科技部補助專題研究計畫報告

臺灣股票市場高頻方向性價格跳躍之研究(第2年)

報 告 類 別 ： 成果報告
計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： MOST 107-2410-H-003-020-MY2
執 行 期 間 ： 108年08月01日至109年07月31日
執 行 單 位 ： 國立臺灣師範大學管理研究所

計 畫 主 持 人 ： 蔡蒔銓

計畫參與人員： 碩士班研究生-兼任助理：吳家儀
碩士班研究生-兼任助理：趙珮伶
碩士班研究生-兼任助理：張譽耀
碩士班研究生-兼任助理：楊育晟
碩士班研究生-兼任助理：曾子維
碩士班研究生-兼任助理：符丞甫
碩士班研究生-兼任助理：林佳旻
碩士班研究生-兼任助理：康佳琳
碩士班研究生-兼任助理：陳奕涵
碩士班研究生-兼任助理：蔡嘉蓉
碩士班研究生-兼任助理：周心安
碩士班研究生-兼任助理：魏如鈺
碩士班研究生-兼任助理：邱亮瑜
碩士班研究生-兼任助理：謝富承
碩士班研究生-兼任助理：邱彥霖
碩士班研究生-兼任助理：蔡侑翰
碩士班研究生-兼任助理：吳冠融

本研究具有政策應用參考價值：■否 □是，建議提供機關

（勾選「是」者，請列舉建議可提供施政參考之業務主管機關）
本研究具影響公共利益之重大發現：□否　□是

中　華　民　國　109　年　10　月　29　日

中 文 摘 要 ： 本研究分析台灣證券交易所完整的逐筆交易資料，探討影響高頻跳躍的主要因素。過去文獻提出造成股價方向性跳躍的可能因素，包括:資訊不對稱、市場流動性、群聚效應、市場情緒等，本研究發現「市場流動性」是影響高頻方向性跳躍的主要因素。同時，最小絕對值收斂和選擇算法(LASSO)、彈性網路分析、主成分分析的結果顯示，「市場流動性」在瞭解金融資產高頻的突然性非連續價格變動，較其他因素如:「資訊不對稱」或「市場情緒」等，更為重要。

中 文 關 鍵 詞 ： 方向性跳躍、資訊不對稱、市場流動性、群聚效應、機器學習

英 文 摘 要 ： We analyze the complete tick-level stock trading records at the Taiwan Stock Exchange and explore what factors are principally associated with jumps in high frequency. Among the potential candidate variables suggested in the literature, liquidity proxies appear to be primarily associated with signed jumps in high frequency. The results from the least absolute shrinkage and selection operator (LASSO), the elastic net method, and principal component analysis further show that liquidity issues are more important than information or sentiment in understanding sudden and discontinuous price innovations to financial assets in high frequency.

英 文 關 鍵 詞 ： signed jumps, information asymmetry, liquidity, herding, machine learning

What Factors Are Associated with Jumps in High Frequency?

Blind Copy

Abstract

We analyze the complete tick-level stock trading records at the Taiwan Stock Exchange and explore what factors are principally associated with jumps in high frequency. Among the potential candidate variables suggested in the literature, liquidity proxies appear to be primarily associated with signed jumps in high frequency. The results from the least absolute shrinkage and selection operator (LASSO), the elastic net method, and principal component analysis further show that liquidity issues are more important than information or sentiment in understanding sudden and discontinuous price innovations to financial assets in high frequency.

JEL classification: G12, G14.

Keywords: signed jumps, information asymmetry, liquidity, herding, machine learning.

Declarations of interest: none.

1. Introduction

Stochastic volatility (Mandelbrot, 1963), jumps (Press, 1967), and market microstructure effects (Niederhoffer and Osborne, 1966) have been long-standing stylized empirical facts within the finance profession. With these breakthroughs, the asset pricing literature has been significantly firmed up from both the theoretical and empirical sides. In this paper, we quantify a catalog of candidate measures that have been believed to affect asset price fluctuations in high frequency, and document that high-frequency jumps are principally associated with liquidity proxies. To do so, we utilize a remarkably unique dataset from the Taiwan Stock Exchange (TWSE), which contains the complete transaction data, order-level tick data, and the identity of each trader in the Taiwan stock market.[1] The use of high frequency tick data is necessary, given that asset prices fluctuations metrics based on low-frequency data tend to detect jumps spuriously (Christensen, Oomen, and Podolskij, 2014). With the tick data from the TWSE, we furnish a comprehensive accounting of the drivers of high-frequency jumps.

The observation that asset prices appear to jump is a well-established regularity in the asset pricing literature, including Aït-Sahalia, Jacod, and Li (2012) (high-frequency jumps in noisy data), Andersen, Bollerslev, and Diebold (2007a) (jump detection, modeling, and forecasting), Andersen, Bollerslev, Diebold, and Labys (2001) (realized exchange rate volatility), Andersen, Bollerslev, Diebold, and Labys (2003a) (modeling realized volatility), Bajgrowicz, Scaillet, and Treccani (2015) (spurious jump detection), Baruník, Kočenda, and Vácha (2016) (asymmetric volatility spillovers), Evans (2011) (macroeconomic news announcements), Lee and Hannig (2010) (jump detection from Levy jump diffusion processes), Lee (2012) (nonparametric tests for jumps), Miao, Ramchander, and Zumwalt (2014) (jumps in the S&P 500 index futures), Patton and Sheppard (2015) (jumps and volatility predictability), and Tauchen and Zhou (2011) (credit spread prediction). We build on these papers to study what factors are reliably associated with jumps in the high-frequency context.

[1] The TWSE data set is also utilized in Barber, Lee, Liu, and Odean (2007) and Barber, Lee, Liu, and Odean (2009), among others.

The drivers of jumps may hinge on the types of jumps that one considers, e.g., Segal, Shaliastovich, and Yaron (2015). The literature has employed the even functions of high-frequency returns, ignoring the information content that may be conveyed in the sign of asset returns. We thus adopt the realized semivariance approach proposed by Barndorff-Nielsen, Kinnebrock, and Shephard (2008) and Patton and Sheppard (2015). The realized semivariance "decomposes the usual realized variance into a component that relates only to positive high-frequency returns and a component that relates only to negative high-frequency returns" (Patton and Sheppard, 2015, p. 683). As such, negative jumps are those associated negative innovations to asset returns, while positive jumps are those associated with positive shocks.

The literature on the driving forces of high-frequency jumps is lacking, mostly because access to comprehensive data is limited. This study overcomes such a hurdle by utilizing a comprehensive tick-level dataset in Taiwan from 2008 to 2014. The TWSE tick data allow us to compute a variety of candidate metrics that have been believed to affect asset price fluctuations in high frequency from a theoretical standpoint. We consider the price impact measure (e.g., Kyle, 1985; Chan, Menkveld, and Yang, 2008), the adverse selection spread component (e.g., Glosten and Harris, 1988), the probability of informed trading (e.g., Easley, Kiefer, and O'Hara, 1996), effective and realized spreads (e.g., Goyenko, Holden, and Trzcinka, 2009; Hasbrouck, 2009), buy-sell imbalances (e.g., Kumar and Lee, 2006), and herding (e.g., Patterson and Sharma, 2007) as candidate proxies. We first examine how signed jumps respond to each of these explanatory variables suggested in theory. We may not rule out the possibility that the aforementioned factors contain some complex correlation structure that can be eliminated without losing valuable information. Hence, in order to identify the key driving forces of jumps, we employ two machine-learning-based feature selection methods: a) the least absolute shrinkage and selection operator (Tibshirani, 1996) and b) the elastic net (Zou and Hastie, 2005). Separately, we employ principal component analysis to check the factor loading of the first few principal components on each of the aforementioned variables.

The evidence we amass suggests that jumps are principally associated with liquidity proxies in high frequency. Most information asymmetry proxies do not square with asset price fluctuations in high frequency. We also provide evidence that the results are not an artifact of data fitting. The results from two machine-learning-based feature selection techniques show that liquidity proxies are indeed primarily correlated with high-frequency jumps. The results from principal component analysis are also consistent, in that the first and second principal components are highly loaded on liquidity proxies.

We contribute to the current understanding of the key drivers of high-frequency jumps by analyzing all tick-level data on the TWSE in the period 2008-2014. This extensive data set and long timespan enable us to furnish compelling evidence that both positive and negative jumps are highly associated with liquidity proxies. Theis price-liquidity relation will have significant implications for regulatory bodies and market participants in the stock market.

The rest of the paper is organized as follows. In Section 2, we summarize how to estimate signed jumps from high-frequency data. In Section 3, we examine the potential determinants of high-frequency jumps. In Section 4, we discuss the institutional background of the TWSE data. In Section 5, we provide the empirical test that we conduct. We recap the merits of this study and conclude in Section 6.

2. Signed Jump Estimation from High-Frequency Data

We briefly describe how to detect signed jumps that are used in our empirical analysis. Let μ be a locally bounded predictable drift process, and let σ be a strictly positive cádlág process. J denotes a pure jump process. We consider a continuous martingale for the log price, p_t, in the following form:

$$p_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_t + J_t \tag{1}$$

in which W represents a standard Wiener process.

For continuous time $t \geq 0$ and for any partition $0 = t_t < ... < t_n = t$ with $\sup_j[t_{j+1} - t_j] \to 0$

as $n \to 0$, the quadratic variation process of a semi-martingale λ_t is:

$$[\lambda_t, \lambda_t] = p\text{-}\lim_{n\to\infty} \sum_{j=0}^{n-1} (\lambda_{t_{j+1}} - \lambda_{t_j})^2. \tag{2}$$

As such, the quadratic variation of the process in Eq.(1) is:

$$[p_t, p_t] = \int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta p_s)^2 \tag{3}$$

where $\Delta p_s = p_s - p_{s-}$ represents a jump.

Andersen et al. (2001) introduce the concept of realized variance and show that the realized variance for discretely observed log stock prices converges to the quadratic variation in probability:

$$\text{RV} = \sum_{i=1}^{n} r_i^2 \xrightarrow{p} [p_t, p_t], \quad as\ n \to \infty$$

in which r_i stands for the log return, i.e., $r_i = p_i - p_{i-1}$.

Barndorff-Nielsen and Shephard (2006) further extend the study of realized variance to bipower variation (BP). For some interval δ of time t, we define the $\delta-$returns on a discretized semi-martingale p_t on intervals with length δ as:

$$r_i^\delta = p_{i\delta} - p_{(i-1)\delta}, \quad i = 1, ..., \lfloor t/\delta \rfloor \tag{4}$$

in which $\lfloor \cdot \rfloor$ is the integer part operator. The realized quadratic variation of the discretized log price process can then be defined as:

$$[p_\delta, p_\delta]_t = \sum_{i=1}^{\lfloor t/\delta \rfloor} (r_i^\delta)^2 \tag{5}$$

The 1,1-order bipower variation, if it exists, is:

$$\{p_t, p_t\} = p\text{-}\lim_{\delta \downarrow 0} \sum_{i=2}^{\lfloor t/\delta \rfloor} |r_{i-1}^{\delta}||r_i^{\delta}| \tag{6}$$

Similar to the realized quadratic variation process of the discretized log price process, the realized bipower variation process can be defined as:

$$\{p_\delta, p_\delta\}_t^{[1,1]} = \sum_{i=2}^{\lfloor t/\delta \rfloor} |r_{i-1}^{\delta}||r_i^{\delta}| \tag{7}$$

It can be shown that the 1,1-order bipower variation converges in probability to the integrated variance:

$$\mathrm{BP} = \frac{\pi^2}{4} \sum_{i=2}^{n} |r_i^{\delta}||r_{i-1}^{\delta}| \xrightarrow{p} \int_0^t \sigma_s^2 ds, \quad as\ n \to \infty \tag{8}$$

Eq.(3), Eq.(4), and Eq.(8) imply that the difference between the bipower variation and the realized variation converges in probability to the jump variation:

$$\text{RV-BP} \xrightarrow{p} \sum_{0<s\leq t} (\Delta p_s)^2. \tag{9}$$

Barndorff-Nielsen et al. (2008) propose an estimator that can discompose the jump component in Eq.(9) into asymmetrical variations:

$$\mathrm{RS}^+ = \sum_{i=1}^{n} r_i^2 \mathbb{I}\{r_i > 0\} \tag{10}$$

$$\mathrm{RS}^- = \sum_{i=1}^{n} r_i^2 \mathbb{I}\{r_i < 0\} \tag{11}$$

where $\mathbb{I}$ is the indicator function. The realized variance can then be easily decomposed as:

$$\mathrm{RV} = \mathrm{RS}^+ + \mathrm{RS}^- \tag{12}$$

Barndorff-Nielsen et al. (2008) further show that the realized semivariance in Eq.(10) and Eq.(11) converges in probability to a quantity consisting of the integrated variance and the sum of squared jumps:

$$\mathrm{RS}^{+} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s > 0\} \tag{13}$$

$$\mathrm{RS}^{-} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s < 0\}. \tag{14}$$

The indicator function in Eq.(13) and Eq.(14) captures the signed jumps that we will utilize. We finally define the signed jump as follows:

$$\Delta J^2 \equiv \mathrm{RS}^{+} - \mathrm{RS}^{-} \tag{15}$$

$$\xrightarrow{p} \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s > 0\} - \sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s < 0\}. \tag{16}$$

The first part in Eq.(16), i.e., $\sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s > 0\}$, denotes the variation only due to positive returns. Similarly, the second part, i.e., $-\sum_{0\le s\le t} \Delta p_s^2 \mathbb{I}\{\Delta p_s < 0\}$, represents the variation only due to negative returns. In our econometric analysis, we use the two signed jump estimators to explore the cause and effect of signed high-frequency returns to gain new insights into the market micro-structure literature.

3. Defining Measures

There are a variety of measures that affect high-frequency stock returns. Because we need to have access to time-stamped data, we consider a group of variables that are available in high frequency for the entire data span: 1) information asymmetry, 2) liquidity, and 3) high-frequency sentiment. As the focus of this study is on the high-frequency implications of signed jumps,

[2]Moreover, the macroeconomic data are day-stamped, making it impossible to keep track of the exact time at the second level. For this line of research, refer to Andersen, Bollerslev, Diebold, and Vega (2003b), Andersen, Bollerslev, Diebold, and Vega (2007b), Faust, Rogers, Wang, and Wright (2007), and Evans (2011).

we do not consider Taiwan's macroeconomic announcements such as gross domestic product, real activity, consumption, investment, government purchases, trade balance, prices, or other forward-looking indicators.[2] Our empirical design is similar to Lee (2012), who keeps track of earnings announcements, analyst recommendations, and dividend announcement at the minute level, and studies the association between the three events and stock return jumps.

3.1 Information Asymmetry

We quantify the following three information asymmetry measures: a) the price impact measure, b) the adverse selection spread component, and c) the probability of informed trading.

3.1.1 Price Impact Measure

The price impact measure bases its foundation on the notion that the price impact of a transaction is most capable of capturing asymmetric information. We skip the finer motivation of the model, which can be found in Kyle (1985), Easley and O'Hara (1987), Glosten (1987), and Glosten and Harris (1988). Chan et al. (2008) also exploit the price impact measure to explain the effects of asymmetric information in China's foreign B-share equity market.

We follow Kyle (1985), Glosten and Harris (1988), and Chan et al. (2008) to estimate the price impact measure. Let m_t denote the true price of the stock at time t. The true price process evolves as:

$$m_t = m_{t-1} + \zeta Q_t V_t + e_t \tag{17}$$

in which ζ is the price impact parameter, Q_t is an indicator for the trade sign (+1 for buys and -1 for sells), V_t is the size of the trade, and e_t is the unobserved public information innovation in the true price process. The transaction price, P_t, can be defined as:

$$P_t = m_t + \theta Q_t. \tag{18}$$

Plugging Eq.(17) into Eq.(18) and replacing m_{t-1} with $P_{t-1} - \theta Q_{t-1}$ yield

$$\Delta P_t = \zeta Q_t V_t + \theta(Q_t - Q_{t-1}) + e_t. \tag{19}$$

It should be recognized that transactions are not always executed at the bid or ask prices. We thus employ Harris' (1989) convention as well as the tick test of Lee and Ready (1991), to classify individual intraday trades as buyer- or seller-initiated transactions. These transactions executed above the prevailing midpoint are classified as buys, and vice versa. The trades exactly at the midpoint of the bid and ask prices are signed based on the tick test, i.e., a buy if 1) the price change is positive or 2) the last price change was positive and the current price change is zero, and vice versa.

In the empirical analysis, we compute the price impact measure (ζ) over a 30-minute interval. We reserve our discussion to the empirical analysis section and proceed to define the other two measures of information asymmetry.

3.1.2 Adverse-Selection Spread Component

Secondly, following Glosten and Harris (1988), we model the adverse selection spread component in a linear specification. In particular, we assume that the adverse selection component (ζ_t) of the bid-ask spread and the transitory spread component (θ_t) evolve over time as follows:

$$\zeta_t = \zeta_0 + \zeta_1 V_t \tag{20}$$

$$\theta_t = \theta_0 + \theta_1 V_t \tag{21}$$

where V_t is the number of shares traded in transaction t and ζ_0, ζ_1, θ_0, and θ_1 are constants. The true price process can be written as $m_t = m_{t-1} + \zeta_t Q_t + e_t$. Similarly, the observed transaction price can be restated as $P_t = m_t + \theta_t Q_t$.

As in Glosten and Harris (1988) and Chan et al. (2008), we incorporate the roundoff error (η_t) caused by price discreteness into our specification. The observed price change can then be written as:

$$\Delta P_t = \theta_0(Q_t - Q_{t-1}) + \theta_1(Q_t V_t - Q_{t-1}V_{t-1}) + \zeta_0 Q_t + \zeta_1 Q_t V_t + e_t. \tag{22}$$

We then define the adverse selection spread component as $\zeta_0 + \zeta_1 V^*$, where V^* is the median order size over the chosen frequency.

3.1.3 Probability of Informed Trading

Thirdly, we also employ the probability of informed trading (PIN) metric proposed by Easley et al. (1996). The PIN model is based on the sequential trade model, e.g., Glosten and Milgrom (1985) and Easley and O'Hara (1987), among others. The intuition behind the PIN model is that the PIN measure aims to capture the expected informed order flow as a percentage of the expected total order flow. This method has been proven to work well in capturing the degree of private information in the market, so that we follow Easley, Kiefer, and O'Hara (1997a), Easley, Kiefer, and O'Hara (1997b), and Chan et al. (2008) to construct the PIN measure in high frequency.

In the Easley et al. (1996) model, informed traders buy when they have positive private information and sell when they have negative private information. At the beginning of each day, nature decides whether or not a private information will arrive with the probability denoted by a. If some private information arrives on that day, informed traders will obtain a positive private signal with probability d. When positive private information arrives in the market, both informed and noise traders buy. Informed traders arrive at the speed of u, whereas noise traders arrive at rate ϵ_b. As such, the buy order flow follows a Poisson distribution with intensity $u + \epsilon_b$. On the contrary, when there is no positive private information, only noise traders sell. The sell order flow hence follows a Poisson distribution with intensity ϵ_s. The probability of an informed trade is calculated by:

$$\text{PIN} = \frac{a \times u}{a \times u + \epsilon_s + \epsilon_b} \tag{23}$$

which is nothing but the ratio of the informed order flow to the total order flow.

In our analysis, we utilize an extended version of the PIN measure, following Duarte and Young (2009). Since buy order flows tend to be more volatile than sell order flows, one can assume that the rate at which informed traders arrive, u, consists of two parts: u_b, the arrival rate of informed buyers, and u_s, the arrival rate of informed sellers. Duarte and Young (2009)'s extended model allows for a symmetric order flow shock whose occurrence probability is denoted by θ. The conditional probability of the symmetric order flow shock upon the arrival of private information is $\theta^{'}$. In particular, more variation is allowed in buys (Δ_b) and sells (Δ_s). Further, the correlation between buys and sells is assumed to be positive. In line with the original PIN model, the probability of informed trade is defined by the ratio of the informed order flow to the total order flow:

$$\text{AdjPIN} = \frac{a \times (d \times u_b + (1-d) \times u_s)}{a \times (d \times u_b + (1-d) \times u_s) + (\Delta_b + \Delta_s) \times (a \times \theta' + (1-a) \times \theta) + \epsilon_s + \epsilon_b}. \quad (24)$$

3.2 Liquidity

3.2.1 Spread Benchmarks

We follow Goyenko et al. (2009) to define two spread benchmarks in high frequency. The effective spread for a stock is defined as:

$$\text{Effective Spread}_i = 2|\log(P_i) - \log(M_i)| \quad (25)$$

in which P_i is the transaction price of the ith transaction, and M_i is the midpoint of the best available bid and offer prices. The stock's effective spread for some time interval k is the value-weighted average of all effective spreads during the time interval.

The realized spread for a stock is defined as:

$$\text{Realized Spread}_i = \begin{cases} 2\big[\log(P_i) - \log(P_{i+30})\big] & \text{if the } i\text{-th transaction is a buy} \\ 2\big[\log(P_{i+30}) - \log(P_i)\big] & \text{if the } i\text{-th transaction is a sell} \end{cases} \quad (26)$$

in which P_{i+30} is the transaction price 30 minutes after the ith transaction. The stock's realized spread for some time interval k is the value-weighted average of all realized spreads during the time interval.

3.2.2 Buy-Sell Imbalances

We employ Kumar and Lee (2006)'s buy-sell imbalance (BSI) over a short time period t and form a measure of order-based liquidity in high frequency. In particular, we define a stock's BSI over a particular time interval k as

$$BSI_k = \frac{\sum_i \left[VB_i - VS_i\right]}{\sum_i \left[VB_i + VS_i\right]} \tag{27}$$

where i denotes a finer time interval and VB_i (VS_i) denotes the value-denominated buy (sell) volume at time interval i.

3.2.3 Price Impact Benchmarks

Even though continuous trading is the canonical trade-matching scheme in major capital markets, Taiwan was the only stock market with call auctions throughout the entire regular trading session during our data span. Due to this call auction trade-matching system in Taiwan, we are not able to compute the N-minute price impact, a widely used price-based liquidity benchmark metric, which can be computed as:

$$\text{N-Minute Price Impact}_i = \begin{cases} 2\left[\log(M_{i+N}) - \log(M_i)\right] & \text{if the } i\text{-th transaction is a buy} \\ 2\left[\log(M_i) - \log(M_{i+N})\right] & \text{if the } i\text{-th transaction is a sell} \end{cases} \tag{28}$$

where M_{i+N} is the midprice of the bid and ask prices N-minutes after the ith trade, and M_i is the mid-price when the ith transaction is executed.

Hasbrouck (2009) proposes a Gibbs approach to quantify the price impact, whose estimate is the slope coefficient of the following regression:

$$r_k = \lambda_k S_k + \epsilon_k \tag{29}$$

where r_n is the stock return for the kth interval, and S_k is the signed volume. One can follow Goyenko et al. (2009) and calculate S_k as

$$S_k = \sum_i \text{sign}(\nu_i)\sqrt{|\nu_i|} \tag{30}$$

in which ν_i is the signed trading volume of the ith transaction during the kth period. ϵ_k is the white noise.

3.3 Sentiment

Investor sentiment affects stock returns, e.g., Baker and Wurgler (2006) and Baker and Wurgler (2007) and can be viewed as an indicator of asset price fluctuations (Baker and Stein, 2004). Investors' sentiment can be measured in several ways. We do not consider market sentiment or its disagreement. Market sentiment or its disagreement is usually defined in low frequency. Thus, applying this low-frequency metric to high-frequency data will be problematic because Yu (2011) shows that the half-life of a stock's sentiment disagreement is approximately one year, implying that low-frequency sentiment measures do not vary significantly in high frequency.

In this study, we quantify a herding measure. In the case that investors herd (e.g., Chang, Cheng, and Khorana, 2000; Hwang and Salmon, 2004), we would observe a series of buy or sell orders in the data. To explore herding in our sample, we take advantage of the bootstrap run test proposed by Patterson and Sharma (2007).[3] We first categorize each trade into buyer- or seller-initiated trade using the tick test.

[3] A run of a sequence is a maximal non-empty equal occurrence of the sequence.

Following Lin, Tsai, and Lung (2013), we define a random variable $x(w, p, j)$ as follows:[4]

$$x(w, p, j) = \frac{(R_w + 0.5) - N \times Pr(w) \times (1 - Pr(w))}{\sqrt{N}} \tag{31}$$

in which R_w is the number of runs, N denotes the number of orders for stock j in the interval p, 0.5 is a constant to adjust discontinuity, as in Mood (1940), and $Pr(w)$ is the probability of buy or sell runs. It can be shown that $x(w, p, j)$ asymptotically follows a normal distribution:

$$x(w, p, j) \sim \mathcal{N}(\mu, \sigma_x^2) \tag{32}$$

$$\sigma_x^2 = \mathrm{Pr(w)}(1 - \mathrm{Pr(w)}) - 3\mathrm{Pr(w)}^2(1 - \mathrm{Pr(w)})^2. \tag{33}$$

We standardize the above herding measure and use it in the remaining analysis.

$$\frac{x(w, p, j)}{\sqrt{\mathrm{Pr(w)}(1 - \mathrm{Pr(w)}) - 3\mathrm{Pr(w)}^2(1 - \mathrm{Pr(w)})^2}}. \tag{34}$$

4. Data

4.1 Institutional Background

The Taiwan Stock Exchange (TWSE), established in 1961, achieved a market capitalization value of TWD 21.53 trillion (≊ USD 656 billion) by the beginning of 2008, rising to TWD 26.89 trillion (≊ USD 887 billion) at the end of 2014. The number of listed companies on the TWSE was 698 at the beginning of 2008 and 854 at the end of 2014. The average daily trading volume was approximately 3.13 billion shares or TWD 104.88 billion (≊ USD 3.19 billion) in 2008, and approximately 2.29 billion shares or TWD 88.3 billion (≊ USD 2.91 billion) in 2014, making the TWSE one of the most important exchanges in the Asia-Pacific region.

The TWSE has no market makers or specialists, and operates in a limit-order book environment, which accepts only limit orders. During the regular trading session, from 9:00 AM to 1:30

[4]Buy orders are denoted by $w = 1$, and sell orders are represented by $w = 2$.

PM, buy and sell orders interact according to strict price and time priorities in the central automated trading system to determine a single market-clearing price. In a call-auction market such as the TWSE, each trade (match) involves multiple buy and sell orders. Buy and sell orders accumulate over a period prior to the market clearing, at which point everyone pays or receives the same price, regardless of their quotes. Although market orders are not permitted, traders can submit an aggressive limit order (a marketable limit order) with a price limit equal to or better than the market price to obtain a matching priority.

To avoid excessive volatility and protect investors by limiting their potential daily losses, the TWSE imposes a daily price change limit of 7%, based on the previous day's closing price in both directions during the sample period.[5] The TWSE price limits are boundaries, and not triggers for trading halts, as stocks that hit price limits can still be traded, as long as the transaction prices are within the limits.

4.2 Data

We have the complete transaction history of all traders on the TWSE from 2008 through 2014. Our sample comprises the underlying common domestic stocks, which are constituents of the FTSE TWSE Taiwan 50 Index. This index selects the 50 largest market capitalization listed stocks as constituents, representing the performance of nearly 70% of the Taiwanese stock market. For an individual stock i to enter our sample, we require the stock to remain a constituent of the index throughout our sample period from January 2008 to December 2014 in order to control for the entry and exit effects. Table 1 lists up to 33 companies that satisfy this criterion.

[Insert Table 1 Here]

The trade data include the date and time of the transaction/order, stock identifier, order type (buy or sell), transaction price/order price limit, number of shares, and identity of the trader. We categorize traders into three groups, namely individuals, domestic institutions, and foreign in-

[5]The daily price change limit has become 10%, effective on June 1, 2015.

stitutions. Individual investors not only account for the majority of investors, but also for more than 60 percent of all trading value. Domestic institutions include Taiwanese corporations, financial institutions, mutual funds and government-owned firms. Foreign institutions are primarily foreign banks, insurance companies, securities firms, and mutual funds.

In addition to transaction data, we have the complete limit-order data, which include the date and time of orders, stock identifier, order type (buy or sell), order price limit, number of shares, identity of the trader, and code facilitating the link between transaction and limit orders. The code for changes in orders and changes in the number of shares is also included within the order data.

We consider 30 minutes as the time interval that will be used for the empirical analysis in the following sections. We reiterate that the TWSE operates from 9:00 AM to 1:30 PM each day, i.e., 4.5 hours. It turns out that frequencies of less than 30 minutes make it impossible to stably compute the metrics discussed in Section 3. Therefore, we choose to quantify the quantities over the 30-minute interval. Table 2 recaps the summary statistics for the variables that we consider.

[Insert Table 2 Here]

5. Empirical Design and Results

5.1 Correlations

In order to have a quick sense of the association structure among variables, we report the Pearson product-moment correlations in Table 3. It appears that both up and down jumps are significantly correlated with the candidate measures, except for the adverse selection spread component. We also observe that there exists a complex correlation structure among candidate variables, making it easy to find out alternative explanations. Thus, in the remaining analysis, we first utilize the canonical regression approach to further check the details of the association, and then employ two feature selection techniques to figure out the relative importance of the candidate variables. After identifying a set of key variables, we will take advantage of principal component analysis to demystify the factor involving most of the information among the candidate metrics.

[Insert Table 3 Here]

5.2 Regression Analysis

Our first objective is to explore how the variables identified in Section 3 are associated with observed signed jumps. For each sample stock i at period t of day j with the signed jump $\pm\Delta J^2$, we estimate the following regression:

$$\begin{aligned}
\pm\Delta {J^2}_{i,j,t} \quad = \quad & \alpha + \underbrace{\beta_1 \Delta \text{PIM}_{i,j,t} + \beta_2 \Delta \text{ASS}_{i,j,t} + \beta_3 \Delta \text{AdjPIN}_{i,j,t}}_{\text{Information Asymmetry}} + \ldots \\
& \underbrace{\beta_4 \Delta \text{ES}_{i,j,t} + \beta_5 \Delta \text{RS}_{i,j,t} + \beta_6 \Delta \text{BSI}_{i,j,t}}_{\text{Liquidity}} + \ldots \\
& \underbrace{\beta_7 \Delta \text{Herding}_{i,j,t}}_{\text{Sentiment}} + \gamma \cdot \text{Controls}_{i,j,t} + \epsilon^i_{j,t} \qquad (35)
\end{aligned}$$

in which PIM is the price impact measure, ASS denotes the adverse selection spread component, AdjPIN is the adjusted PIN, ES stands for effective spreads, RS represents realized spreads, BSI is the buy-sell imbalance, and Herding is the herding metric. Controls include each stock's return, volatility, and trading volume. We also control for the return, volatility, and trading volume for the Taiwan Capitalization Weighted Stock Index, which covers all of the listed common stocks at the TWSE. We take the logarithm of the monetary values. As the information set of investors is evolving over time, and traders are responding to the innovations in the information set (e.g., Collin-Dufresne, Goldstein, and Martin, 2001), we make use of the first difference of explanatory variables, denoted by Δ in Eq.(35).[6]

[Insert Table 4 Here]

[6] We do not consider the data from 9am to 9:30am because the first 30-minute bin may contain overnight events.

Table 4 reports the results for up jumps. Column (1) shows that, among information-related measures, both the adverse selection spread components and Adjusted PIN are subsumed by the price impact measure. As for the liquidity-related variables in Column (2), it appears that both realized spreads and buy-sell imbalances are significantly associated with up jumps. The herding metric is not related to up jumps in Column (3). Finally, Column (4) shows that the price impact measure, realized spreads, and buy-sell imbalances turn out to be significantly associated with up jumps.

[Insert Table 5 Here]

The results for down jumps are summarized in Table 5. In Column (1), the price impact measures have statistical power in estimating the magnitude of down jumps (as well as up jumps). We also observe that the realized spreads are highly significantly associated with down jumps. Unlike the case of up jumps, the herding metric turns out to be statistically significant. Column (4) shows that down jumps are significantly associated with price impact measures, realized spreads, and herding.

The results in Tables 4 and 5 imply that price impact measures, realized spreads, buy-sell imbalances, and possibly herding behaviors are likely to be associated with signed jumps in high frequency. With this set of results, we proceed to check whether the four variables are indeed the keys to understanding up and down jumps in high frequency.

5.3 Feature Selection Techniques

Our second approach takes advantage of important recent advances in the statistical learning literature. Similarly to Rapach, Strauss, and Zhou (2013), who draw on feature selection techniques in order to examine the relative predictability of the United States in the global financial market, we employ Tibshirani (1996)'s least absolute shrinkage and selection operator (LASSO), which performs variable selection and regularization with an ℓ_1 penalty term. The LASSO coefficients minimize the below Lagrangian equation:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j|. \tag{36}$$

When the tuning parameter λ is large enough, the ℓ_1 penalty term shrinks some less important or negligible coefficient estimates toward zero. Therefore, the LASSO regression is able to identify a key subset of the variables.

We also employ the elastic net method (Zou and Hastie, 2005), which can complement the LASSO regression. One drawback of the LASSO method is that it does not take into account the correlation structure of the explanatory variable set. In the case that a group of explanatory variables are highly intertwined, the LASSO method selects only one variable in the group. The elastic net method can overcome this caveat of the LASSO method when a group of variables are highly correlated. If there are high correlations among the key explanatory variables, the elastic net method can select more than one variable.[7] The elastic net method assumes two penalty terms and can be written in the following Lagrangian form:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}(\beta_j)^2. \tag{37}$$

[Insert Table 6 Here]

Table 6 shows the results. Consistent with the results from the regressions, the LASSO method finds that up jumps are primarily associated with price impact measures, realized spreads, and buy-sell imbalances, and that down jumps are reliably correlated with herding (to some extent), in addition to the three variables that are chosen for up jumps. The elastic net algorithm reveals a very similar pattern, which ascertains the key variable set selected by the LASSO method.

[7] The elastic net may end up with the same results as those of the LASSO method, which will further firm up the results.

5.4 Principal Component Analysis

The regression in Eq.(35) examines whether the variables suggested by the literature are significant, both statistically and economically, in explaining signed jumps. The LASSO and elastic net methods further clarify the driving forces of the association. In order to check whether there is a low-dimensional representation of the results that capture as much of the information as possible, we conduct principal component analysis. As is usual in the literature (e.g., Manso, Balsmeier, and Fleming, 2019), we take a threshold of eigenvalues above one. It appears that only two principal components satisfy this criterion. Table 7 shows that 58% of the joint variation involving the four key variables identified from the regression, LASSO, and elastic net can be explained by the first two principal components.

[Insert Table 7 Here]

In order to enhance each factor's interpretability, we apply a Varimax rotation of the two principal components.[8] Table 8 shows how much and in which direction each key variable loads on the two principal components. In Table 8, values are multiplied by 100 and rounded off to the nearest integer. Values greater than 0.4 are flagged by an asterisk. We confirm that liquidity proxies are heavily loaded on the first principal component, and both the price impact and herding variables are loaded on the second principal component. It is now safe to claim that jumps are mainly associated with liquidity issues in the stock market.

[Insert Table 8 Here]

6. Concluding Remarks

This study applies recent signed jump detection technique to examine the association between such jumps and the variables that have been believed to induce asset price fluctuations

[8]The results are consistent across other orthogonal rotations.

in high frequency. We further investigate the extent to which signed jumps are associated with proxies for the potential explanatory variables suggested by the literature. It turns out that liquidity proxies are highly associated with both up and down jumps in high frequency in Taiwan's stock market. As high-frequency data tend to be noisy, we further conduct robustness checks using machine-learning-based feature selection techniques and principal component analysis. The three econometric tools that we employ all lend support to the conclusion that liquidity issues play a key role in understanding sudden and discontinuous price innovations to financial assets in high frequency.

We conclude this study by articulating a few interesting future research avenues. Taiwan's stock market is characterized by its unique call-auction-based trade matching system. Unlike most other major financial markets, the TWSE matches trades by call auctions throughout the entire trading hours. A new continuous trade matching scheme was enacted as of March 23, 2020. How this new continuous trade matching system will change the aspects of asset price fluctuations (in high frequency) would be an interesting and important research question. Individual investors account for approximately 60∼70% of the total trading volume at the TWSE. Nonetheless, they lose as much as 2.2% of Taiwan's gross domestic product (Barber et al., 2009). Also, retail investors in Taiwan tend to trade stocks just for fun (Gao and Lin, 2015), which might create excessive asset price fluctuations in the stock market. Hence, demystifying the association between trader identity and asset price fluctuations would be another interesting research avenue.

References

Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. Journal of Econometrics 168, 207–222.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007a. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. The Review of Economics and Statistics 89, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003a. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Vega, C., 2003b. Micro effects of macro announcements: Real-time price discovery in foreign exchange. The American Economic Review 93, 38–62.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Vega, C., 2007b. Real-time price discovery in global stock, bond and foreign exchange markets. Journal of International Economics 73, 251–277.

Bajgrowicz, P., Scaillet, O., Treccani, A., 2015. Jumps in high-frequency data: Spurious detections, dynamics, and news. Management Science 62, 2198–2217.

Baker, M., Stein, J. C., 2004. Market liquidity as a sentiment indicator. Journal of Financial Markets 7, 271–299.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. The Journal of Economic Perspectives 21, 129–151.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., Odean, T., 2007. Is the aggregate investor reluctant to realise losses? evidence from taiwan. European Financial Management 13, 423–447.

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

Barndorff-Nielsen, O. E., Kinnebrock, S., Shephard, N., 2008. Measuring downside risk—realised semivariance .

Barndorff-Nielsen, O. E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. Journal of Financial Econometrics 4, 1–30.

Baruník, J., Kočenda, E., Vácha, L., 2016. Asymmetric connectedness on the us stock market: Bad and good volatility spillovers. Journal of Financial Markets 27, 55–78.

Chan, K., Menkveld, A. J., Yang, Z., 2008. Information asymmetry and asset prices: Evidence from the china foreign share discount. The Journal of Finance 63, 159–196.

Chang, E. C., Cheng, J. W., Khorana, A., 2000. An examination of herd behavior in equity markets: An international perspective. Journal of Banking & Finance 24, 1651–1679.

Christensen, K., Oomen, R. C., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. Journal of Financial Economics 114, 576–599.

Collin-Dufresne, P., Goldstein, R. S., Martin, J. S., 2001. The determinants of credit spread changes. The Journal of Finance 56, 2177–2207.

Duarte, J., Young, L., 2009. Why is pin priced? Journal of Financial Economics 91, 119–138.

Easley, D., Kiefer, N. M., O'Hara, M., 1996. Cream-skimming or profit-sharing? the curious role of purchased order flow. The Journal of Finance 51, 811–833.

Easley, D., Kiefer, N. M., O'Hara, M., 1997a. The information content of the trading process. Journal of Empirical Finance 4, 159–186.

Easley, D., Kiefer, N. M., O'Hara, M., 1997b. One day in the life of a very common stock. Review of Financial Studies 10, 805–835.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90.

Evans, K. P., 2011. Intraday jumps and us macroeconomic news announcements. Journal of Banking & Finance 35, 2511–2527.

Faust, J., Rogers, J. H., Wang, S.-Y. B., Wright, J. H., 2007. The high-frequency response of exchange rates and interest rates to macroeconomic announcements. Journal of Monetary Economics 54, 1051–1068.

Gao, X., Lin, T.-C., 2015. Do individual investors treat trading as a fun and exciting gambling activity? evidence from repeated natural experiments. The Review of Financial Studies 28, 2128–2166.

Glosten, L. R., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. The Journal of Finance 42, 1293–1307.

Glosten, L. R., Harris, L. E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

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, 71–100.

Goyenko, R. Y., Holden, C. W., Trzcinka, C. A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Harris, L., 1989. A day-end transaction price anomaly. Journal of Financial and Quantitative Analysis 24, 29–45.

Hasbrouck, J., 2009. Trading costs and returns for us equities: Estimating effective costs from daily data. The Journal of Finance 64, 1445–1477.

Hwang, S., Salmon, M., 2004. Market stress and herding. Journal of Empirical Finance 11, 585–616.

Kumar, A., Lee, C., 2006. Retail investor sentiment and return comovements. The Journal of Finance 61, 2451–2486.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society pp. 1315–1335.

Lee, C., Ready, M. J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46, 733–746.

Lee, S. S., 2012. Jumps and information flow in financial markets. Review of Financial Studies 25, 439–479.

Lee, S. S., Hannig, J., 2010. Detecting jumps from lévy jump diffusion processes. Journal of Financial Economics 96, 271–290.

Lin, W. T., Tsai, S.-C., Lung, P.-Y., 2013. Investors' herd behavior: Rational or irrational? Asia-Pacific Journal of Financial Studies 42, 755–776.

Mandelbrot, B., 1963. The variation of certain speculative prices. The Journal of Business 36, 394–419.

Manso, G., Balsmeier, B., Fleming, L., 2019. Heterogeneous innovation and the antifragile economy .

Miao, H., Ramchander, S., Zumwalt, J. K., 2014. S&p 500 index-futures price jumps and macroeconomic news. Journal of Futures Markets 34, 980–1001.

Mood, A. M., 1940. The distribution theory of runs. The Annals of Mathematical Statistics 11, 367–392.

Niederhoffer, V., Osborne, M. F. M., 1966. Market making and reversal on the stock exchange. Journal of the American Statistical Association 61, 897–916.

Patterson, D., Sharma, V., 2007. Did herding cause the stock market bubble of 1998-2001? Tech. rep., working papers, Social Science Research Network.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Press, S. J., 1967. A compound events model for security prices. Journal of Business pp. 317–335.

Rapach, D. E., Strauss, J. K., Zhou, G., 2013. International stock return predictability: what is the role of the united states? The Journal of Finance 68, 1633–1662.

Segal, G., Shaliastovich, I., Yaron, A., 2015. Good and bad uncertainty: Macroeconomic and financial market implications. Journal of Financial Economics 117, 369–397.

Tauchen, G., Zhou, H., 2011. Realized jumps on financial markets and predicting credit spreads. Journal of Econometrics 160, 102–118.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

Yu, J., 2011. Disagreement and return predictability of stock portfolios. Journal of Financial Economics 99, 162–183.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: series B (statistical methodology) 67, 301–320.

Table 1: The List of Companies in the Sample

Ticker	Company Name
1101	Taiwan Cement Corporation
1102	Asia Cement Corporation
1216	Uni-President Enterprises Corporation
1301	Formosa Plastics Corporation
1303	Nan Ya Plastics Corporation
1326	Formosa Chemicals & Fibre Corporation
1402	Far Eastern New Century Corporation
2002	China Steel Corporation
2301	Lite-on Technology Corporation
2303	United Microelectronics Corporation
2308	Delta Electronics Inc.
2311	Advanced Semiconductor Engineering Inc.
2317	Hon Hai Precision Industry Co., Ltd.
2325	Siliconware Precision Industries
2330	Taiwan Semiconductor Manufacturing Company
2354	Foxconn Technology
2382	Quanta Computer Inc.
2409	Au Optronics Corporation
2454	Media Tek Inc.
2498	HTC Corporation
2801	Chang Hwa Commercial Bank Ltd.
2880	Hua Nan Financial Holdinds Co. Ltd.
2881	Fubon Financial Holding Co. Ltd.
2882	Cathay Financial Holdings Co. Ltd.
2883	China Development Financial Holding Corporation
2885	Yuanta Financial Holding Co. Ltd.
2886	Mega Financial Holding Company
2890	Sinopac Financial Holdings Company Ltd.
2891	CTBC Financial Holding Co., Ltd.
2892	First Holdings
2912	President Chain Store Corporation
3481	Innolux Corporation
6505	Formosa Petrochemical Corporation

Table 2: Summary Statistics

N denotes the number of observations. Std. Dev. represents the standard deviation.

Variable	N	Mean	Std. Dev.	Min	25th	Median	75th	Max
Return	501696	-0.0001	0.0058	-0.1304	-0.0027	0	0.0024	0.1106
Volatility	501696	0.0014	0.0008	0	0.0009	0.0012	0.0018	0.0368
Skewness	501696	0.0262	0.6992	-6.63	-0.0901	0.0046	0.1130	6.6332
Up Jumps	509935	0.0001	0.0001	0	0.0000	0.0000	0.0001	0.0075
Down Jumps	509984	0.0001	0.0002	0	0.0000	0.0000	0.0001	0.0942
Trading Volume in Shares	501696	2149	3706	0	481	1087	2402	261300
Price Impact Measure	451512	0.0084	0.2379	-25	-0.0079	0	0.0166	24
Adverse Selection Spread	501696	0.0863	62.83	-19410	-0.0111	0	0.0160	30034
Adjusted PIN	515383	0.3023	0.2728	0	0.0847	0.1991	0.4994	0.9965
Effective Spreads	423846	-0.0000	0.0005	-0.0153	-0.0003	0	0.0002	0.0110
Realized Spreads	517374	0.0086	0.0773	-0.6900	-0.0003	0.0006	0.0017	0.7691
BSI	517373	-0.0248	0.3636	-1	-0.2793	-0.0344	0.2150	1.0000
Herding	517373	19.84	10.81	-72.08	12.78	18.34	25.13	137.47

Table 3: Correlations

$+\Delta J^2$ denotes up jumps; $-\Delta J^2$ represents down jumps in absolute values; PIM is the price impact measure, ASS is the adverse selection spread component; AdjPIN is the adjusted PIN; ES stands for effective spreads; RS represents realized spreads; BSI is the buy-sell imbalance; and Herding is the herding metric. ^{a}, ^{b}, and ^{c} represent significance at the 1%, 5%, and 10% levels, respectively.

Variable	$+\Delta J^2$	$-\Delta J^2$	PIM	ASS	AdjPIN	ES	RS	BSI	Herding
$+\Delta J^2$	1.0000	0.9204^a	0.0051^a	-0.0003	-0.0169^a	0.0055^a	0.0069^a	0.0206^a	0.1800^a
$-\Delta J^2$		1.0000	0.0045^a	-0.0005	-0.0185^a	-0.0130^a	-0.0064^a	-0.0124^a	0.1920^a
PIM			1.0000	0.0057^a	0.0018	-0.0069^a	-0.0003	-0.0025	0.0208^a
ASS				1.0000	0.0014	0.0011	-0.0002	0.0010	-0.0001
AdjPIN					1.0000	-0.0061^a	0.0022	-0.0073^a	-0.0268^a
ES						1.0000	0.0752^a	0.4969^a	-0.0246^a
RS							1.0000	0.3190^a	-0.1223^a
BSI								1.0000	-0.0507^a
Herding									1.0000

Table 4: Regression Results for Up Jumps

$+\Delta J^2$ denotes up jumps; PIM is the price impact measure; ASS is the adverse selection spread component; AdjPIN is the adjusted PIN; ES stands for effective spreads; RS represents realized spreads; BSI is the buy-sell imbalance; and Herding is the herding metric. Controls include each stock's return, volatility, and trading volume. We also control for the return, volatility, and trading volume for the Taiwan Capitalization Weighted Stock Index, which covers all of the listed common stocks at the TWSE. We take the logarithm of the monetary values. N denotes the number of observations. Heteroscedasticity-consistent test statistics are between brackets. ^{a}, ^{b}, and ^{c} represent significance at the 1%, 5%, and 10% levels, respectively.

		$+\Delta J^2$		
Variables	(1)	(2)	(3)	(4)
ΔPIM	0.000002			0.000002
	[2.65]^a			[2.37]^b
ΔASS	0.000000			0.000000
	[0.04]			[0.05]
ΔAdjPIN	-0.000000			-0.000000
	[-0.29]			[-0.23]
ΔES		0.000294		0.000484
		[0.30]		[0.42]
ΔRS		-0.000033		-0.000040
		[-4.58]^a		[-5.71]^a
ΔBSI		-0.000002		-0.000002
		[-2.53]^b		[-2.15]^b
ΔHerding			-0.000001	0.000001
			[-0.77]	[2.60]^a
N	357831	346601	438962	279127
F-statistic	2978^a	2750^a	4594^a	1576^a
Adjusted R^2	0.0697	0.0666	0.0683	0.0683

Table 5: Regression Results for Down Jumps

$-\Delta J^2$ represents down jumps in absolute values; PIM is the price impact measure; ASS is the adverse selection spread component; AdjPIN is the adjusted PIN; ES stands for effective spreads; RS represents realized spreads; BSI is the buy-sell imbalance; and Herding is the herding metric. Controls include each stock's return, volatility, and trading volume. We also control for the return, volatility, and trading volume for the Taiwan Capitalization Weighted Stock Index, which covers all of the listed common stocks at the TWSE. We take the logarithm of the monetary values. N denotes the number of observations. Heteroscedasticity-consistent test statistics are between brackets. ^{a}, ^{b}, and ^{c} represent significance at the 1%, 5%, and 10% levels, respectively.

	$-\Delta J^2$			
Variables	(1)	(2)	(3)	(4)
ΔPIM	0.000001			0.000001
	[1.87]^c			[1.87]^c
ΔASS	-0.000000			-0.000000
	[-0.10]			[-0.09]
ΔAdjPIN	-0.000000			-0.000000
	[-0.28]			[-0.11]
ΔES		0.000181		0.000287
		[0.24]		[0.32]
ΔRS		-0.000043		-0.000048
		[-4.38]^a		[-4.61]^a
ΔBSI		0.000000		0.000000
		[0.62]		[0.67]
ΔHerding			-0.000000	0.000001
			[-1.44]	[2.68]^a
N	357815	346588	438939	279114
F-statistic	2420^a	2205^a	3692^a	1276^a
Adjusted R^2	0.0574	0.0541	0.0556	0.0561

Table 6: Feature Selection Techniques

ΔJ^2 represents up jumps and $-\Delta J^2$ represents down jumps in absolute values. N denotes the number of observations. ^{a}, ^{b}, and ^{c} represent significance at the 1%, 5%, and 10% levels, respectively.

	ΔJ^2		-ΔJ^2	
Variables	LASSO	Elastic Net	LASSO	Elastic Net
ΔPIM	0.000001	0.000001	0.000001	0.000001
ΔASS				
ΔAdjPIN				
ΔES				
ΔRS	-0.000037	-0.000040	-0.000040	-0.000037
ΔBSI	-0.000001	0.000000		-0.000001
ΔHerding			0.000000	0.000001
N	279127	279127	279114	279114
F-statistic	2049^a	1841^a	1841^a	2048^a
Adjusted R^2	0.0683	0.0560	0.0560	0.0684

Table 7: Principal Component Analysis

The principal components with eigenvalues above one are reported. The variables that we use for the PCA are PIM, RS, BSI, and Herding.

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.3081	0.2963	0.3270	0.3270
Factor 2	1.0117		0.2529	0.5800

Table 8: Varimax-rotated Factor Pattern

Values are multiplied by 100 and rounded off to the nearest integer. Values greater than 0.4 are flagged by an asterisk.

variables	Factor 1	Factor 2
PIM	14	84*
RS	79*	-8
BSI	78*	0
Herding	-20	55*

107年度專題研究計畫成果彙整表

計畫主持人：蔡蒔銓	**計畫編號：**107-2410-H-003-020-MY2
計畫名稱：臺灣股票市場高頻方向性價格跳躍之研究	

成果項目			量化	單位	質化 （說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等）
國內	學術性論文	期刊論文	0	篇	
		研討會論文	0		
		專書	0	本	
		專書論文	0	章	
		技術報告	0	篇	
		其他	0	篇	
國外	學術性論文	期刊論文	1	篇	已投稿至 Pacific-Basin Finance Journal
		研討會論文	1		15th Conference on Asia-Pacific Financial Markets (in Korea)
		專書	0	本	
		專書論文	0	章	
		技術報告	0	篇	
		其他	0	篇	
參與計畫人力	本國籍	大專生	0	人次	
		碩士生	17		聘任跨校多名碩士生，提供相關研究、程式、大數據、資料庫管理等訓練，並提供團隊合作的成功經驗與機會，對學生未來就業及發展有很大的助益。
		博士生	0		
		博士級研究人員	0		
		專任人員	0		
	非本國籍	大專生	0		
		碩士生	0		
		博士生	0		
		博士級研究人員	0		
		專任人員	0		

其他成果 （無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。）	本研究後續與韓國籍教授 Yongkil Ahn (Seoul National University of Science and Technology) 進行國際合作，共同研究與發表論文，擴大研究成果之國際影響力。