The Information Content of the Limit Order Book and the Corresponding Trading Strategy

William T. Lin

Department of Banking and Finance, Tamkang University

Shih-Chuan Tsai

College of Management, National Taiwan Normal University

Zhenlong Zheng[1]

Department of Finance, Xiamen University

Pei-Yau Lung[2]

Department of Banking and Finance, Tamkang University

ABSTRACT

This paper explores the relationship between orders and performance in the Taiwan futures market and aims to predict the futures price change using the order aggressiveness and information content in the limit order book. The empirical results show that the performance of market orders of TAIEX futures is significantly positive, indicating that the market orders contain information. The five quotes of the limit order book can predict the change in futures prices, especially when there is an uptrend in the market. The predictability of the change in futures prices also increases when the imbalance in the price impact between the demand and supply schedules is extreme. We also use the five quotes of the limit order book to propose a trading strategy. This trading strategy could earn positive returns even when transaction costs are taken into account.

Keywords: Information content; Limit order book; Order aggressiveness; Trading strategy

[1] Address correspondence to: Zhenlong Zheng, Dept. of Finance, Xiamen University, Xiamen, Fujian, 361005, PRC. Email: zlzheng@xmu.edu.cn

[2] Address correspondence to: Pei-Yau Lung, Tamkang University, No.151, Yingzhuan Rd., Tamsui Dist., New Taipei City 25137, Taiwan. Email: 899530108@s99.tku.edu.tw

JEL Classification: G1, G14, G17

1. Introduction

The financial market trading mechanism, based on how trading should be organized, can be divided into two categories: the quote-driven market and the order-driven market. In an order-driven market, there are no market makers and the market relies on orders to provide immediacy or liquidity. Investors can select whether to trade using market orders or limit orders based on information regarding the underlying asset. As a result, the orders convey information (Cao, Hansch, and Wang, 2008). This study examines the information content of the orders submitted by investors in the Taiwan futures market to benefit predictions of future price changes in that market.

The advantages of utilizing market orders include a high probability of transactions taking place. However, the transaction price may increase with market uncertainty, which may subsequently increase the investor's transaction costs. In comparison, limit orders limit the transaction price to within a certain range, as expected by the investor. Thus, the investor can set an ideal price based on the information available. However, limit orders may require longer transaction times, or may even not occur if the market price does not reach the price range specified in the limit order. Previous studies typically assumed that investors trading with information favor market orders to earn higher returns, whereas investors trading without information favor limit orders, playing the role of liquidity traders in the market.[3] However, Bloomfield, O'Hara, and Saar (2005) suggested the adoption of a theoretical model where informed traders used both limit and market orders. Subsequently, Anand, Chakravarty, and Martell (2005) used the market mid-quote when the investor submitted an order and the mid-quote 5 min and 60 min after the order was submitted to gauge the order performance. They found that limit orders submitted by institutional investors outperformed limit orders submitted by individual investors, further indicating that limit orders contain information. Therefore, this study investigates whether the investors had actual information when they issued the order, or whether they were uninformed noise traders who believed they had information.

Easley and O'Hara (1992) stated that institutional investors can obtain beneficial information more easily than individual traders. Therefore, when institutional traders make a purchase, it indicates a promising future market. Conversely, when institutional investors sell, it indicates that a downward market should be expected in the future. This trading behavior causes the number of transactions to suddenly increase; thus private information results in increased trading

[3] The related literature includes Glosten (1994), Rock (1996), and Seppi (1997).

volume. Cao, Hansch, and Wang (2009) indicated that when the imbalance between demand and supply in the market becomes more severe, the likelihood of informed trading increases. When an investor submits a limit order, it is entered into the limit order book before the transaction; simultaneously, the market reveals the different prices that have not yet been filled according to the order price to enable the investor to understand the current market situation. Therefore, this study observes the orders submitted by investors and predicts future price changes by investigating the best five quotes in terms of price and volume as revealed in the limit order book.

In addition, this study further classifies investors into the following categories: individuals, dealers, foreign institutions, and other domestic institutions. We also investigate the order patterns for different categories of investors to know whether any differences in informativeness exist among them. Lee, Lin, and Liu (1999) divided investors into three categories: institutional investors, small individual investors, and large individual investors. They then verified which category of investors consisted of informed investors in the Taiwan stock market using the VAR model. The results revealed that small individual investors were not informed traders, big individual investors were the most informed traders, and institutional investors were somewhere between the two. They also discovered that informed traders tend to place orders of large quantities, a finding which was consistent with the results reported by Easley and O'Hara (1987). Anand and Martell (2001) examined the difference in performance between limit orders submitted by informed and uninformed traders. The results showed that after controlling the characteristics of the order, the price trend after the transaction was beneficial to limit order traders. In addition, limit orders submitted by institutional investors showed a better performance than those submitted by individual investors. Therefore, they inferred that when institutional investors obtain extremely beneficial private information on the value of the securities, they use limit orders to maximize their trading profits. This is because institutional investors hold more information that is beneficial to predictions and the use of limit orders reduces the risk of uncertain trading prices.

Lee, Liu, Roll, and Subrahmanyam (2004) further categorized investors on the Taiwan Stock Exchange as either large or small investors in the three major categories according to the number of orders each investor submitted. Order patterns were also analyzed to determine which categories of investors were informed traders. The results showed that, compared to investors in other categories, domestic institutional investors can profit from immediate transactions and their trading strategies can effectively reduce the prices' impact. Therefore, institutional investors were categorized as informed traders. Barber, Lee, Liu, and Odean (2009) divided investors into aggressive investors and passive investors according to the pricing of their placed orders and verified their profits in different holding periods. The empirical results showed that institutional investors achieved better performance than that of individual investors. The more aggressive an individual investor was in submitting orders, the poorer the performance. This result suggests that institutional investors hold more information compared to individual investors.

The empirical results in this study show that market orders can obtain a significant and positive return 5 minutes after they are submitted, regardless of whether the order is submitted by institutional or individual investors. This shows that market orders contain temporary information. Investors who hold temporary information tend to use market orders for immediate transactions and profits. Information revealed from limit orders, such as the order quote (length) and amount (length), can be used to predict future price changes. When the market condition is also considered, the information revealed in the limit order book is more beneficial for predictions than the information for the entire sample period when the market exhibits an upward trend. However, when the market exhibits a downward trend, the ability of information from the limit order book to predict is lower than that of the information for the entire sample period. When the imbalance of the price impact between supply and demand is extreme, the prediction ability of limit order book increases significantly. This shows that at times of market supply and demand imbalance, informed traders are more likely to exist. Therefore, this study aims to develop trading strategies that facilitate a positive return, after accounting for the price slippage and transaction costs, based on information revealed in the limit order book.

The next section introduces the information and empirical models used in this study. The third section presents the empirical test results, and the fourth section shows the trading strategy and its profitability based on the empirical results provided in the third section. Section Five presents the conclusion.

2. Data and Methodology

2.1 Data

Our data comprise Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) futures contracts sourced from the Taiwan Futures Exchange (TAIFEX). The sample period extends from January 2, 2008 to September 30, 2009. The trading units of the TAIEX futures are the index value of the TAIEX multiplied by NT$200. Compared to other contracts, nearby contracts are the most liquid and contain more information. Therefore, we use the nearby contracts in our analysis. We also exclude data from three days prior to the contract expiry date to eliminate the expiration effect.

Our data set includes detailed order flow, order book, and transaction data. For each order, the data include the order arrival date and time, the order direction (buy or sell), the price and quantity, the order type (limit or market order), the order code, the account identification details, and trader type. The transaction data include the transaction date and time, the transaction direction (buy or sell), the quantity, the order code, the transaction price, and the traders' identity.

The account identification details and order code allow us to trace the trading activity of each account by matching the orders and transactions. Orders with price that differ significantly from the market price have a low probability of being executed and contain less information. Therefore, we only selected the orders that were completely executed or partially executed to ensure that our sample contained more information.

2.2 Order performance

To determine whether an investor is an informed or uninformed trader, this study first categorizes the investor as either an aggressive or passive investor based on the order price he or she submitted. The performance of the order after transaction is then considered. If the investor submits a buying order and the order price is higher than the best price revealed in the market at the time, the order is considered to be aggressive, while the opposite order behavior is regarded as passive. The aggressiveness of selling orders is also defined in a similar manner.

The performance of orders is measured using the calculation employed by Anand, Chakravarty, and Martell (2005). The formulae are as follows:

$$\textit{Performance of purchase orders} = m_t - m_{t+j} \quad (1)$$

$$\textit{Performance of sell orders} = m_{t+j} - m_t \quad (2)$$

where m_t is the quote midpoint at the time that the order was submitted. The measuring principle mainly uses the price reversal amount after the order was submitted. This study calculates the price reversal at 15 s, 30 s, 1 min, and 5 min.

2.3 Forecasting future returns

This study uses the best five bid and ask prices and quantities from the limit order book to observe the order behavior of various investors and predict price changes in the futures market to identify opportunities for earning profits. This study employs the model used by Cao, Hansch, and Wang (2009) to predict the returns of future price. The price difference between each quote is considered to be the order book's height, and the market depth of each quote price is considered to be its length. We believe that the more severe the imbalance between the buyers and sellers, the higher is the likelihood of informed trading occurring. The concept of an imbalance within the limit order book is used to predict the future returns.

Cao, Hansch, and Wang (2009) studied the limit order book from the perspective of market supply and demand; therefore, they viewed orders as market demand, with P^d_j and Q^d_j representing the order price and demand quantity for the *j*th quote, respectively. Conversely, P^s_j and Q^S_j represent the best selling price and supply quantity for the *j*th quote. Furthermore, Δu^d_j and Δn^s_j are defined as $(P^d_j - P^d_{j-1})$ and $(P^s_j - P^s_{j-1})$, where j is 2…5.

This study calculates the daily returns of futures contracts every 15 seconds.[4] In addition, this study uses the selling and buying price midpoints for the best quote instead of the futures transaction price to calculate the return and prevent the results from being affected by missing data or transactions mispriced by investors (Anand et al., 2005). The calculation formula for the return is as follows:

$$m_t = \frac{P^s_1 + P^d_1}{2} \tag{3}$$

$$mid_t = \ln\left(\frac{m_t}{m_{t-1}}\right) \tag{4}$$

where mid_t is the mid-quote return. This study considers that the returns may exhibit autocorrelation; therefore, we have configured the AR(5) model based on the principle of the smallest AIC[5]. The residual ε of the model is then used as the unpredicted mid-quote return. This study also uses the relative spread as control variables, which are defined using the following:

$$spread = \frac{P^s_1 - P^d_1}{P^s_1 + P^d_1} \tag{5}$$

When setting the order book height and length, previous studies considered whether the buying or selling aspect of the market had greater market depth, encouraging investors to issue more aggressive orders.[6] For example, if the buying side of the futures market has greater market depth, the investors are aggressive when issuing buying orders, therefore significantly increasing the likelihood of raising the futures price. This concept connects investors' aggressiveness in submitting orders with the direction of the market price movement. Based on a similar concept, Cao, Hansch, and Wang (2009) reported that *HR* and *QR* can capture the aggressiveness of orders using the following formulae:

[4] This study also examined return rates in the 5 s, 30 s, 1 min, and 5 min periods. We found that calculations in 15 s periods provided the best predictions of return rates.

[5] Akaike Information Criterion.

[6] Biais, Hillion, and Spatt (1995), Cao, Hansch, and Wang (2008), Foucault (1999), Griffiths, Smith, Turnbull, and White (2000), Hollifield, Miller, and Sandås (2004), Parlour (1998), and Ranaldo (2004).

$$HR_j = \frac{(P_j^s - P_{j-1}^s) - (P_{j-1}^d - P_j^d)}{(P_j^s + P_{j-1}^s) + (P_{j-1}^d - P_j^d)} \quad j = 2 \cdots 5 \tag{6}$$

$$QR_j = \frac{Q_j^s - Q_j^d}{Q_j^s + Q_j^d} \quad j = 1 \cdots 5 \tag{7}$$

where *HR* represents the revealed price differences between different quotes. The price difference of the buyers should be smaller when the investors submit orders more aggressively; this signifies that greater competition exists among the buyers. In these instances, the *HR* value is positive, simultaneously predicting a greater chance of an increase in subsequent futures prices. The opposite scenario indicates that there exists a greater chance of a decline in futures prices. *QR* predicts the direction of movement in futures prices from the perspective of market supply and demand. When the buying side has greater depth, more investors want to buy futures contracts. Under such conditions, the *QR* value is positive, and the price after the transaction increases. Conversely, the price decreases in the opposite scenario. The model configuration below summarizes these variables:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{5} \beta_j HR_{j,t-1} + \sum_{j=2}^{5} \gamma_j QR_{j,t-1} + \eta_t \tag{8}$$

During the empirical analysis in this study, variables from different quotes are progressively integrated into the model. After the model is adjusted, tests are conducted to identify increases in the R-square to assist in understanding whether the quotes in the limit order book contain additional information. Thus, the best quote from the limit order book is slowly added to the subsequent models for the empirical analysis. Equation (8) is then rewritten after the status of the limit order book is obtained by observing the order aggressiveness and order quantity of both the buying and selling sides. The new empirical model is shown below:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=1}^{5} \gamma_{d,j} Q_{j,t-1}^d + \sum_{j=1}^{5} \gamma_{s,j} Q_{j,t-1}^s + \sum_{j=2}^{5} \beta_{d,j} \Delta p_{j,t-1}^d + \sum_{j=2}^{5} \beta_{s,j} \Delta p_{j,t-1}^s + \eta_t \tag{9}$$

Intuitively, when no significant imbalance exists between the supply and demand, the opportunities for informed trading are minimal. Conversely, the chances of informed trading are significant in the opposite scenario. Therefore, this study measures the scope of imbalance through price impact variables. The key assumption of the calculation method is that, for every match, there are some market orders that deplete the limit orders on the other side, of which m_1 and m_2 are the limit order book quotes where the market orders have depleted the limit orders. The greater the market depth on the other side, the lower the price impact. This also signifies relatively good liquidity at that particular point in time. Therefore, the price impact is a preconceived concept to calculate the necessary trading costs when q units of futures are sold or bought. The equations are as follows:

$$LD(q) = \frac{(P_1^s + P_1^d)}{2} - \frac{\sum_{j=1}^{m_1-1} P_j^d Q_j^d + P_{m_1}^d Q_{m_1}^d}{q} \qquad q = \overline{Q}、1.5\overline{Q} \cdots 3\overline{Q} \tag{10}$$

$$LS(q) = \frac{\sum_{j=1}^{m_2-1} P_j^s Q_j^s + P_{m_2}^s Q_{m_2}^s}{q} - \frac{(P_1^s + P_1^d)}{2} \qquad q = \overline{Q}、1.5\overline{Q} \cdots 3\overline{Q} \tag{11}$$

$$LR(q) = \frac{LS(q) - LD(q)}{LS(q) + LD(q)} \qquad q = \overline{Q}、1.5\overline{Q} \cdots 3\overline{Q} \tag{12}$$

where LD and LS represent the price impact of the demand and supply sides, respectively, and $\overline{Q}$ is the average futures quantity traded within the quote. This study attaches a different measuring weight to the average price impact value. Therefore, the greater the LD, the lower the market demand and the more likely the price will increase in the next quote. Conversely, the greater the LS, the more likely the price will decline in the next quote. The variable LR summarizes the price impact of both the market supply and demand. This study selects a time with extreme imbalance and analyzes whether the model can predict price changes more accurately under such market conditions.[7]

Considering the price impact and to predict the movement of the mid-quote, a model for buying and selling orders is used. The model is as shown below:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=2}^{5} \beta_{d,j} LD(j\overline{Q}/2)_{t-1} + \sum_{j=2}^{5} \beta_{s,j} LS(j\overline{Q}/2)_{t-1} + \eta_t \tag{13}$$

3. Empirical Results

Table 1 shows the empirical results for order performance prediction using the method employed by Anand, Chakravarty, and Martell (2005). In Panel A, the performance of orders that were partially or entirely executed is calculated. The results show that the predictions for the performance of orders, regardless of whether they are submitted by institutional or individual investors, are all significantly positive, whereas the predictions for limit orders are all significantly negative. Because an investor's motives cannot be identified through their orders, predicting an investor's realized profit and loss according to the performance of limit orders is impossible.

[7] LR of the greatest and least 5% of the sample.

However, as regarding the positive performance of market orders, this study contends that the market orders contain information within a short period of time. Therefore, temporary information can be captured at times when investors are most likely to place market orders. The information obtained can then be used to buy or sell futures to profit from price reversal.

Panel B consists of executed day trade orders because day trading is more likely to be arbitraged and to contain information. The results show that market orders in day trading perform better than other transacted orders. This verifies that day trade orders contain more short-term information. The predicted performance indicates that, in the Taiwan Futures Market, market orders possess short-term information. This is consistent with the conclusion reported by Rock (1996), namely, that informed traders with temporary information are inclined to issue market orders for immediate transactions.

Table 2 shows the analysis results of the effects of unbalanced order book height (*HR)* and length (*QR)* on future returns. Order book height refers to the price difference between the buying and selling sides. A small price difference indicates the aggressiveness of the investors' orders. In other words, when investors place aggressive orders, the price order for the quotes should be relatively close. If the *HR* is positive, the buying orders are more aggressive and the likelihood of a positive return on the next quote is higher. Conversely, if the *HR* is negative, the selling side is more aggressive and the likelihood of a negative return on the next quote is higher. The imbalance in the measurement of the order book length is primarily based on measuring the difference in market depth between the buying and selling sides. If the *QR* is positive, the selling orders have a greater quantity currently and a greater chance for aggressive orders from the sellers in the future, thereby creating a higher likelihood of a negative return in the next quote. Conversely, if the *QR* is negative, the quantity on the buying side is currently greater and has a bigger chance of more aggressive orders from buyers in the future. This means that the likelihood of positive returns occurring in the next quote is higher. In addition, this study adds the information revealed by the order book into the regression quotes individually to examine whether the order book from different quotes contained extra information.

The adjusted R-square shown in Table 2 indicates that when the variables from the best quote are used, the value is 8%, when variables from the second quote are added, the explanatory power increases by only 0.13% to 8.13%, and when five quotes of revealed information are added, the explanatory power is 8.24%, only 0.24% more than the value obtained using only the best quote. Therefore, the height and length information from the best quote in the order book explains most of the information. From the perspective of the coefficient from the regression model, both the *QR* and *HR* in the first and second quotes produced the expected results. However, the signs reveal the opposite phenomenon from the third quote onwards. Nevertheless, the transaction rate in these quotes is relative and, therefore, does not have a significant effect on the price.

Table 3 shows the price changes as predicted by the height and length of the market supply and demand. Q^d and Q^s represent the order quantity of the demand and supply, respectively. A greater Q^d suggests the possibility of attracting more aggressive orders from the demand side in the future and, thus, predicts a positive relationship between Q^d and the return. Conversely, the greater the Q^s is, the greater the supply quantity is; this leads to more aggressive orders from the supply side and can be used to predict a negative relationship between Q^s and the return. Δa^s is the selling price difference; smaller differences suggest more aggressive selling orders, which leads to a positive relationship between Δa^s and the return.

The results in Table 3 show that the adjusted R-square is 8.37% when the best quote is used; the coefficient increases to 16.04% with the addition of the second quote; and after the implementation of information from all five quotes, the adjusted R-square is 16.34%. We discovered that the first two quotes contain more information when the height and length or market supply and demand are used to predict price changes. We also discover that the significance of the coefficients matches the expected direction. The demand height in the order book, however, shows a significantly lower coefficient value than that of the supply. This indicates that buying orders reveal more information than selling orders.

The results in Table 3 show that the effects produced by the order book height of the supply and demand sides differ. In an empirical study, Ranaldo (2004) discovered that the aggressive order behavior on the buying and selling sides is asymmetrical, which suggests that investors have different order behaviors depending on whether the price is increasing or decreasing. The disposition effect proposed by Shefrin and Statman (1985) also explains that investors hope to sell the shares as quickly as possible when they are earning profits to retain the profit; however, when a loss occurs, investors are reluctant to sell and tend to wait for a price reversal. The method used in this study is based on a method employed by Ranaldo (2004) to understand whether the investors in the futures market have different responses to buying and selling behaviors. The return from the previous quote is divided into increasing and decreasing trends. If the return from the previous quote is positive, then an upward trend exists; the opposite suggests a downward trend. Equations (8) and (9) are then repeated for the empirical analysis.

Tables 4 and 5 show the empirical results of the upward trend derived using equations (8) and (9). The adjusted R-square indicates that the model has more prediction ability than the entire sample body during an upward trend. Using the variables of the best quote in Table 4 as an example, when the entire sample body is implemented, the model possesses only 8% explanatory power. However, the explanatory power rises to 48% during an upward trend. Because the information from the order book is concentrated in the first two quotes, the regression coefficient also matches expectations. Tables 6 and 7 show the empirical results of equations (5) and (6). The explanatory power of the regression indicates that the explanatory power of the model is significantly smaller than that of the entire sample body during downward trends. For example, when the variables from the best quote are used in Table 6, the explanatory power of the model decreases to 3.76% from 8%. However, the majority of the information in the order book remains concentrated in the first two quotes, and the regression coefficient still matches expectations.

This study then utilizes the imbalance between the price impacts of market supply and demand to predict the return for the next quote. Using the demand side as an example, the price impact is defined as follows: if the supply is hypothesized to provide $\overline{Q}$, the number of market orders, which subsequently consume the limit orders from the demand side and produce changes in transaction prices, and if the demand quantity is significant enough during the best quote, the transaction price should be consistent with the demand price for the best quote. Conversely, the transaction price increases if the demand is insufficient. In other words, the price impact measures the liquidity of supply and demand; the greater the price impact, the poorer the liquidity. This also signifies a greater chance for a price drop in the next quote and a negative relationship between the price impact and return. The opposite situation produces a positive relationship between the price impact and return of the suppliers.

In addition, $\overline{Q}$ is given a different weight of $\frac{j}{2}$ and reintegrated into the regression to identify the most suitable expected quantity of market orders. Because *LD* and *LS* are primarily based on whether the limit orders can satisfy the consumption of market orders, if more limit orders are issued, the effect of price impacts is reduced. Therefore, the price impacts can be considered to measure the quality of market liquidity. Therefore, when more limit orders are issued, the *LD* coefficient is positive; the opposite scenario results in a negative *LS*. Table 8 displays the expected market order quantity after different weights are integrated. Because the explanatory power only increases by a total of 0.03% from weights 1 to 5, the results show a limited ability to increase the return for the next quote. Therefore, this study considers using the average quantity of brokered transactions to be sufficient for predicting the market order quantity in the next quote.[8]

When the imbalance between market supply and demand becomes more severe, the chance of information existing in the market increases. Subsequently, price changes during such times are more dramatic. Thus, the model can provide a better prediction. This study uses the imbalance between the price impact of market supply and demand (*LR*) to differentiate the market equilibrium conditions. The greatest and smallest 5% of *LR* were selected as samples to incorporate into equations (8) and (9) to assess whether the model provides better predictions when the market is imbalanced. The results in Table 9 show that when measuring the imbalance in order book height, only the coefficient from the second quote reaches the 10% level of significance. However, measurements for imbalance in the order book length are significant for the first 4 quotes. From the adjusted R-square in the model, we can infer that the majority of the information from the order book is concentrated in the best quote. Table 10 shows that the heights are insignificant in both the supply and demand order book. This may be because the increasing trading volume in the futures market and the rapid speed of delivery of the prices are more significant in a severely imbalanced market. Subsequently, the minimum price variation in the futures cannot completely reflect the aggressiveness of the investors' orders. The aspect of the order book length indicates that all five quotes are significant and consistent with the expected direction. From the aspect of the model's adjusted R-square, one quote from the order book contains the majority of the information. Therefore, when the price impacts between the supply and demand are severely unbalanced, the majority of information is concentrated in one quote, and the R-square increases significantly. This shows that when the market is extremely imbalanced, the chance of informed trading is higher. When the liquidity is more uneven, the transaction cost increases, resulting in a tendency among informed traders to trade using limit orders. This is consistent with that reported by Kaniel and Liu (2006).

4. Trading Strategies

In the previous section, we discussed how an order book with height (*HR*) and depth (*QR*) possesses explanatory power for future returns. Based on this conclusion, trading strategies can be created using the length and height variables of the limit order book and the spread in the best buying and selling quotes in individual futures contracts throughout the entire market. In addition, to avoid abnormal significance in statistical testing and to be more reasonable in practice, this study configures the return with the AR(1) model and uses the residual ε as the explained variable. This

[8] This study used the expected market order quantity of different weights and found no significant changes.

enables the model to predict the return on the next quote more accurately, making this trading strategy feasible. Changes in the *QR* and *HR* in the lagged period, the spread in the best buying and selling quote, and the return on the previous quote are used as explanatory variables. The model used for the trading strategy is as follows:

$$\begin{aligned}\varepsilon_t = {} & \alpha_0 + \beta_0 QR_{t-1,1} + \beta_1 HR_{t-1,1} + \beta_2 QR_{t-1}^{up,1} + \beta_3 QR_{t-1}^{down,1} + \beta_4 HR_{t-1}^{up,1} + \beta_5 HR_{t-1}^{down,1} \\ & + \beta_6 Spread_{t-1,1} + \beta_7 Spread_{t-1}^{up,1} + \beta_8 Spread_{t-1}^{down,1}\end{aligned} \tag{14}$$

Of which,

$$QR^{up,1} = \begin{cases} 1, if\ QR_t - QR_{t-1} > 0 \\ 0, else \end{cases} \quad , \quad QR^{down,1} = \begin{cases} 1, if\ QR_t - QR_{t-1} < 0 \\ 0, else \end{cases}$$

$$HR^{up,1} = \begin{cases} 1, if\ HR_t - HR_{t-1} > 0 \\ 0, else \end{cases} \quad , \quad HR^{down,1} = \begin{cases} 1, if\ HR_t - HR_{t-1} < 0 \\ 0, else \end{cases}$$

$$Spread^{up,1} = \begin{cases} 1, if\ spread_t - spread_{t-1} > 0 \\ 0, else \end{cases} \quad , \quad Spread^{down,1} = \begin{cases} 1, if\ spread_t - spread_{t-1} < 0 \\ 0, else \end{cases} \tag{15}$$

Using this model to conduct regression analysis, calculate the expected return, and compare the expected return with the transaction costs increases the number of bullish or bearish signals. The investor can then trade according to the signals and generate returns.[9] After calculating the expected return on the current quote, if the expected return exceeds the transaction costs and is a positive value, the bullish signal is established and the investor can purchase futures contracts in the following quote. If the expected return for the following quote continues to be greater than zero, the bullish signal remains, and the investor can hold the futures. Because the transaction cost has already been considered at the time of purchase, futures can be held if the expected return is greater than zero; this saves transaction costs by eliminating excessive buying and selling actions. The futures can be held until the expected return on the next quote declines below zero, when the futures should be sold. Conversely, if the expected return on the next quote is negative and lower than the transaction cost, the bearish signal is established. This signal prompts the investor to sell the futures contracts. If the expected return on the following quote is still less than zero, the bearish signal remains. When the expected return on the next quote becomes greater than zero, futures can be purchased.

In practice, because only past information can be used to predict future price

[9] Please see the appendix for details of the transaction costs considered in this study.

changes, the regression for the entire sample period in the previous section cannot be applied to a practical trading strategy. Therefore, this study divides the quote into either one-day, two-day, and three-day quotes for the regression. In other words, the information from the past one day, two days, and three days is integrated into the regression after the regression coefficient from the following day to calculate the expected return and the adjusted R-square. This study investigates which quote has the best R-square and the greatest residual after integrating the regression coefficients from different quotes into the current quote. This quote can be used as the base quote for prediction and can form the basis of the trading strategy.

Table 11 is the regression analysis with different day-units. The results in the table indicate that the adjusted R-square decreases with an increase in the day-unit regardless of the type of futures contract. Because the R-square of the *t-1* day regression is the highest, using the regression coefficient on the previous day to calculate the return for the current rate is the most accurate approach.

This study also calculates the average coefficient of regressions with different day-units and compares the acquired average coefficient with the average coefficient of the current day regression. The closer the absolute values of the average coefficients are, the more stable the model, and the more accurate the expected return of the current quote derived using the regression coefficient of the previous quote is. The residual difference represents the current-quote regression residual minus the previous-quote regression residual. The smaller the difference, the closer is the regression error term and the more stable is the model. Table 11 also shows the residual calculated using the regression coefficient from the previous day with information from the current quote. When the residual from the current regression is at its smallest level, the model is the most secure. Therefore, this study uses the information from the previous day to conduct regression analysis and integrate the acquired coefficient with current information to calculate the expected return.

Table 12 is the regression analysis of the trading strategy. The regression results are primarily consistent with our expectations. The coefficient direction of the limit order book length is the opposite of the return; the coefficient direction of the order book height is consistent with the return; and the spread and return exhibit a positive relationship. When spread changes increase, the buying price of the best quote decreases or the selling price of the best quote increases, resulting in an increase in returns.

Table 13 shows the returns of the trading strategy. After obtaining the transaction and price slippage, only the bullish strategy of the major and MiNi-TAIEX Futures are found to result in negative returns; the other futures contracts all result in positive returns regardless of whether the bullish or bearish strategy is used. During the study, the number of TAIEX Futures contracts traded using the bullish strategy was 26,300, with an average holding period of 2.3 days. The quantity of contracts traded using the bearish strategy was 22,782, with an average holding period of 2.28 days. The quantity of Electronics Sector Index Futures contracts traded using the bullish strategy was 30,823, with an average holding period of 3.11 days. Another 30,865 futures contracts were traded using the bearish strategy, with an average holding period of 3.5 days. Finance Sector Index Futures contracts traded using the bullish strategy totaled 30,196, with an average holding period of 3.14 days. A further 30,284 contracts were traded using the bearish strategy, with an average holding period of 4.22 days. MiNi-TAIEX Futures contracts traded using the bullish strategy totaled 8,722, with an average holding period of 2.35 days; the quantity of contracts traded using the bearish strategy totaled 7,112, with an average holding period of 1.96 days.

This study also calculated the return obtained after subtracting the transaction and price slippage, as well as the success rates for both the bullish and bearish strategies. As shown in Table 13, regardless of the strategy type, the success rate is the highest for the Finance Sector Index Futures and Electronics Sector Index Futures, followed by TAIEX Futures, and then MiNi-TAIEX Futures. Although the success rate for using the bullish strategy for the TAIEX Futures is higher than 50%, the return obtained for a successful condition is less than the return obtained for an unsuccessful condition. In comparison, MiNi-TAIEX Futures have a success rate of less than 50% both when the bullish or bearish strategies are used. However, for successful conditions, the return generated is significantly greater than the return generated for unsuccessful conditions. The quantity of MiNi-TAIEX Futures traded is less, which is the reason for the higher return after the transaction costs are subtracted.

As shown in the trading strategy return table, the model generated by this study produced a better performance using the bearish strategy for TAIEX Futures and MiNi-TAIEX Futures. In comparison, because the bullish strategy may involve a higher number of trades, the accumulated return is easily cancelled by the transaction costs. However, the bullish strategy provides a positive return if transaction costs are not considered. This suggests that the model can predict the returns in the next period. The average holding periods and the transaction quantities of Electronics Sector Index Futures and Finance Sector Index Futures are higher, signifying that the returns from these futures are more persistent, and the prediction ability of the model is greater than when used with TAIEX Futures and MiNi-TAIEX Futures.

5. Conclusion

This study first categorized investors as informed investors with actual information or noise investors by observing the order behavior pattern of investors. The method suggested by Anand, Chakravarty, and Martell (2005) was used to understand the relationship between the daily order behavior of investors and their order performance. We found that investors who placed market orders were able to receive significant and positive returns at 15 s, 30 s, 1 min, or 5 min after the order was placed. The investors were then divided into institutional investors and individual investors for comparison purposes. Because both groups still showed a significant and positive market order performance, this suggests that market orders contain temporary information. This is consistent with the suggestion made by Rock (1996) that informed traders with temporary information are inclined to place market orders to engage in immediate transactions.

The revealed quote (height) and order quantity (length) in the futures market limit order book are integrated into the model used by Cao, Hansch, and Wang (2009) to predict the price change in the following quote. The depth and length of different quotes were individually calculated and reintegrated into the regression model. The explanatory power of the model was used to investigate the amount of information in different quotes. The results show that under normal circumstances, and considering the imbalance between the height and length of supply and demand, only the revealed information for the best quote in the limit order book possesses higher explanatory power for predicting future price changes. However, if separate considerations of the supply and demand height and length are obtained, this indicates that the first two quotes in the order book possess explanatory power for predicting future price changes. A comparison of the explanatory power of the two models shows that separate consideration of the supply and demand height and length results in greater explanatory power than that derived by calculating the imbalance between height and length. We also discovered that when the market was divided into upward and downward trends, the model during upward trends had more explanatory power than the entire sample, whereas the opposite was true during downward trends. This finding is consistent with that reported by Ranaldo (2004) where an asymmetrical phenomenon exists between the buying and selling sides of the market.

This study then used the imbalance in the price impact to measure the imbalance between market supply and demand. When market imbalance becomes more severe, the chance of informed trading in the market increases. Therefore, the greatest and smallest 5% of the price impact imbalance were selected as samples for verification. The results showed that in a severely imbalanced market, the height of the limit order book does not possess prediction ability. This may be because the tick size of futures in the Taiwan Futures Market cannot reflect the order behavior of investors in a severely imbalanced market. However, the explanatory power of the model showed that the model's prediction ability increased significantly in an imbalanced market. This demonstrates that when a severe imbalance occurs between the market supply and demand, the chance for informed traders to be present in the market is higher.

The second part of the study used the variables obtained in the first part, along with the variables for changes in limit order book height and length in previous and following quotes, to predict the return on the following quote and establish trading strategies. The results of the trading strategies show that without considering the transaction cost, all four types of futures contracts generated returns greater than zero, regardless of whether the bearish or bullish strategies were used. This shows that the model can predict future returns. When the actual price slippage and transaction costs were considered, only the bullish trading of TAIEX Futures and MiNi-TAIEX Futures generated returns less than zero; the returns for other futures and strategies remained positive. This suggests that this model can be used to provide a practical tool for investors to earn profits.

Appendix: Transaction Cost

To establish trading strategies that are as practical and realistic as possible to earn profit from actual transactions, the calculation of transaction costs was considered. The transaction costs include the securities dealers' handling fees, futures tax, and price slippage. This appendix introduces the calculation of the three transaction costs mentioned in this study and integrates them into the returns to obtain a result as close to reality as possible.

1. Handling fees and futures tax

Table A1 shows the transaction costs associated with the four types of futures contracts: TAIEX Futures, Electronics Sector Index Futures, Finance Sector Index Futures, and MiNi-TAIEX Futures. The figures include the futures tax (4/100000) and handling fee of the securities dealers. For the convenience of calculation, this study set the total average transaction price as the basis for calculating the futures tax because the futures tax increases and decreases according to the contract value.

2. Price slippage

When trading strategy models produce a buy or sell signal, the investor must buy or sell immediately, making the most aggressive order. Immediate transactions are guaranteed if the buying price is lower than the selling price of the best quote and the selling price is lower than the buying price of the best quote.

Large orders can cause the market price to increase or decline and partial orders may not be transacted in the best buying and selling quote. The remaining quantities can be transacted at a level higher than the selling price of the best quote or lower than the buying price of the best quote, thereby resulting in an increase in cost; this is referred to as a price slide cost. To ensure the results are as close to the market condition as possible, the returns from trading strategies were calculated with the price slide cost.

The buying price is the selling price of the best quote plus one tick size of futures contract.

The selling price is the buying price of the best quote minus one tick size of futures contract.

The calculation of the return including the price slippage is shown below:

$$R_{jt} = \ln\left(\frac{P_{j,t}^{d} - k}{P_{j,t-1}^{s} + k}\right) \times 100 \tag{A.1}$$

where $P_{j,t}^{d}$ is the buying price of the best quote, $P_{j,t-1}^{s}$ is the selling price of the best quote in the futures contract, and *j*=0, 1, 2, and 3 represent the contract futures of TAIEX Futures, Electronics Sector Index Futures, Finance Sector Index Futures, and MiNi-TAIEX Futures, respectively. *k* = tick size of different futures contracts, with TAIEX Futures being 1, Electronics Sector Index Futures being 0.05, Finance Sector Index Futures being 0.2, and MiNi-TAIEX Futures being 1.

References

Anand, A., Chakravarty, S., & Martell, T. (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets, 8*(3), 288-308.

Anand, A., & Martell, T. (2001). 'Informed' limit order trading. Working paper.

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2009). Just how much do individual investors lose by trading? *Review of Financial Studies*, *22*(2), 609-632.

Biais, B., Hillion, P., & Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris bourse. *Journal of Finance, 50*(5), 1655-1689.

Bloomfield, R., O'Hara, M., & Saar, G. (2005). The 'make or take' decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics, 75*(1), 165-199.

Cao, C., Hansch, O., & Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets, 29*(1), 16-41.

Cao, C., Hansch, O., & Wang, X. (2008). Order placement strategies in a pure limit order book market. *Journal of Financial Research, 31*(2), 113-140.

Easley, D., & O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance, 47*(2), 577-605.

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics, 19*(1), 69-90.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets, 2*(99), 134.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *Journal of Finance*, *49*(4), 1127-1161.

Griffiths, M. D., Smith, B. F., Turnbull, D. A. S., & White, R. W. (2000). The costs and determinants of order aggressiveness. *Journal of Financial Economics,56*(1), 65-88.

Hollifield, B., Miller, R. A., & Sandås, P. (2004). Empirical analysis of limit order markets. *Review of Economic Studies*, *71*(4), 1027-1063.

Kaniel, R., & Liu, H. (2006). So what orders do informed traders use? *Journal of Business*, *79*(4), 1867-1913.

Lee, Y. T., Lin, J. C., & Liu, Y. J. (1999). Trading patterns of big versus small players in an emerging market: An empirical analysis. *Journal of Banking & Finance*, *23*(5), 701-725.

Lee, Y. T., Liu, Y. J., Roll, R., & Subrahmanyam, A. (2004). Order imbalances and market efficiency: Evidence from the Taiwan stock exchange. *Journal of Financial and Quantitative Analysis, 39*(2)

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies, 11*(4), 789-816.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets, 7*(1), 53-74.

Rock, K. (1996). The specialist's order book and price anomalies. *Review of Financial Studies*, *9*, 1-20.

Seppi, D. J. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies, 10*(1), 103-150.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, *40*(3), 777-790.

Table 1. Performance prediction of different investors

Panel A and Panel B show the expected transacted order and day trade order performance, respectively. The method of prediction used is to primarily compare the quote midpoint at 15 s, 30 s, 1 min, and 5 min after the order is placed. For buying orders, the calculation is the buying and selling quote midpoint of the best quote when the order is placed, minus the best buying and selling quote midpoint of the best quote some time after the order is placed. For selling orders, the calculation is the buying and selling quote midpoint of the best quote some time after the order is placed minus the buying and selling quote midpoint of the best quote when the order is placed.

	15 s	30 s	1 min	5 min
Panel A Transacted orders				
Market orders				
Institutional investors	1.054***	1.135***	1.202***	1.324***
Individual investors	0.883***	0.958***	1.091***	1.161***
Limit orders	-1.545***	-1.599***	-2.009***	-2.500***
Panel B Day trade orders				
Market orders				
Institutional investors	2.689***	2.837***	2.787***	2.310***
Individual investors	1.412***	1.542***	1.606***	1.679***
Limit orders	-1.415***	-1.430***	-1.691***	-1.758***

Note: *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 2. Regression analysis of the order book imbalance and returns

This table uses the information revealed by the order book in arrears to predict future returns using the following regression formula:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{5} \beta_j HR_{j,t-1} + \sum_{j=2}^{5} \gamma_j QR_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, QR_j is the market imbalance for the jth quote, and HR_j is the price imbalance for the jth quote.

j	1	2	3	4	5
α	-0.307***	-0.288***	-0.273***	-0.267***	-0.267***
$Spread_{t-1}$	23.962***	23.952***	23.983***	24.011***	24.024***
$HR_{2,t-1}$		0.002	0.003**	0.004**	0.004**
$HR_{3,t-1}$			-0.036***	-0.034***	-0.034***
$HR_{4,t-1}$				-0.031***	-0.031***
$HR_{5,t-1}$					-0.013***
$QR_{1,t-1}$	0.006***	0.005***	0.004***	0.004***	0.004***
$QR_{2,t-1}$		0.010***	0.009***	0.008***	0.006***
$QR_{3,t-1}$			0.006***	0.005***	0.005***
$QR_{4,t-1}$				0.002***	0.002***
$QR_{5,t-1}$					-0.001**
Adj-R^2	8.00%	8.13%	8.20%	8.23%	8.24%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 3. Relationship between the order book height, length, and the return.

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=1}^{5} \gamma_{d,j} Q^d_{j,t-1} + \sum_{j=1}^{5} \gamma_{s,j} Q^s_{j,t-1} + \sum_{j=2}^{5} \beta_{d,j} \Delta p^d_{j,t-1} + \sum_{j=2}^{5} \beta_{s,j} \Delta p^s_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, Q^d_j is the quantity of market demand for the *j*th quote, Q^s_j is the quantity of market supply for the *j*th quote, Δb^d_j is the price of market demand for the *j*th quote, and Δb^s_j is the price of market supply for the *j*th quote.

j	1	2	3	4	5
α	0.364***	0.397***	0.513***	0.680***	0.576***
$Spread_{t-1}$	32.493***	29.952***	29.916***	29.662***	29.321***
$Q^s_{1,t-1}$	-0.058***	-0.035***	-0.034***	-0.034***	-0.034***
$Q^s_{2,t-1}$		-0.013***	-0.016***	-0.015***	-0.015***
$Q^s_{3,t-1}$			-0.007***	-0.010***	-0.010***
$Q^s_{4,t-1}$				-0.004***	-0.005***
$Q^s_{5,t-1}$					-0.001
$Q^d_{1,t-1}$	0.021***	0.017***	0.016***	0.016***	0.016***
$Q^d_{2,t-1}$		0.020***	0.019***	0.019***	0.019***
$Q^d_{3,\ t-1}$			0.010***	0.010***	0.009***
$Q^d_{4,\ t-1}$				0.003***	0.002***
$Q^d_{5,t-1}$					0.004***
Δ .$^{d}_{2,t-1}$		-0.227***	-0.227***	-0.227***	-0.227***
Δ *$^{d}_{3,t-1}$			0.0202***	0.020***	0.020***
Δ .$^{d}_{4,t-1}$				0.016***	0.016***
Δ *$^{d}_{5,t-1}$					0.001
Δ .$^{s}_{2,t-1}$		-0.171***	-0.155***	-0.158***	-0.151***
Δ *$^{s}_{3,t-1}$			-0.136	-0.124	-0.111
Δ 0$^{s}_{4,t-1}$				-0.126	-0.102
Δ 0$^{s}_{5,t-1}$					0.003
Adj-R^2	8.37%	16.04%	16.20%	16.33%	16.34%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 4. Regression analysis of the order book imbalance and return in an upward trend.

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{5} \beta_j HR_{j,t-1} + \sum_{j=2}^{5} \gamma_j QR_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, QR_j is the imbalance between the market supply and demand for the jth quote, and HR_j is the price imbalance for the jth quote.

j	1	2	3	4	5
α	1.1014***	1.236***	1.245***	1.228***	1.210***
$Spread_{t-1}$	-69.300***	-69.245***	-69.326***	-69.431***	-69.539***
$HR_{2,t-1}$		-0.068***	-0.069***	-0.070***	-0.070***
$HR_{3,t-1}$			0.075***	0.072***	0.070***
$HR_{4,t-1}$				0.107***	0.100***
$HR_{5,t-1}$					0.116***
$QR_{1,t-1}$	0.015***	0.012***	0.011***	0.011***	0.011***
$QR_{2,t-1}$		0.013***	0.012***	0.012***	0.013***
$QR_{3,t-1}$			0.007***	0.006***	0.007***
$QR_{4,t-1}$				0.003***	0.003***
$QR_{5,t-1}$					-0.001
Adj-R^2	48.32%	48.51%	48.57%	48.65%	48.75%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 5. Relationship between the order book height, length, and the return in an upward trend.

This table used the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=1}^{5}\gamma_{d,j}Q^{d}_{j,t-1} + \sum_{j=1}^{5}\gamma_{s,j}Q^{s}_{j,t-1} + \sum_{j=2}^{5}\beta_{d,j}\Delta p^{d}_{j,t-1} + \sum_{j=2}^{5}\beta_{s,j}\Delta p^{s}_{j,t-1} + \eta_t$$

where *j* is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, Q^d_j is the quantity of market demand for the *j*th quote, Q^s_j is the quantity of market supply for the *j*th quote, Δb^d_j is the price of market demand for the *j*th quote, and Δs^s_j is the price of market supply for the *j*th quote.

j	1	2	3	4	5
α	-0.108	1.243***	1.859***	1.608***	1.221***
$Spread_{t-1}$	-95.515***	-93.038***	-93.586***	-94.110***	-93.028***
$Q^s_{1,t-1}$	-0.018***	-0.075***	-0.075***	-0.075***	-0.075***
$Q^s_{2,t-1}$		-0.013***	-0.013***	-0.013***	-0.013***
$Q^s_{3,t-1}$			-0.007***	-0.007***	-0.006***
$Q^s_{4,t-1}$				-0.005***	-0.005***
$Q^s_{5,t-1}$					-0.001
$Q^d_{1,t-1}$	0.039***	0.036***	0.035***	0.035***	0.035***
$Q^d_{2,t-1}$		0.026***	0.025***	0.025***	0.025***
$Q^d_{3, t-1}$			0.009***	0.008***	0.008***
$Q^d_{4, t-1}$				0.004***	0.004***
$Q^d_{5,t-1}$					-0.001
$\Delta\ 0^d_{2,t-1}$		0.380***	0.379***	0.379***	0.379***
$\Delta\ *^d_{3,t-1}$			0.003***	0.003**	0.003**
$\Delta\ .^d_{4,t-1}$				0.002	0.001
$\Delta\ .^d_{5,t-1}$					0.001
$\Delta\ .^s_{2,t-1}$		-1.128***	-1.096***	-1.101***	-1.112***
$\Delta\ 1^s_{3,t-1}$			-0.644***	-0.655***	-0.679***
$\Delta\ 0^s_{4,t-1}$				0.267	0.225
$\Delta\ .^s_{5,t-1}$					0.506
Adj-R^2	48.28%	60.65%	60.66%	60.67%	60.67%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 6. Regression analysis of the order book imbalance and return in a downward trend.

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{5} \beta_j HR_{j,t-1} + \sum_{j=2}^{5} \gamma_j QR_{j,t-1} + \eta_t$$

where *j* is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, QR_j is the imbalance between the market supply and demand for the *j*th quote, HR_j is the price imbalance for the *j*th quote.

j	1	2	3	4	5
α	-0.523***	-0.586***	-0.593***	-0.593***	-0.588***
$Spread_{t-1}$	27.207***	28.651***	28.840***	28.833***	28.555***
$HR_{2,t-1}$		-0.038***	-0.037***	-0.037***	-0.038***
$HR_{3,t-1}$			-0.010***	-0.010***	-0.010***
$HR_{4,t-1}$				0.012***	0.011***
$HR_{5,t-1}$					0.021***
$QR_{1,t-1}$	0.016***	0.013***	0.013***	0.013***	0.013***
$QR_{2,t-1}$		0.012***	0.011***	0.011***	0.012***
$QR_{3,t-1}$			0.005***	0.005***	0.005***
$QR_{4,t-1}$				0.001	0.001
$QR_{5,t-1}$					-0.003
Adj-R^2	3.76%	6.65%	6.92%	6.95%	7.12%

Notes: 1 The coefficient is the result multiplied by 100.
2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 7. Relationship between the order book height, length, and the return in a downward trend.

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=1}^{5} \gamma_{d,j} Q^d_{j,t-1} + \sum_{j=1}^{5} \gamma_{s,j} Q^s_{j,t-1} + \sum_{j=2}^{5} \beta_{d,j} \Delta p^d_{j,t-1} + \sum_{j=2}^{5} \beta_{s,j} \Delta p^s_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), s*pread* is the relative price difference for the best quote, Q^d_j is the quantity of market demand for the jth quote, Q^s_j is the quantity of market supply for the jth quote, Δb^d_j is the price of market demand for the jth quote, and Δb^s_j is the price of market supply for the jth quote.

j	1	2	3	4	5
α	-0.191***	1.125***	1.532***	1.179***	0.776***
$Spread_{t-1}$	42.146***	42.058***	42.255***	42.188***	42.395***
$Q^s_{1,t-1}$	-0.040***	-0.037***	-0.036***	-0.036***	-0.036***
$Q^s_{2,t-1}$		-0.020***	-0.019***	-0.019***	-0.019***
$Q^s_{3,t-1}$			-0.009***	-0.009***	-0.009***
$Q^s_{4,t-1}$				-0.004***	-0.004***
$Q^s_{5,t-1}$					-0.001
$Q^d_{1,t-1}$	0.033***	0.030***	0.029***	0.029***	0.029***
$Q^d_{2,t-1}$		0.018***	0.017***	0.016***	0.016***
$Q^d_{3, t-1}$			0.010***	0.009***	0.009***
$Q^d_{4, t-1}$				0.004***	0.004***
$Q^d_{5,t-1}$					0.001**
$\Delta *^d_{2,t-1}$		0.004***	0.003***	0.003***	0.003***
$\Delta .^d_{3,t-1}$			0.001*	0.001*	0.001*
$\Delta .^d_{4,t-1}$				0.001***	0.001***
$\Delta .^d_{5,t-1}$					0.001*
$\Delta .^s_{2,t-1}$		-1.285***	-1.252***	-1.254***	-1.266
$\Delta 1^s_{3,t-1}$			-0.446***	-0.446***	-0.456***
$\Delta 0^s_{4,t-1}$				0.347***	0.330***
$\Delta *^s_{5,t-1}$					0.431***
Adj-R^2	2.74%	4.72%	5.17%	5.28%	5.30%

Notes: 1 The coefficient is the result multiplied by 100.
2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 8. Relationship between the buying and selling price impact imbalance and the return

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=2}^{5} \beta_{d,j} LD(j\overline{Q}/2)_{t-1} + \sum_{j=2}^{5} \beta_{s,j} LS(j\overline{Q}/2)_{t-1} + \eta_t$$

$\overline{Q}$ is the average quantity per transaction in the current quote, *j* is the revealed quote of the orders, ε_t is the residual of quote midpoint AR(5), *spread* is the relative price difference for the best quote, and $LD(j\overline{Q}/2)$ and $LS(j\overline{Q}/2)$ are the price impacts of the supply and demand, with the condition of a hypothesized transacted quantity of $j\overline{Q}/2$.

j	1	2	3	4	5
α	-0.131^{***}	-0.113^{***}	-0.107^{***}	-0.119^{***}	-0.118^{***}
$Spread_{t-1}$	26.597^{***}	22.085^{***}	22.096^{***}	22.139^{***}	22.190^{***}
$LD(1.0\overline{Q})_{t-1}$	-0.142^{**}	-0.943^{***}	-0.652^{***}	-0.652^{***}	-0.656^{***}
$LD(1.5\overline{Q})_{t-1}$		0.650	-0.269	0.430	0.430
$LD(2.0\overline{Q})_{t-1}$			0.630^{***}	-1.320^{**}	-0.085
$LD(2.5\overline{Q})_{t-1}$				1.251^{***}	-1.985^{***}
$LD(3.0\overline{Q})_{t-1}$					2.008^{***}
$LS(1.0\overline{Q})_{t-1}$	-0.113	0.687	0.298	0.294	0.295
$LS(1.5\overline{Q})_{t-1}$		-0.604^{***}	0.620	0.234	0.235
$LS(2.0\overline{Q})_{t-1}$			-0.837^{***}	0.260	-0.998^{*}
$LS(2.5\overline{Q})_{t-1}$				-0.708^{***}	2.580^{***}
$LS(3.0\overline{Q})_{t-1}$					-2.034^{***}
Adj-R^2	7.93%	7.94%	7.95%	7.95%	7.96%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 9. Regression analysis of the order book imbalance and return of extremely large and small *LR*

This table utilizes the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{5} \beta_j HR_{j,t-1} + \sum_{j=2}^{5} \gamma_j QR_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, QR_j is the imbalance between the market supply and demand for the jth quote, and HR_j is the price imbalance for the jth quote.

j	1	2	3	4	5
α	-0.283***	-0.267***	-0.248***	-0.241***	-0.242***
$Spread_{t-1}$	23.997***	23.986***	23.984***	23.973***	23.965***
$HR_{2,t-1}$		0.003*	0.003*	0.003*	0.003*
$HR_{3,t-1}$			-0.005	-0.004	-0.004
$HR_{4,t-1}$				0.007	0.006
$HR_{5,t-1}$					0.009*
$QR_{1,t-1}$	0.001*	-0.001*	-0.001***	-0.001***	-0.001**
$QR_{2,t-1}$		0.010***	0.009***	0.008***	0.008***
$QR_{3,t-1}$			0.008***	0.007***	0.007***
$QR_{4,t-1}$				0.004***	0.004***
$QR_{5,t-1}$					0.001
Adj-R^2	79.84%	79.97%	80.04%	80.06%	80.06%

Notes: 1 The coefficient is the result multiplied by 100.
2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 10. Relationship between order book height, length, and the return on extremely large and small *LR*

This table uses the information revealed by the order book in arrears to predict future returns. The regression formula is as follows:

$$\varepsilon_t = \alpha_0 + \beta_0 spread_{t-1} + \sum_{j=1}^{5}\gamma_{d,j} Q^d_{j,t-1} + \sum_{j=1}^{5}\gamma_{s,j} Q^s_{j,t-1} + \sum_{j=2}^{5}\beta_{d,j}\Delta p^d_{j,t-1} + \sum_{j=2}^{5}\beta_{s,j}\Delta p^s_{j,t-1} + \eta_t$$

where j is the revealed quote in the order book, ε_t is the residual of the quote midpoint AR(5), *spread* is the relative price difference for the best quote, Q^d_j is the quantity of the market demand for the *j*th quote, Q^s_j is the quantity of the market supply for the *j*th quote, Δb^d_j is the price of the market demand for the *j*th quote, and Δb^s_j is the price of the market supply for the *j*th quote.

j	1	2	3	4	5
α	0.654***	-0.867***	-0.1606***	0.901***	1.853***
$Spread_{t-1}$	49.737***	-11.254***	12.749***	-2.256***	68.997***
$Q^s_{1,t-1}$	-0.018***	-0.016***	-0.015***	-0.016***	-0.016***
$Q^s_{2,t-1}$		-0.016***	-0.015***	-0.015***	-0.015***
$Q^s_{3,t-1}$			-0.010***	-0.008***	-0.008***
$Q^s_{4,t-1}$				-0.007***	-0.007***
$Q^s_{5,t-1}$					-0.002***
$Q^d_{1,t-1}$	0.007***	0.006*	0.005***	0.005***	0.005***
$Q^d_{2,t-1}$		0.019***	0.017***	0.016***	0.016***
$Q^d_{3,t-1}$			0.014***	0.011***	0.011***
$Q^d_{4,t-1}$				0.011***	0.010***
$Q^d_{5,t-1}$					0.006***
$\Delta\ .^d_{2,t-1}$		4.52E+12	1.59E+14	-3.58E+11	5.38E+13
$\Delta\ .^d_{3,t-1}$			7.67E+12	1.16E+14	4.33E+13
$\Delta\ .^d_{4,t-1}$				-2.99E+15	-1.32E+13
$\Delta\ 1^d_{5,t-1}$					5.05E+14
$\Delta\ .^s_{2,t-1}$		-4.52E+12	-1.59E+14	3.58E+11	-5.38E+13
$\Delta\ 5^s_{3,t-1}$			-7.67E+12	-1.16E+14	-4.33E+13
$\Delta\ 4^s_{4,t-1}$				2.99E+15	1.32E+13
$\Delta\ .^s_{5,t-1}$					-5.05E+14
Adj-R^2	79.87%	79.97%	80.04%	80.08%	80.09%

Notes: 1 The coefficient is the result multiplied by 100.

2 *** 1% level of significance, ** 5% level of significance, and * 10% level of significance.

Table 11. Analysis of futures contracts

		t-1	*t-1~t-2*	*t-1~t-3*
TAIEX Futures	Average residual difference	4.53E-06	6.31E-06	5.75E-06
	Average coefficient difference	4.47E-04	2.42E-02	3.75E-02
	Adj-R^2	23%	18%	9%
Electronics Sector Index Futures	Average residual difference	3.94E-06	5.87E-06	4.35E-06
	Average coefficient difference	2.63E-04	5.49E-03	8.88E-03
	Adj-R^2	21%	6%	5%
Finance Sector Index Futures	Average residual difference	2.40E-06	6.24E-06	4.04E-06
	Average coefficient difference	3.54E-04	6.94E-03	1.09E-02
	Adj-R^2	21%	7%	5%
MiNi-TAIEX Futures	Average residual difference	4.06E-06	8.41E-06	5.58E-06
	Average coefficient difference	3.95E-04	1.37E-02	2.54E-02
	Adj-R^2	19%	16%	8%

Note: Average coefficient difference, average residual difference, and Adj-R^2 are all daily averages, with *t-1* as the regression coefficient of the previous day that is integrated into current data, *t-1~t-2* as the regression coefficients of the previous two days that are integrated into current data, and *t-1~t-3* as the regression coefficients of the previous three days that are integrated into current data.

Table 12. Regression analysis of trading strategies

This table uses the information from the previous day to conduct following regression:

$$\eta_t^* = \alpha_0 + \beta_1 QR_{t-1,1} + \beta_2 HR_{t-1,1} + \beta_3 Spread_{t-1} + \beta_4 QR_{t-1,1}^{up,1} + \beta_5 QR_{t-1,1}^{down,1} + \beta_6 HR_{t-1,1}^{up,1}$$
$$+ \beta_7 HR_{t-1,1}^{down,1} + \beta_8 Spread_{t-1}^{up,1} + \beta_9 Spread_{t-1}^{down,1}$$

	TAIEX Futures	Electronics Sector Index Futures	Finance Sector Index Futures	MiNi-TAIEX Futures
Intercept	-0.013	-0.037	-0.043	-0.014
	(231/349)*	(346/349)*	(345/349)*	(296/349)*
QR of the first quote	-0.004	-0.011	-0.012	-0.007
	(200/349)*	(261/349)*	(259/349)*	(236/349)*
HR of the first quote	0.050	0.022	0.027	0.032
	(206/349)*	(313/349)*	(307/349)*	(190/349)*
Spread	0.009	0.233	0.074	0.008
	(280/349)*	(347/349)*	(346/349)*	(300/349)*
QR with an upward trend	-0.001	-0.004	-0.006	-0.001
	(150/349)*	(132/349)*	(150/349)*	(135/349)*
QR with a downward trend	0.001	0.001	0.001	0.001
	(156/349)*	(135/349)*	(123/349)*	(98/349)*
HR with an upward trend	0.007	0.010	0.015	0.001
	(196/349)*	(141/349)*	(162/349)*	(113/349)*
HR with a downward trend	-0.010	-0.001	-0.004	-0.003
	(218/349)*	(118/349)*	(138/349)*	(129/349)*
Spread with an upward trend	-0.001	0.003	0.001	-0.001
	(150/349)*	(100/349)*	(166/349)*	(158/349)*
Spread with a downward trend	-0.001	-0.010	-0.001	0.001
	(109/349)*	(113/349)*	(192//349)*	(140/349)*
Day $_{t-1}$ Adj-R^2	22.04%	20.97%	21.11%	18.71%
Adj-R^2	66.08%	36.36%	96.47%	27.36%

Note: The number of significance is the sum of the days with 10%, 5%, and 1% significance, η_t is the residual item of the mid-quote return after subtracting serial correlation, and Adj-R^2 is the daily average.

Table 13. Returns generated with trading strategies after subtracting transaction and price slippage

Panel A	TAIEX Futures	Electronics Sector Index Futures	Finance Sector Index Futures	MiNi-TAIEX Futures
Average return per contract	0.018%	0.048%	0.061%	0.062%
Average daily return	2.623%	8.754%	10.904%	2.929%
Total average range retained	2.31	3.23	3.32	2.23
Total transacted contracts	49373	61826	60596	16019
Average return per contract using the bullish strategy	-0.005%	0.056%	0.072%	-0.042%
t-value	-34.14***	39.18***	31.16***	-21.18***
Contracts of bullish trades	26527	30910	30287	8840
Total average range retained using the bullish strategy	2.39	2.95	3.05	2.45
Average return per contract using the bearish strategy	0.045%	0.039%	0.049%	0.192%
t-value	5.59***	17.84***	13.79***	9.08***
Contracts of bearish trades	22846	30916	30309	7179
Total average range retained using the bearish strategy	2.21	3.51	3.58	1.99

Panel B	TAIEX Futures	Electronic Sector Index Futures	Finance Sector Index Futures	MiNi-TAIEX Futures
Bullish				
Success rate	57.08%	71.19%	69.98%	53.51%
Average success rate per contract	0.08%	0.13%	0.17%	0.10%
Standard deviation	0.0004	0.0006	0.0010	0.0008
Average loss per contract	-0.11%	-0.06%	-0.07%	-0.18%
Standard deviation	0.0005	0.0007	0.0006	0.0033
Bearish				
Success rate	43.19%	67.11%	65.23%	52.77%
Average success rate per contract	0.21%	0.12%	0.16%	0.38%
Standard deviation	0.0004	0.0013	0.0022	0.0063
Average loss per contract	-0.05%	-0.05%	-0.06%	-0.08%
Standard deviation	0.0067	0.0003	0.0004	0.0008

Table A1. Transaction costs of futures contracts

Products	TAIEX Futures	Electronics Sector Index Futures	Finance Sector Index Futures	MiNi-TAIEX Futures
Fee (buying + selling)	0.0120%	0.0154%	0.0196%	0.0301%
Futures tax (buying + selling)	0.0080%	0.0080%	0.0080%	0.0080%
Total transaction costs	0.0200%	0.0234%	0.0276%	0.0381%