 500, the information declines moderately as we move farther from the BAS. The IS measure

different results when comparing the informativeness of different variables. PT and MIS result in a relatively higher level of informativeness for the steps beyond the BAS, while IS seems to give a higher share to the variables reflecting top of the book. Hasbrouck (2002) shows that while the IS bounds sometimes define a wide range, the correct share value is always included in the range, whereas the PT estimate may be far from the correct value in certain market scenarios.

Our MIS estimates, even though they seem to be closer to the PT estimates, are always included within the IS bounds, whereas in some cases the PT estimates happen to be higher than the IS upper limit bound. The aggregate information share measures for the four-variable model are reported in table 3. Similar to the three-variable model, the information share differs across measures, with higher shares at the top of the book obtained using IS and higher shares on the LOB beyond the BAS obtained using PT and MIS. Table 3 also shows that in general, the information share measures are higher for the steps closer to the BAS for livestock and E-mini S&P 500. From the two meats, live cattle steps closer to the BAS have relatively more information based on the three metrics results. For grains, the measures generate mixed results but it seems that the two steps beyond the BAS are as informative as the seven steps further away. For corn, all measures suggest a slightly higher share for $WP^{2\text{-}3}$ than $WP^{4\text{-}10}$, whereas for soybeans the opposite is true. For E-mini S&P 500, $WP^{2\text{-}5}$ has a higher information share than $WP^{6\text{-}10}$.[16]

Cao, Hansch, and Wang (2009) also examined the distribution of information across the LOB steps beyond the BAS using two variables for Australian stocks. These authors divided their LOB steps beyond the BAS into two variables, one from steps 2 to 4, and the other from steps 5 to 10. Despite considerable higher depth in the steps closer to the BAS, these authors find that the contribution of the steps further from the BAS to price discovery is higher than it is from the steps closer to the BAS. The authors link this phenomenon to the existence of two types of orders in the steps at or near the top of the book: orders that are used to search for hidden orders, or orders that are "faked" and deleted in a few seconds and are used to "spoof" the opposite side of the market in order to improve the price levels in the trader's favor. This "fishing" and "spoofing" may be less present in the steps further from the top of the book and thus those steps may be less noisy and more informative. We find that in the futures markets, this is not the case and the steps closer to the top of the book contain more information than the steps further from the BAS. This can be an indication that faked and hidden orders are used to a lesser degree in the futures markets.

Conclusions

We reconstruct the LOB for five major agricultural commodities, namely live cattle, lean hogs, corn, wheat, and soybeans, as well as a very active futures contract, namely E-mini S&P 500. We find that there is a large number of contracts existing on the bids and asks beyond the BAS for all the products. For agricultural commodities, most contracts exist at steps two and three compared to the BAS and the steps farther. For E-mini S&P 500, we find that the contracts are uniformly distributed along the LOB steps.

An MIS information share metric is used to assess whether informed traders use the steps of the LOB beyond the BAS in futures markets, and especially in agricultural commodity markets in their trading strategies. The MIS results are also compared with those of PT and IS. Our results show a substantial contribution of the steps of the LOB beyond the BAS to price discovery in futures markets,

results, for agricultural commodities, are in line with those of MIS. Ranking of the price series, according to PT, is consistent with both MIS and IS. We also calculated the information share measures using MID instead of WP^1 as in Cao, Hansch, and Wang (2009). The results of the IS midpoint metric for E-mini S&P 500 when MID is used instead of WP^1 are similar to Cao, Hansch, and Wang (2009) for the IS midpoint for the one hundred active Australian stocks. The information share of MID appears to be considerably higher than that of WP^1 according to all three measures for all the products studied. This may be due to the lower level of noise in MID than WP^1, and not necessarily a reflection of a higher level of informativeness (Yan and Zivot 2010; Putnins 2013).

[16] The daily estimates of the information share metrics show that the steps beyond the BAS but closer to the top of the book contain more information during the early and late weekdays and less information in the middle of the week, with mainly Wednesdays creating a V-shaped pattern for the LOB information share for E-mini S&P 500. The pattern exhibits a reverse V shape for the steps beyond the BAS but farther from the top of the book. This means that informed traders in the E-mini S&P 500 futures markets use the steps closer to the top of the book more on the early and late days of the week, as well as the steps farther in the middle of the week. The behavior of the LOB steps for agricultural commodities appears to be less systematic.

one that is higher than previous findings for stock markets (i.e., Cao, Hansch, and Wang 2009). Across the products studied, the MIS for the LOB steps beyond the BAS is about 33% to 35% for the meats, 40% to 42% for the grains, and 45% for E-mini S&P 500. The PT information share suggests that the contribution of the LOB steps beyond the BAS is in the 30% to 40% range for meats, the 40% to 50% range for soybeans and wheat, and the 50% to 60% range for corn and E-mini S&P 500. IS indicates a 27% share for the contribution of the steps beyond the BAS to price discovery for meats and E-mini S&P 500, and about 29% to 31% for the grains. Our results suggest that the LOB beyond the BAS is informative and could be used by traders to develop trading algorithms and strategies.

Along the LOB steps, our results show that for live cattle, lean hogs, corn, and E-mini S&P 500, the steps closer to the top of the book have more information relative to the steps farther from the best bid and ask. Using IS, Cao, Hansch, and Wang (2009) find higher shares for the steps further from the top of the LOB, which is in line with our findings for soybeans (using all three measures) and wheat (using IS and MIS). This suggests that faked and spoofing trades may be present to a higher extent in these two markets relative to the rest of the markets studied.

Supplementary Material

Supplementary material are available at *American Journal of Agricultural Economics* online.

References

Aidov, A. 2013. Three Essays on Market Depth in Futures Markets. PhD Dissertation, Florida International University. FIU Electronic Theses and Dissertations, Paper 974. http://digitalcommons.fiu.edu/etd/974 (last accessed on 24/01/2016).

Aidov, A., and R.T. Daigler. 2015. Depth Characteristics for the Electronic Futures Limit Order Book. *Journal of Futures Markets* 35 (6): 542–60.

Anand, A., and A. Subrahmanyam. 2008. Information and the Intermediary: Are Market Intermediaries Informed Traders in Electronic Markets? *Journal of Financial and Quantitative Analysis* 43 (01): 1–28.

Anand, A., V.A. Gatchev, L. Madureira, C.A. Pirinsky, and S. Underwood. 2011. Geographic Proximity and Price Discovery: Evidence from NASDAQ. *Journal of Financial Markets* 14 (2): 193–226.

Arnade, C., and L. Hoffman. 2015. The Impact of Price Variability on Cash/ Futures Market Relationships: Implications for Market Efficiency and Price Discovery. *Journal of Agricultural and Applied Economics* 47 (4): 539–59.

Aznar, A., and M. Salvador. 2002. Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criterion. *Econometric Theory* 18: 926–47.

Baillie, R.T., G.G. Booth, Y. Tse, and T. Zabotina. 2002. Price Discovery and Common Factor Models. *Journal of Financial Markets* 5 (3): 309–21.

Baruch, S. 2005. Who Benefits from an Open Limit-Order Book? *Journal of Business* 78 (4): 1267–306.

Beveridge, S., and C.R. Nelson. 1981. A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle". *Journal of Monetary Economics* 7 (2): 151–74.

Biais, B., T. Foucault, and S. Moinas. 2015. Equilibrium Fast Trading. *Journal of Financial Economics* 116 (2): 292–313.

Bloomfield, R., M. O'Hara, and G. Saar. 2005. The Make or Take Decision in an Electronic Market: Evidence on the Evolution of Liquidity. *Journal of Financial Economics* 75 (1): 165–99.

Boehmer, E., G. Saar, and L. Yu. 2005. Lifting the Veil: An Analysis of Pre-Trade Transparency at the NYSE. *Journal of Finance* 60 (2): 783–815.

Booth, G., J.C. Lin, T. Martikainen, and Y. Tse. 2002. Trading and Pricing in Upstairs and Downstairs Markets. *Review of Financial Studies* 15 (4): 1111–35.

Brorsen, B.W. 1989. Liquidity Costs and Scalping Returns in the Corn Futures. *Journal of Futures Markets*.9 (3): 225–36.

Bryant, H., and M. Haigh. 2004. Bid-Ask Spreads in Commodity Futures Markets. *Applied Financial Economics* 14 (13): 923–36.

Cao, C., O. Hansch, and X. Wang. 2009. The Information Content of an Open Limit-

Order Book. *Journal of Futures Markets* 29 (1): 16–41.
Chen, H., and P.M.S. Choi. 2012. Does Information Vault Niagara Falls? Cross-Listed Trading in New York and Toronto. *Journal of Empirical Finance* 19 (2): 175–99.
Chen, H., P.M.S. Choi, and Y. Hong. 2013. How Smooth Is Price Discovery? Evidence from Cross-Listed Stock Trading. *Journal of International Money and Finance* 32: 668–99.
Chen, W.P., and H. Chung. 2012. Has the Introduction of S&P 500 ETF Options Led to Improvements in Price Discovery of SPDRs? *Journal of Futures Markets* 32 (7): 683–711.
Chen, Y.L., and Y.F. Gau. 2010. News Announcements and Price Discovery in Foreign Exchange Spot and Futures Markets. *Journal of Banking & Finance* 34 (7): 1628–36.
Chu, Q.C., W.G. Hsieh, and Y. Tse. 1999. Price Discovery on the S&P 500 Index Markets: An Analysis of Spot Index, Index Futures, and SPDRs. *International Review of Financial Analysis* 8 (1): 21–34.
Cont, R., A. Kukanov, and S. Stoikov. 2014. The Price Impact of Order Book Events. *Journal of Financial Econometrics* 12 (1): 47–88.
de Jong, F. 2002. Measures of Contributions to Price Discovery: A Comparison. *Journal of Financial Markets* 5 (3): 323–8.
Eisler, Z., J.P. Bouchaud, and J. Kockelkoren. 2012. The Price Impact of Order Book Events: Market Orders, Limit Orders and Cancellations. *Quantitative Finance* 12 (9): 1395–419.
Engle, R.F., and C.W.J. Granger. 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55 (2): 251–76.
Engle, R.F., and J.R. Russell. 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* 66 (5): 1127–62.
Frank, J., and P. Garcia. 2011. Bid-Ask Spreads, Volume, and Volatility: Evidence from Livestock Markets. *American Journal of Agricultural Economics* 93 (1): 209–25.
Fricke, C., and L. Menkhoff. 2011. Does the “Bund” Dominate Price Discovery in Euro Bond Futures? Examining Information Shares. *Journal of Banking and Finance* 35 (5): 1057–72.
Frijns, B., A. Gilbert, and A. Tourani-Rad. 2010. The Dynamics of Price Discovery for Cross-Listed Shares: Evidence from Australia and New Zealand. *Journal of Banking & Finance* 34: 498–508.
Frino, A., A. Lepone, and G. Wearin. 2008. Intraday Behavior of Market Depth in a Competitive Dealer Market: A Note. *Journal of Futures Markets* 28 (3): 294–307.
Glosten, L. 1994. Is the Electronic Open Limit Order Book Inevitable? *Journal of Finance* 49 (4): 1127–61.
Gonzalo, J., and C. Granger. 1995. Estimation of Common Long-Memory Components in Cointegrated Systems. *Journal of Business & Economic Statistics* 13 (1): 27–35.
Gonzalo, J., and J.Y. Pitarakis. 1998. Specification via Model Selection in Vector Error Correction Models. *Economics Letters* 60 (3): 321–8.
Gould, M.D., M.A. Porter, S. Williams, M. McDonald, D.J. Fenn, and S.D. Howison. 2013. Limit Order Books. *Quantitative Finance* 13 (11): 1709–42.
Grammig, J., and F.J. Peter. 2013. Telltale Tails: A New Approach to Estimating Unique Market Information Shares. *Journal of Financial and Quantitative Analysis* 48 (02): 459–88.
Harris, F.H., T.H. McInish, and R.A. Wood. 2002. Security Price Adjustment Across Exchanges: An Investigation of Common Factor Components for Dow Stocks. *Journal of Financial Markets* 5 (3): 277–308.
Harris, L., and V. Panchapagesan. 2005. The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions. *Journal of Financial Markets* 8 (1): 25–67.
Hasbrouck, J. 1991. Measuring the Information Content of Stock Trades. *Journal of Finance* 46 (1): 179–207.
———. 1995. One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance* 50 (4): 1175–99.
———. 1996. Order Characteristics and Stock Price Evolution. An Application to Program Trading. *Journal of Financial Economics* 41 (1): 129–49.
———. 2002. Stalking the “Efficient Price” in Market Microstructure Specifications: An Overview. *Journal of Financial Markets* 5 (3): 329–39.

———. 2004. Liquidity in the Futures Pits: Inferring Market Dynamics from Incomplete Data. *Journal of Financial and Quantitative Analysis* 39 (02): 305–26.

Hautsch, N., and R. Huang. 2012. The Market Impact of a Limit Order. *Journal of Economic Dynamics and Control* 36 (4): 501–22.

Huang, R. 2002. The Quality of ECN and NASDAQ Market Maker Quotes. *Journal of Finance* 57 (3): 1285–319.

Irwin, S.H., and D.R. Sanders. 2012. Financialization and Structural Change in Commodity Futures Markets. *Journal of Agricultural and Applied Economics* 44 (03): 371–96.

Janzen, J.P., and M.K. Adjemian. 2017. Estimating the Location of World Wheat Price Discovery. *American Journal of Agricultural Economics* 99 (5): 1188–207.

Kaniel, R., and H. Liu. 2006. So What Orders Do Informed Traders Use? *Journal of Business* 79 (4): 1867–913.

Korczak, P., and K. Phylaktis. 2010. Related Securities and Price Discovery: Evidence from NYSE-Listed Non-U.S. Stocks. *Journal of Empirical Finance* 17 (4): 566–84.

Lehecka, G.V., X. Wang, and P. Garcia. 2014. Gone in Ten Minutes: Intraday Evidence of Announcement Effects in the Electronic Corn Futures Market. *Applied Economic Perspectives & Policy* 36 (3): 504–26.

Lien, D., and K. Shrestha. 2009. A New Information Share Measure. *Journal of Futures Markets* 29 (4): 377–95.

———. 2014. Price Discovery in Interrelated Markets. *Journal of Futures Markets* 34 (3): 203–19.

Lien, D., and Z. Wang. 2016. Estimation of Market Information Shares: A Comparison. *Journal of Futures Markets* 36 (11): 1108–24.

Liu, Q., and Y. An. 2011. Information Transmission in Informationally Linked Markets: Evidence from U.S. and Chinese Commodity Futures Markets. *Journal of International Money and Finance* 30 (5): 778–95.

Lo, D., and A. Hall. 2015. Resiliency of the Limit Order Book. *Journal of Economic Dynamics and Control* 61: 222–44.

Madhavan, A., D. Porter, and D. Weaver. 2005. Should Securities Markets Be Transparent? *Journal of Financial Markets* 8 (3): 265–87.

Martinez, V., P. Gupta, Y. Tse, and J. Kittiakarasakun. 2011. Electronic Versus Open Outcry Trading in Agricultural Commodities Futures Markets. *Review of Financial Economics* 20 (1): 28–36.

Martinez, V.H., and I. Roşu. 2011. High Frequency Traders, News, and Volatility. *Working Paper*, Baruch College and HEC Paris.

Narayan, S., and R. Smyth. 2015. The Financial Econometrics of Price Discovery and Predictability. *International Review of Financial Analysis* 42: 380–93.

Putnins, T.J. 2013. What Do Price Discovery Metrics Really Measure? *Journal of Empirical Finance* 23: 68–83.

Rittler, D. 2012. Price Discovery and Volatility Spillovers in the European Union Emissions Trading Scheme: A High-Frequency Analysis. *Journal of Banking & Finance* 36 (3): 774–85.

Seppi, D. 1997. Liquidity Provision with Limit Orders and a Strategic Specialist. *Review of Financial Studies* 10 (1): 103–50.

Shah, S., and B.W. Brorsen. 2011. Electronic vs. Open Outcry: Side-by-Side Trading of KCBT Wheat Futures. *Journal of Agricultural and Resource Economics* 36 (1): 48–62.

Stock, J.H., and M.W. Watson. 1988. Testing for Common Trends. *Journal of the American Statistical Association* 83 (404): 1097–107.

Tse, Y. 1999. Round-the-Clock Market Efficiency and Home Bias: Evidence from the International Japanese Government Bond Futures Markets. *Journal of Banking and Finance* 23 (12): 1831–60.

Vo, M.T. 2007. Limit Orders and the Intraday Behavior of Market Liquidity: Evidence from the Toronto Stock Exchange. *Global Finance Journal* 17 (3): 379–96.

Wang, X., P. Garcia, and S.H. Irwin. 2014. The Behavior of Bid-Ask Spreads in the Electronically-Traded Corn Futures. *American Journal of Agricultural Economics* 96 (2): 557–77.

Yan, B., and E. Zivot. 2010. A Structural Analysis of Price Discovery Measures. *Journal of Financial Markets* 13 (1): 1–19.