Received: 5 May 2022 | Accepted: 11 July 2022

DOI: 10.1002/fut.22368

RESEARCH ARTICLE

WILEY

Trades or quotes: Which drives price discovery? Evidence from Chinese index futures markets

Liwei Jin[1] | Xianghui Yuan[1] | Shihao Wang[1] | Peiran Li[1] | Feng Lian[2]

[1]School of Economics and Finance, Xi'an Jiaotong University, Xi'an, Shaanxi, China

[2]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China

Correspondence
Xianghui Yuan, School of Economics and Finance, Xi'an Jiaotong University, 74# Yanta West Rd, 710061 Xi'an, Shaanxi, China.
Email: xhyuan@mail.xjtu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 11801433, 11971372, 72101199

Abstract

In this paper, we analyze the role that trades and quotes play in price discovery. Based on tick-level data for CSI 300 stock index futures, we find that the contribution of quotes to price discovery does not differ from trades at low resolutions, but dominates at high resolutions. This difference is influenced by spreads and volume. Further analysis reveals that the intraday price contribution of quotes and trades is trending downward, up to 31% in the first half-hour. The adverse selection and liquidity supply cost components of spreads significantly contribute to and dampen the difference in intraday contribution, respectively.

KEYWORDS

index futures, price discovery, quote, spread, trade

1 | INTRODUCTION

Price discovery, the most important function of the securities market, refers to the efficient and timely incorporation of new information into the price of a security (He et al., 2009). According to the traditional view of price discovery, trades and quotes reflect investors' private information and rational expectations (Glosten & Milgrom, 1985; Kyle, 1985). Brogaard et al. (2019) pointed out that the choice of market and limit orders made by informed and uninformed investors determine price discovery. Thus in order-driven markets, quoted and traded prices imply different information. In this paper, we systematically quantify the contribution of quotes and trades to price discovery and their difference, and further analyze intraday trend changes. The results are explained through the lens of bid–ask spread theory.

Many important issues regarding market microstructure depend on the joint dynamics of quoted and actual traded prices. These dynamics changes are helpful in identifying innovations in the market microstructure (Hasbrouck, 2021). They are directly related to the process of incorporating information into prices, i.e. the price discovery process. Theoretically, the last transaction price of a security reflects traders' timely response to the latest market information. At the same time, the quotation behavior of traders also implies the wide spreads of market information, which serves as the basis for other investors to make trading decisions. It is worth noting that a trader's quote may not be completely executed. In other words, some orders may not ultimately be fulfilled, although these orders also contain information that drives the price formation. Therefore, both quotes and trades play an important role in promoting the price discovery process, and discovering which one is more powerful in price formation deserves in-depth study.

With the rapid development of financial computer technology, high-frequency transaction and order data provide more possibilities for research on price discovery. Hasbrouck (2021) explored the price discovery with a resolution of 10 microseconds. This sparked our interest in exploring the differences in price discovery at different resolutions. Meanwhile, the influencing factors of the difference in price discovery contribution between trade and quote are also issues that are studied in this paper. Furthermore, we aim to explore the proportion of price discovery at different times in a day, and provide quantitative evidence by decomposing the bid–ask spread into two parts: information asymmetry; and liquidity supply cost.

By addressing the above research questions, we aim to gain new insights into price discovery from a high-frequency perspective. It is worth noting that relatively few studies have focused on the influencing factors of price discovery. Revealing the influencing factors helps us to deepen our understanding of the potential mechanisms of price discovery. In addition, addressing these questions helps us to understand investor trading behavior, as the answers not only show price discovery differences between quotes and trades, but also provide insights into intraday trend changes. This is of great value to investors in formulating trading strategies and to decision makers in maintaining the stability and effectiveness of the market.

The questions addressed are both interesting and exciting, but answering them is not easy. First of all, as detailed and reliable data are the basis for our research, the collection and preprocessing of high-frequency data are major problems that need to be overcome. Second, choosing a suitable model to quantify the contribution of price discovery among various methods is also an arduous but necessary task.

Our main findings are threefold. First, the contribution of quotes to price discovery is indistinguishable from trades at low resolution, while it dominates at high resolution. Second, spreads and trading volume are important factors that influence the difference in price contribution between quotes and trades. Last but not least, the information asymmetry and liquidity supply cost components of the spread decomposition have opposite but significant effects on the intraday variation in price discovery.

To present our work logically and clearly, the remainder of this paper is organized as follows. Section 2 reviews relevant literature on price discovery. Section 3 describes the methods and data in detail. Section 4 empirically analyzes the contribution of quotes and trades to price discovery and the factors influencing their differences. Section 5 further discusses the contribution of price discovery in different time periods in a day, and also provides robustness tests. Finally, Section 6 provides conclusions.

2 | LITERATURE REVIEW

When homogenous or closely related securities are traded in multiple markets, the importance of determining where the price discovery takes place has aroused great interest among researchers in the recent years. Hasbrouck (1995) used an information share (IS) model to empirically analyze on 30 Dow Jones stocks and found that price discovery occurs primarily in the New York Stock Exchange (NYSE). Patel et al. (2020) used a new empirical method for information leadership and found that options now play a role in price discovery five times greater than previously. Karmakar and Inani (2019) investigated price discovery in Indian stock markets and found that the spot market played a dominant role throughout the sample period. In addition, similar research has been extended to Bitcoin (Hung et al., 2021), treasury bonds (Mizrach & Neely, 2008), exchange rates (Chen & Xu, 2021), exchange-traded funds (ETFs) (Buckle et al., 2018; Hasbrouck, 2003), and commodity futures markets (Hu et al., 2020; Jing et al., 2021; Li & Xiong, 2021).

With the development of the theory of market microstructure, many studies have focused on the quotes and trades of the same asset. Pascual and Pascual-Fuster (2014) found that quotes on one side of the market are usually more informative than those on the other side, and that this asymmetry is more pronounced in small-cap stocks. Chen and Gau (2014) found that bid quotes in the Forex (FX) electronic limit order market contribute more to price discovery than ask quotes. In addition, Frijns and Tse (2015) found that trades in the FTSE 100 index futures market contain more information than quotes, explaining about 80% of the innovation in the efficient price. However, Brogaard et al. (2019) found that the IS of quotes was generally increasing, and that high-frequency trading (HFT) bids and offers were more informative than non-HFT. Hasbrouck (2021) examined the price contribution between quotes, lit trades, and dark trades, revealing that the quotes price contribution was as high as 65%. The classical information asymmetry model assumes that public information is reflected by quotes, while private information is reflected by trades; however, this view is overly simplistic. With high-frequency exchange gradually replacing the traditional liquidity providers, the rapid update of quotes enhances the liquidity of information, so it is necessary to re-examine the relative price contribution of quotes and trades.

Moreover, there are various approaches to explore price discovery. In addition to the IS model, they include the component share model (Harris et al., 1995), the permanent transient (PT) model (Gonzalo & Granger, 1995), the weighted price contribution (WPC) method (Barclay & Hendershott, 2008), the tail dependence measure (Grammig & Peter, 2013), and the information leadership share (Putnins, 2013), among others. The vector error correction (VEC) model explains long-run equilibrium and short-run shocks well (Engle & Patton, 2004), and the IS model proposed by Hasbrouck (1995) on its basis has become a popular choice for studying price discovery (Chen & Gau, 2014; Pascual & Pascual-Fuster, 2014). Furthermore, Baillie et al. (2002) demonstrated the feasibility of the IS model when the correlation between series is low, and Hasbrouck (2021) successfully applied the IS model to high-frequency sequences. Thus, in this paper, the IS model is chosen for experiments and the PT model is used for robustness checks.

3 | METHODOLOGY AND DATA

3.1 | Methodology

This study employs IS and PT models to investigate the price discovery process, and both models are based on vector VEC models (Davidson et al., 1978).

Hasbrouck (1995) attributed the sources of effective prices change to different markets. In this context, price discovery is measured by the variance of innovation to the common factor. Thus, the share of the contribution of price discovery can be defined as the proportion of the variance of innovation in effective prices in each market.

First, considering two first-order cointegrated price series $Y_t = (y_{1t}, y_{2t})'$, the error correction term can be denoted as $Z_t = \beta' Y_t = y_{1t} - y_{2t}$, where $\beta = (1, -1)'$ is the cointegration vector. Thus, the VEC model can be expressed as

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{k} A_j \Delta Y_{t-j} + e_t, \tag{1}$$

where α is the error correction vector and e_t is the serially uncorrelated zero-mean vector with covariance matrix $\mathbf{\Omega}$. The $\alpha\beta' Y_{t-1}$ term represents the long-term equilibrium of the variables in the system, while the $\sum_{j=1}^{k} A_j \Delta Y_{t-j}$ term represents the short-term dynamics of the system.

Hasbrouck (1995) transformed Equation (1) into a vector moving average (VMA):

$$\Delta Y_t = \boldsymbol{\psi}(L) e_t \tag{2}$$

and its integrated form:

$$Y_t = \boldsymbol{\psi}(1) \sum_{s=1}^{t} e_s + \boldsymbol{\psi}^*(L) e_t, \tag{3}$$

where $\boldsymbol{\psi}(L)$ and $\boldsymbol{\psi}^*(L)$ are matrix polynomials in the lag operator L. $\boldsymbol{\psi}(1)$ as the impact matrix can be calculated from the sum of the moving average coefficients, and the estimation of $\boldsymbol{\psi}(1)$ is crucial to the results of IS. Denoting $\psi = (\psi_1, \psi_2)$ as the common row vector in $\boldsymbol{\psi}(1)$, Equation (3) can be re-written as

$$Y_t = \iota\psi \sum_{s=1}^{t} e_s + \boldsymbol{\psi}^*(L) e_t, \tag{4}$$

where ι is the column unit vector. The above equation reveals that the price consists of the common factor $\iota\psi\sum_{s=1}^{t} e_s$ and the transitory portion $\boldsymbol{\psi}^*(L)e_t$. The component of the price change can be represented by ψe_t, which reflects the new information. By calculating the variance $\psi\Omega\psi'$ of ψe_t, the ISs of market j can be expressed as

$$IS_j = \frac{\psi_j^2 \Omega_{jj}}{\psi\Omega\psi'}. \tag{5}$$

However, there may be contemporaneous correlation in price innovations, especially when using low-frequency data (Tse, 2000). To address this issue, Hasbrouck (1995) expressed the covariance matrix as $\Omega = FF'$ through Cholesky decomposition. Thus, Equation (5) can be re-written as

$$IS_j = \frac{([\psi F]_j)^2}{\psi \Omega \psi'}, \tag{6}$$

where F is a lower triangular matrix. $[\psi F]_j$ is the jth element of the row of matrix $[\psi F]_j$; it can be denoted as

$$F = \begin{pmatrix} f_{11} & 0 \\ f_{21} & f_{22} \end{pmatrix}. \tag{7}$$

Baillie et al. (2002) showed that the values of ψ are directly related to $\alpha_{\perp} = (\gamma_1, \gamma_2)'$, so the IS of market j can be expressed as

$$IS_j = \frac{([\gamma F]_j)^2}{\gamma \Omega \gamma'}. \tag{8}$$

Furthermore, Baillie et al. (2002) provided the IS calculation method for two markets:

$$IS_1 = \frac{(\gamma_1 f_{11} + \gamma_2 f_{21})^2}{(\gamma_1 f_{11} + \gamma_2 f_{21})^2 + (\gamma_2 f_{22})^2}, \tag{9}$$

$$IS_2 = \frac{(\gamma_2 f_{22})^2}{(\gamma_1 f_{11} + \gamma_2 f_{21})^2 + (\gamma_2 f_{22})^2}. \tag{10}$$

Equations (9) and (10) show that the calculation of the IS depends only on the error correction vector α and the covariance matrix $\mathbf{\Omega}$. We can find that IS_1 contains not only the contribution of its own series, but also the correlation of other series. However, in IS_2 only pure contributions that are not correlated with other series are considered. Thus the upper (lower) bond of the IS value is obtained when the variable is the first (last) in the Cholesky ordering.

However, the PT model proposed by Gonzalo and Granger (1995) only focuses on the error correction process, and only considers the permanent impact that leads to imbalance. In this model, the contribution of each market is defined as a function of the market error correction coefficient. The PT value of each market is expressed as

$$PT_1 = \frac{|\alpha_2|}{|\alpha_1| + |\alpha_2|} \text{ and } PT_2 = \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|}. \tag{11}$$

When $PT_1 = 1$, the price discovery is completely dominated by market 1. If the two markets have equal contributions to price discovery, then their PT values should both be equal to 0.5. Therefore, a larger PT value indicates a greater contribution to price discovery by the corresponding market.

In addition, compared with IS or PT methods, the WPC method proposed by Barclay and Warner (1993) allows us to observe the price discovery in different periods of the day. Referring to the adaptation of the WPC method proposed by Buckle et al. (2018), we first calculate the return of the price series according to the following equation:

$$\begin{aligned} R_k &= \ln(P_k) - \ln(P_{k'}) \\ R_{k,t} &= \ln(P_{k,t}) - \ln(P_{k,t-1})' \end{aligned} \tag{12}$$

where R_k is the return on day k, P_k and $P_{k'}$ represent the close and open price of the day k, respectively, and $R_{k,t}$ represents the return of period t in day k. Using these returns, the WPC is given by

$$WPC_t = \sum_{k=1}^{T'} \left(\frac{|R_k|}{\sum_{k=1}^{T'} |R_k|} \right) \times \left(\frac{R_{k,t}}{R_k} \right). \tag{13}$$

The first item on the right side of the equation is the weighting coefficient of each day during T'. The second item is the relative contribution of period t on day k. We set T' to the number of trading days included in a month following Buckle et al. (2018). It is important to note that, although the WPC method does not reveal who dominates the market as the IS or PT methods do, it can reveal price contributions between different times of the day.

3.2 | Data

To conduct our study, we used tick-level data for CSI 300 stock index futures. As the earliest and most representative stock index futures contracts listed in China, its good liquidity has attracted the attention of a large number of investors and researchers (Han & Liang, 2017; Sohn & Zhang, 2017).

It should be underlined that the update frequency of the tick-level data we obtained is 0.5 s, that is, the interval of each tick is 0.5 s. This is the most frequent data set we can obtain currently. Our data set contains all trading days from January 1, 2021 to January 1, 2022. For each trading day, we obtained all intraday trade and quote data from 9:30 a.m. to 3:00 p.m. (trading is suspended from 11:30 to 13:00). Table 1 presents descriptive statistics of the experimental data.

Theoretically, we could obtain 28,800 observations at a frequency of 0.5 s during a 4-h trading day. Note, however, that in our study only price updates are considered, that is, the change of price compared to the previous moment (Frijns & Tse, 2015). Observe that Table 1 shows that the average number of updates per day for trades and quotes are 19,209 and 27,034, respectively, and it is clear that the quotes are updated more frequently. Also, we can find that the quotes have a smaller variance compared to the trades, which indicates a greater degree of fragmentation for daily trades data. The "Duration" column represents the average duration of changes in trades and quotes, which are 0.75 and 0.53 s, respectively. Given the higher update frequency, we can assume that the quotes data contain more information.

For outliers presented in the data, Verousis and Gwilym (2010) recommended removing excessive price changes that exceed a 5% threshold. The removal of outliers is controversial in many studies, and we believe that post-opening trades and quotes are informative, even for these outliers. Therefore, to better evaluate of price discovery, we adopted the approach of Buckle et al. (2018), which does not remove these outliers. All of our data came from the JoinQuant database (https://www.joinquant.com/).

4 | EMPIRICAL RESULTS

4.1 | Price discovery at different resolutions

In this section, we quantify the price discovery contribution of quotes and trades. Note that we use the midpoints of bid and ask quotes as the quote series. Midpoint quotes are typically used as a proxy for fundamental value. They contain less microstructural noise than trades prices, which are affected by bid–ask bounces. The feasibility of using the bid–ask midpoint as a proxy for quotes has also been confirmed by previous studies (Hasbrouck, 2003; Papavassiliou & Kinateder, 2021; Pascual et al., 2006). On the basis of the data we obtained for the 0.5 s frequency, we further obtained a sequence of trade and quote prices at 1, 5, 10, and 60 s frequencies. Taking each day as a subsample, based on the estimation results for Equation (1), we calculated the IS and PT values at different sampling frequencies. For the IS model, we further calculated the upper and lower bounds, as well as their mean values. Table 2 reports the mean values of the above results over the total sample period.

Table 2 shows that, at low resolution (60 s), the information contribution of trades and quotes calculated by the IS model is about 50%, and there is no significant difference between them. However, as the resolution increases, the price contribution percentage of the quotes also increases. The quotes price discovery contribution is as high as 66% at a 0.5 s sampling frequency.

TABLE 1 Descriptive statistics of daily trade and quote innovations

	Mean	Max	Min	Median	Std	Duration
Trade	19,209	26,049	7368	20,605	4306.17	0.75
Quote	27,034	28,710	20,316	27,744	1809.36	0.53

TABLE 2 IS and PT values at different resolutions

	IS						PT	
	Trade			Quote				
Resolution	Upper	Lower	Mean	Upper	Lower	Mean	Trade	Quote
60 s	0.9786	0.0134	0.4960	0.9866	0.0214	0.5040	0.4250	0.5750
10 s	0.9605	0.0066	0.4835	0.9934	0.0395	0.5165	0.2546	0.7454
5 s	0.9306	0.0096	0.4701	0.9904	0.0694	0.5299	0.2454	0.7546
1 s	0.7137	0.0364	0.3750	0.9636	0.2863	0.6250	0.2447	0.7553
0.5 s	0.6220	0.0552	0.3386	0.9448	0.3780	0.6614	0.2629	0.7371

Note: "Upper," "Lower," and "Mean" indicate the upper and lower bounds calculated by the IS model and their average values, respectively. All analyses are based on a sampling frequency of 0.5 s.

Abbreviations: IS, information share; PT, permanent transient.

To clearly observe the change in information contribution, we report the upper, lower, and average values of the IS model for quotes and trades at a sampling frequency of 0.5 s, as well as in the results of the PT model in Figure 1. We calculate the corresponding upper and lower bounds as the first series for the sequence of quotes and trades, respectively. As shown in Figure 1, although the price discovery of quotes and trades fluctuate, all values of quotes are higher than those of trades, and quotes show a more stable trend over the sampled period. This confirms our assumption.

Methodologically speaking, our experimental results support Baillie et al. (2002)'s view regarding the measurement aspects of the IS and PT models. They argued that the residual correlation depends partly in the market information flow and partly on the frequency of the price data, and that data with very high frequency are usually less correlated. In this case, the average of the upper and lower bounds of the IS model can be used as a reasonable estimate of the effective price discovery in the market, which is also consistent with the PT model results.

As for price discovery, the quoted price provides a higher contribution compared to the traded price. Thus, we can conclude that quotes face a lower risk of adverse selection, and a higher proportion is provided by informed traders, especially in HFT (Rindi, 2008). Subsequently, the quote may adapt to the innovation of fundamental values faster than the traded price (Goettler et al., 2009). The higher contribution to price discovery by the quote under high-frequency price series has also been proven in a recent study by Hasbrouck (2021).

4.2 | Determinants of the difference between trades and quotes on price discovery

In this subsection, we focus on exploring factors that significantly affect the price discovery difference between quotes and trades.

The bid–ask spread cannot be ignored as an important indicator. When a buyer (seller) initiates a trade that may lead to a rise (fall) in the midpoint of the quote, the liquidity provider observes the change in the transaction price and quickly makes an adjustment to the quote price, resulting in a change in the bid–ask spread compared to the previous moment. At the same time, the change in the quotation affects the price of the next moment of trading. In this case, the changes in the quoted and traded prices and the size of the bid–ask spread drive each other. Therefore, it is necessary to ascertain how the spread affects the price discovery of quotes and trades.

Moreover, information asymmetry and the costs of liquidity supply also influence investors' choice of market and limit orders. Therefore, we also investigate the impact of spread components on price discovery differences. Here, we use the spread decomposition method proposed by Madhavan et al. (1997) to make inferences about the information asymmetry component and the cost of liquidity supply component, as detailed in Appendix A.

In addition, we also consider the influence of market liquidity and momentum. We therefore introduce daily returns as their proxy variable (Chen & Gau, 2014). The daily return is calculated by the midpoint of the opening and closing quotes.

It should be noted that informed traders of HFT update their quotes frequently, and part of the quotes may only be made to induce the completion of a trade rather than for it to be filled, thus steering the price movement in its own favor. This results in a huge difference between the quote and the trade volume. It is therefore necessary to consider the impact of quote volume versus trade volume on price discovery differences.

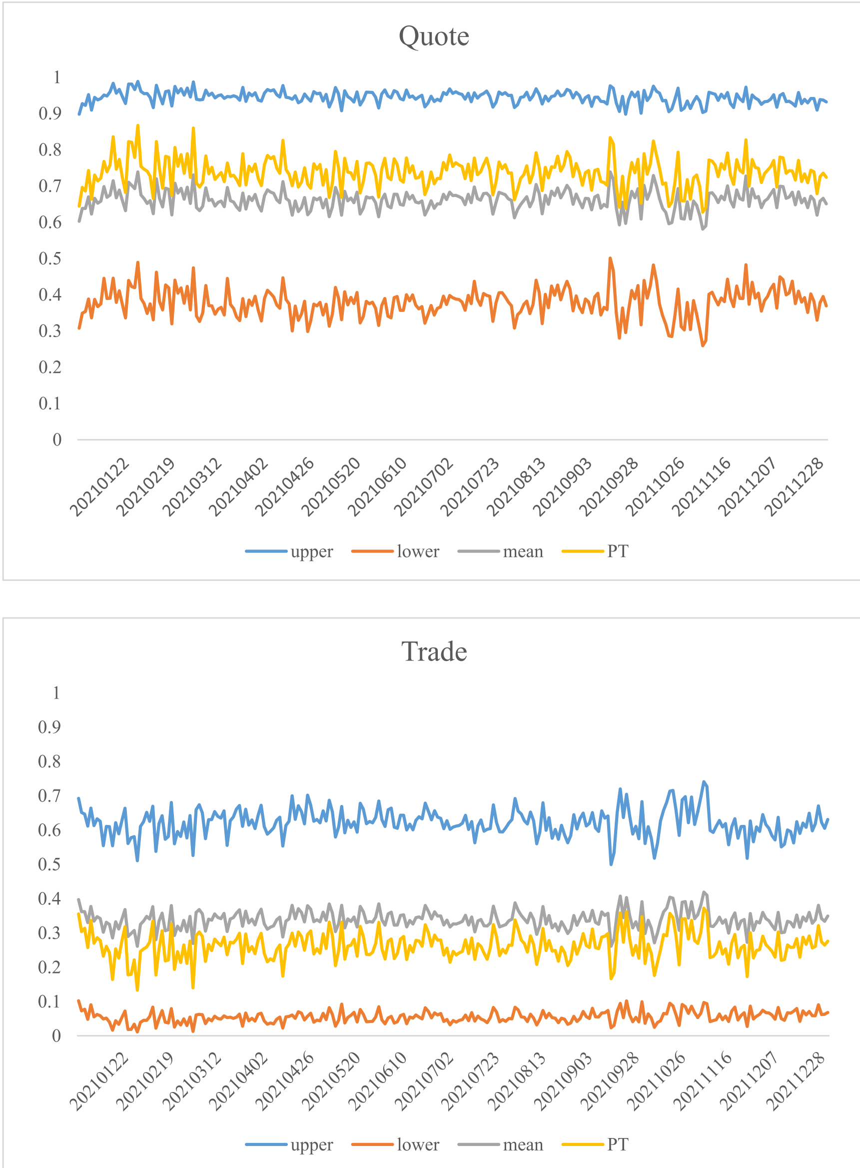

FIGURE 1 The IS and PT results for quotes and trades. The figure shows the upper bound, lower bound, and average values of the IS model calculations and the PT model values for each day from January 1, 2021 to January 1, 2022. IS, information share; PT, permanent transient.

Table 3 reports the descriptive statistics for each variable. RCIS and RCPT denote the price discovery difference between quotes and trades under the IS and PT models, respectively. The *t*-values in Table 3 indicate that price discovery is significantly higher for quotes than for trades. In addition, we can also find that the size of the quoted volume is noticeably higher than the traded volume. For the spread, the sum of the decomposed information asymmetry cost (Theta) and the liquidity supply cost (Phi) is close to one-half of the spread, which is consistent with the results for the spread decomposition found by Madhavan et al. (1997).

To determine whether the variables in Table 3 are related to the difference in price discovery for quotes and trades, we estimate the following regression:

$$RC_t = \alpha + \beta X_t + \varepsilon_t, \tag{14}$$

where RC_t is used as the dependent variable, indicating RCIS or RCPT, α and ε_t are the constant term and error terms, respectively, and X_t is the influencing factor in Table 3. Note that Trade-vol and Quote-vol are not directly included in our model, but the logarithmic difference between them is used. In addition, the information asymmetry and liquidity

TABLE 3 Summary statistics

	Min	Max	Mean	Std	Skew	Kurt	T
RCIS	0.1619	0.4776	0.3228	0.0548	−0.0765	0.4872	91.83***
RCPT	0.2556	0.7337	0.4742	0.0810	0.0412	0.6109	91.24***
Trade-Vol	16429	165628	77113.7	30612.3	−0.1903	−0.3879	
Quote-Vol	33040	207402	125918	35263.7	−0.9201	0.4930	
Return	−0.0468	0.0325	−0.0003	0.0112	−0.3588	1.5986	
Spread	0.4616	1.0085	0.6275	0.1266	1.3032	0.7047	
Theta	0.1142	0.3297	0.2013	0.0466	0.5396	−0.1399	
Phi	0.0528	0.1610	0.0917	0.0186	0.5653	0.4380	

Note: "RCIS" and "RCPT," respectively, represent the price discovery difference between quotes and trades under the IS model and the PT model. The *t*-values in the table are tests for the corresponding series, which are significantly greater than 0. "Trade-vol" and "Quote-vol" represent the volume of trades and quotes, respectively. "Theta" and "Phi" are the information asymmetry cost and liquidity supply cost decomposed from the bid–ask spread, respectively. ***indicates significant at the 1% level.

TABLE 4 Regression results for factors influencing price discovery differences between quotes and trades

	Constant	Ln-diff	Return	Spread	Theta	Phi	Adj-R^2
Dependent variable: RCIS							
(1)	−0.4340*** (0.159)	0.0656*** (0.014)	−0.9126*** (0.302)	0.0818** (0.034)			0.097
(2)	−0.4418*** (0.184)	0.0631*** (0.016)	−0.8251*** (0.301)		0.2117** (0.106)	0.4735** (0.183)	0.108
Dependent variable: RCPT							
(3)	−0.3950 (0.240)	0.0753*** (0.020)	−1.1580** (0.457)	0.0949* (0.051)			0.057
(4)	−0.5627** (0.277)	0.0854*** (0.024)	−1.0700*** (0.455)		0.3609** (0.160)	0.5041* (0.276)	0.070

Note: "Ln-diff" indicates the logarithmic difference between Quote-vol and Trade-vol. The value in parentheses are robust standard errors, and ***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.

supply cost components are highly correlated because they are derived from the decomposition of the spread, so we include the spread and spread decomposition components in the regression model, respectively. The regression results are reported in Table 4.

Table 4 shows that both RCIS and RCPT have similar results, implying the robustness of the experimental results. It is not difficult to find that Ln-diff significantly promotes the difference in price discovery, suggesting that a greater volume of quotations contains more information, which leads to higher price discovery for quotes than trades. This finding provides strong support for the claim that informed trades select more limit orders in HFT.

The results in Table 4 also show that daily returns significantly affect RCIS and RCPT, and on days marked by lower daily returns, the dominance of quotes in price discovery is more likely to be found. In addition, spreads also significantly increase RCIS and RCPT, and the decomposition of spreads further suggests a significant role both for information asymmetry and liquidity supply costs. This suggests that greater information asymmetry and the higher cost of liquidity supply lead to a more common tendency for investors to trade with limit orders, which results in more market information being reflected by quote orders.

5 | FURTHER ANALYSIS

In Section 4, we found that quotes drive price discovery more than trades because of the influence of daily returns, volume, and bid–ask spreads. In this section, we further explore the variation in price discovery across time during the day and the factors affecting it.

5.1 | Intraday price discovery at different times of the day

To gain further insights into the intraday price discovery process, we adopt the WPC method introduced in Section 3, which is widely used to calculate the intraday price contribution of the same asset (Barclay & Hendershott, 2008; Cheng et al., 2004; Lien et al., 2020).

We divide each day using half-hour intervals and obtain eight time periods. For each time period, WPC is calculated as a percentage of the current time period in the price discovery of the whole day. We take T to be 22 in calculating WPC (a month contains roughly 22 trading days). Figure 2 reports the average WPC results for each half-hour of each day in the sample period.

In Figure 2, it can be observed that the WPC values do not differ much within each half-hour, both for quotes and trades, which is consistent with our finding in Table 2 that there is no significant difference in price discovery between quotes and trades at low frequencies. In addition, it is not difficult to see a decreasing trend in price discovery over the eight half-hours of the day, with a high price discovery of 31% in the first half-hour, accounting for almost one-third of the whole day, and less than 3% in the last half-hour. This indicates that price discovery is concentrated in the first half-hour, which is due to the accumulation of overnight information that is released after the opening and is digested as trading proceeds, thus creating a phenomenon of decreasing price discovery during the day.

5.2 | Spreads and intraday price discovery

In this subsection we explore the relationship between spreads and intraday price discovery at different times of the day. The accumulation of information overnight is released in the early morning session and then fades away during the day. Therefore, the information asymmetry is highest in the opening session, and for this reason we also decompose the spreads using the method proposed by Madhavan et al. (1997) to investigate how information asymmetry and liquidity supply costs affect intraday price discovery. It should be stressed that we calculate bid–ask spreads and their components in the same period as WPC. A graphical depiction of spreads and their components over the eight half-hour intraday periods is plotted in Figure 3.

The gray curve in Figure 3 represents the bid–ask spread, corresponding to the right-hand axis. The orange and blue bars represent the information asymmetry cost (Theta) and the liquidity supply cost (Phi), respectively. The results presented in Figure 3 are the average values over the sample period. Figure 3 clearly shows that the spread is highest in the first half-hour and then gradually drops, showing a decreasing trend during the day. The decomposition of the spreads shows that the information asymmetry component is the main reason for the decreasing trend of the bid–ask spreads during the day. Information asymmetry is highest in the opening session; in other words, informed traders prefer to trade in the opening period. The liquidity supply cost shows a slow upward trend at different times of the day.

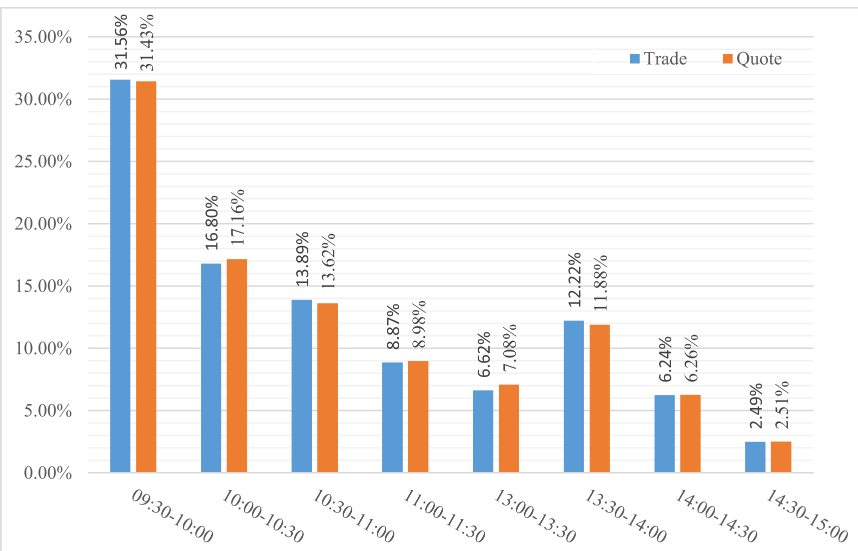

FIGURE 2 Weighted price contribution (WPC) results

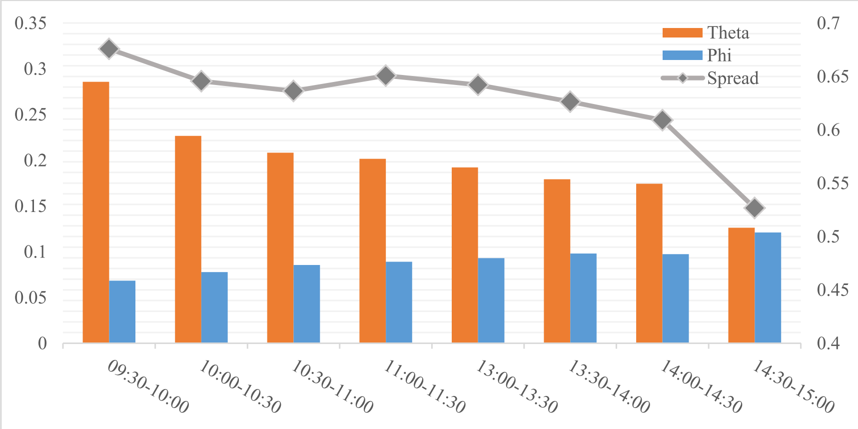

FIGURE 3 Half-hourly spreads and their components. The spreads correspond to the right axis and the components (Theta and Phi) to the left axis.

TABLE 5 Regression results of spreads and decomposition components on intraday price discovery

	Constant	Spread	Theta	Phi	Adj-R^2
Dependent variable: Trade					
(1)	−0.0251 (0.017)	0.2366*** (0.026)			0.044
(2)	0.1082*** (0.026)		0.7206*** (0.057)	−1.4215** (0.176)	0.306
Dependent variable: Quote					
(3)	−0.0274* (0.016)	0.2409*** (0.026)			0.047
(4)	0.1112** (0.025)		0.7223*** (0.056)	−1.4546*** (0.173)	0.318

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Hagstromer et al. (2016) further divided liquidity supply costs into inventory pressure costs and order processing costs, and the weak upward trend here may be due to the increasing inventory pressure cost as trading proceeds, which will not be explored further in this paper because of the small inventory pressure cost.

To quantitatively prove that spreads affect intraday price discovery, we use the share of price discovery in different periods of the day as dependent variables, and spreads and their decomposition components as independent variables. The regression results are reported in Table 5.

The regression results show that spreads significantly promote price discovery, but the decomposition component of spreads exhibits distinct effects on prices in different intraday periods. Among them, the information asymmetry component significantly promotes prices discovery in the intraday period, while the liquidity supply cost significantly inhibits price discovery. This suggests that a higher probability of informed trading and lower liquidity supply costs are conductive to facilitating price discovery. In addition, we also notice a substantial improvement in the fit of models (2) and (4) in Table 5, suggesting that the spread decomposition component provides stronger explanatory power for price discovery compared with the spread.

5.3 | Robustness test

To ensure the robustness of our experimental results, here we perform the same operation on CSI 500 stock index futures as the previous analysis did on CSI 300 stock index futures. The regressions are conducted separately on the factor influencing the price discovery differences between quotes and trades, and on the factors influencing the price discovery differences at different times of the day. The experimental results are reported in Tables B1 and B2, respectively.

Through the two tables, we can clearly see that the experimental results for CSI 500 stock index futures are generally consistent with those for CSI 300 stock index futures. This proves that our conclusions are not affected by this change in the stock index futures analyzed and that our experimental results are robust and credible.

6 | CONCLUSION

Based on tick-level data for CSI 300 stock index futures, this paper has explored the price discovery of quotes and trades using IS and PT models. Our main findings include the following aspects.

First, the price discovery contributions both of quotes and trades at low resolution are around 50%, thus not significantly differing from each other. However, as the sampling frequency increases, we find that quotes dominate in price discovery compared with trades. Second, we explored the factors influencing the price discovery difference between quotes and trades. The experimental results showed that spreads, daily returns, and volume all have different degrees of influence on these differences. Finally, we found that price discovery shows a decreasing trend within the day through the WPC method, and that the spread is an important influencing factor for the intraday price discovery discrepancy. Further, the decomposition of spreads revealed that the information asymmetry component is the main source contributing to this discrepancy, while the liquidity supply cost significantly suppresses this discrepancy. Our experimental results were unaffected by changes in the data used (CSI 300 vs. CSI 500 stock index futures).

These findings may help us to gain a deeper understanding of market microstructure and price discovery theory. Additionally, the present study has implications for investors, especially high-frequency traders, in developing quantitative strategies. It also provides empirical evidence and useful information for market regulatory structures to further improve market efficiency.

ACKNOWLEDGMENT

This study was supported by the National Natural Science Foundation of China (Nos. 72101199, 11971372, and 11801433).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable requests.

ORCID

Liwei Jin http://orcid.org/0000-0002-1043-8052
Xianghui Yuan http://orcid.org/0000-0003-1466-5268
Peiran Li http://orcid.org/0000-0002-7029-4900

REFERENCES

Baillie, R. T., Booth, G. G., Tse, Y., & Zabotina, T. (2002). Price discovery and common factor models. *Journal of Financial Markets*, *5*(3), 309–321.

Barclay, M. J., & Hendershott, T. (2008). A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance*, *15*(5), 839–849.

Barclay, M. J., & Warner, J. B. (1993). Stealth trading and volatility—Which trades move prices. *Journal of Financial Economics*, *34*(3), 281–305.

Brogaard, J., Hendershott, T., & Riordan, R. (2019). Price discovery without trading: Evidence from limit orders. *Journal of Finance*, *74*(4), 1621–1658.

Buckle, M., Chen, J., Guo, Q., & Tong, C. (2018). Do ETFs lead the price moves? Evidence from the major US markets. *International Review of Financial Analysis*, *58*, 91–103.

Chen, Y. L., & Gau, Y. F. (2014). Asymmetric responses of ask and bid quotes to information in the foreign exchange market. *Journal of Banking & Finance*, *38*, 194–204.

Chen, Y. L., & Xu, K. (2021). The impact of RMB's SDR inclusion on price discovery in onshore-offshore markets*. *Journal of Banking & Finance*, *127*. https://doi.org/10.1016/j.jbankfin.2021.106124

Cheng, L. T. W., Jiang, L., & Ng, R. W. Y. (2004). Information content of extended trading for index futures. *Journal of Futures Markets*, *24*(9), 861–886.

Davidson, J. E. H., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modeling of the aggregate time-series relationship between consumers expenditure and income in the United Kingdom. *Economic Journal*, *88*(352), 661–692.

Engle, R. F., & Patton, A. J. (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets*, *7*(1), 1–25.

Frijns, B., & Tse, Y. M. (2015). The informativeness of trades and quotes in the ftse 100 index futures market. *Journal of Futures Markets*, *35*(2), 105–126.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1), 71–100.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2009). Informed traders and limit order markets. *Journal of Financial Economics*, *93*(1), 67–87.

Gonzalo, J., & Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, *13*(1), 27–35.

Grammig, J., & Peter, F. J. (2013). Telltale tails: A new approach to estimating unique market information shares. *Journal of Financial and Quantitative Analysis*, *48*(2), 459–488.

Hagstromer, B., Henricsson, R., & Norden, L. L. (2016). Components of the Bid-Ask spread and variance: A unified approach. *Journal of Futures Markets*, *36*(6), 545–563.

Han, Q., & Liang, J. F. (2017). Index futures trading restrictions and spot market quality: evidence from the recent Chinese stock market crash. *Journal of Futures Markets*, *37*(4), 411–428.

Harris, F. H. D., McInish, T. H., Shoesmith, G. L., & Wood, R. A. (1995). Cointegration, error correction, and price discovery on informationally linked security markets. *Journal of Financial and Quantitative Analysis*, *30*(4), 563–579.

Hasbrouck, J. (1995). One security, many markets—Determining the contributions to price discovery. *Journal of Finance*, *50*(4), 1175–1199.

Hasbrouck, J. (2003). Intraday price formation in US equity index markets. *Journal of Finance*, *58*(6), 2375–2399.

Hasbrouck, J. (2021). Rejoinder on: price discovery in high resolution. *Journal of Financial Econometrics*, *19*(3), 465–471.

He, Y., Lin, H., Wang, J. B., & Wu, C. C. (2009). Price discovery in the round-the-clock US treasury market. *Journal of Financial Intermediation*, *18*(3), 464–490.

Hu, Z. P., Mallory, M., Serra, T., & Garcia, P. (2020). Measuring price discovery between nearby and deferred contracts in storable and nonstorable commodity futures markets. *Agricultural Economics*, *51*(6), 825–840.

Hung, J. C., Liu, H. C., & Yang, J. J. (2021). Trading activity and price discovery in Bitcoin futures markets. *Journal of Empirical Finance*, *62*, 107–120.

Jing, H., Feng, H., Baiao, L. C., & Li, Z. H. (2021). Price discovery and its determinants for the Chinese soybean options and futures markets. *Finance Research Letters*, *40*. https://doi.org/10.1016/j.frl.2020.101689

Karmakar, M., & Inani, S. (2019). Information share and its predictability in the Indian stock market. *Journal of Futures Markets*, *39*(10), 1322–1343.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

Lee, C. M. C., & Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance*, *46*(2), 733–746.

Li, M., & Xiong, T. (2021). Dynamic price discovery in Chinese agricultural futures markets. *Journal of Asian Economics*, *76*, 101370.

Lien, D., Hung, P. H., & Lin, Z. W. (2020). Whose trades move stock prices? Evidence from the Taiwan Stock Exchange. *International Review of Economics & Finance*, *66*, 25–50.

Madhavan, A., Richardson, M., & Roomans, M. (1997). Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies*, *10*(4), 1035–1064.

Mizrach, B., & Neely, C. J. (2008). Information shares in the US treasury market. *Journal of Banking & Finance*, *32*(7), 1221–1233.

Papavassiliou, V. G., & Kinateder, H. (2021). Information shares and market quality before and during the European sovereign debt & nbsp; crisis & nbsp. *Journal of International Financial Markets Institutions & Money*, *72*. https://doi.org/10.1016/j.intfin.2021.101334

Pascual, R., & Pascual-Fuster, B. (2014). The relative contribution of ask and bid quotes to price discovery. *Journal of Financial Markets*, *20*, 129–150.

Pascual, R., Pascual-Fuster, B., & Climent, F. (2006). Cross-listing, price discovery and the informativeness of the trading process. *Journal of Financial Markets*, *9*(2), 144–161.

Patel, V., Putnins, T. J., Michayluk, D., & Foley, S. (2020). Price discovery in stock and options markets. *Journal of Financial Markets*, *47*, 100524.

Putnins, T. J. (2013). What do price discovery metrics really measure? *Journal of Empirical Finance*, *23*, 68–83.

Rindi, B. (2008). Informed traders as liquidity providers: Anonymity, liquidity and price formation. *Review of Finance*, *12*(3), 497–532.

Sohn, S. B., & Zhang, X. F. (2017). Could the extended trading of CSI 300 index futures facilitate its role of price discovery? *Journal of Futures Markets*, *37*(7), 717–740.

Tse, Y. (2000). Further examination of price discovery on the Nyse and regional exchanges. *Journal of Financial Research*, *23*(3), 331–351. https://doi.org/10.1111/j.1475-6803.2000.tb00746.x

Verousis, T., & Gwilym, O. A. (2010). An improved algorithm for cleaning ultra high-frequency data. *Journal of Derivatives and Hedge Funds*, *15*(4), 323–340.

How to cite this article: Jin, L., Yuan, X., Wang, S., Li, P., & Lian, F. (2022). Trades or quotes: Which drives price discovery? Evidence from Chinese index futures markets. *Journal of Futures Markets*, 42, 2235–2247. https://doi.org/10.1002/fut.22368

APPENDIX A

We use the model proposed by Madhavan et al. (1997) to decompose the bid–ask spread into a component associated with information asymmetry and a liquidity supply cost component. The liquidity supply cost component here is also associated with transaction costs and can be further divided into order manipulation costs and inventory pressure costs.

The decomposed information asymmetry captures the permanent price impact of the transaction, while the liquidity supply component is related to the temporary price. We use the method proposed by Lee and Ready (1991) to discriminate the direction of trade initiation. The spread decomposition model shows the relationship between price changes and trades, and is given by

$$P_t - P_{t-1} = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + \varepsilon_t, \tag{A1}$$

where P_t is the transaction price of the security at time t, and x_t is an indicator variable for trade initiation. When $x_t = 1$, trade t is buyer initiated; when $x_t = -1$, trade is seller initiated. When a cross-trade occurs within the prevailing bid–ask spread, it can be considered as a simultaneous initiation by both buyers and sellers, in this case $x_t = 0$, and $\Pr[x_t = 0] = \lambda$ is denoted. ϕ and θ represent the liquidity supply cost and the information asymmetry component, respectively, and ρ denotes the autocorrelation coefficient of the x_t series. Let:

$$u_t = P_t - P_{t-1} - (\phi + \theta)x_t + (\phi + \rho\theta)x_{t-1}. \tag{A2}$$

Each parameter in Equation (A1) and a constant α can be calculated using the generalized method of moments (GMM), and the model-implied moment condition is expressed as

$$E\begin{pmatrix} x_t x_{t-1} - x_t^2\rho \\ |x| - (1-\lambda) \\ u_t - \alpha \\ (u_t - \alpha)x_t \\ (u_t - \alpha)x_{t-1} \end{pmatrix} = 0. \tag{A3}$$

APPENDIX B

TABLE B1 Regression results of factors influencing price discovery differences between quotes and trades for CSI 500 stock index futures

	Constant	Ln-diff	Return	Spread	Theta	Phi	Adj-R^2
Dependent variable: RCIS							
(1)	−1.5347*** (0.276)	0.1636*** (0.023)	−1.1403** (0.578)	0.1309*** (0.041)			0.231
(2)	−1.5728*** (0.267)	0.1589*** (0.022)	−0.7411 (0.568)		0.4654*** (0.117)	0.8006** (0.169)	0.280
Dependent variable: RCPT							
(3)	−1.9497*** (0.393)	0.2095*** (0.023)	−1.5843* (0.822)	0.1716*** (0.058)			0.193
(4)	−2.2503** (0.377)	0.2231*** (0.032)	−0.9387 (0.803)		0.7370*** (0.166)	1.1023*** (0.238)	0.254

Note: The sample period selected for CSI 500 stock index futures is from January 1, 2021 to January 1, 2022. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

TABLE B2 Regression results of spreads and decomposition components on intraday price discovery for CSI 500 stock index futures

	Constant	Spread	Theta	Phi	Adj-R^2
Dependent variable: Trade					
(1)	−0.0257* (0.014)	0.1576*** (0.014)			0.064
(2)	0.1285*** (0.024)		0.3183*** (0.045)	−1.0760*** (0.135)	0.204
Dependent variable: Quote					
(3)	−0.0289** (0.014)	0.1615*** (0.014)			0.067
(4)	0.1285** (0.024)		0.3157*** (0.046)	−1.0620*** (0.136)	0.200

*, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.