ADVANCED APPLICATIONS THROUGH INTRADAY DATA:

BEHAVIORAL FINANCE AND DYNAMICS OF TRADING

BIN ZHAO
WASIN SIWASARIT

https://mif.tbs.tu.ac.th/
https://www.cmdf.or.th/

Advanced Applications through Intraday Data: Behavioral Finance and Dynamics of Trading

Bin Zhao

Wasin Siwasarit

2025

Thammasat Business School

Advanced Applications through Intraday Data: Behavioral Finance and Dynamics of Trading

Authors: Bin Zhao and Wasin Siwasarit

This book is part of a project titled "Market Microstructure of the Stock Exchange of Thailand: Data platform and applications" under financial support from the Capital Market Development Fund (CMDF)

ISBN 978-616-629-862-8

Published by: Bin Zhao and Wasin Siwasarit

Editor Pantisa Pavabutr

Printed by: Thammasat University Press

Rangsit Center, Thammasat University Press Building

Khlong Neung, Khlong Luang, Pathum Thani 12121

Tel: 02564-4440-79 ext 1204-1205

Cover Design: Nattha Klieothong

nattha.klieothong@gmail.com

Contents

Acknowledgements

The authors have benefited from helpful suggestions from colleagues at the Capital Market Development Fund (CMDF), the Stock Exchange of Thailand (SET), and at Thammasat Business School. Special thanks go to Chanant Charnchainarong, Juckchai Boonyawat, Soraphol Tulayasathien, and Thanaporn Chanakularee. We appreciate data assistance and advice from Sumitra Tangsomworapong. This work has received financial support from CMDF.

About the Authors

Dr. Bin Zhao

is an Associate Professor of Finance at Thammasat Business School. Before this, she has worked as assistant professor in finance at NYU Shanghai and Shanghai Advanced Institute of Finance, Shanghai Jiaotong University. Her research interest is on behavioral finance and stock markets. She has been in the joint research plan of Shanghai Stock Exchange, she also had consulting experience for an investment bank and hedge fund. Bin completed her Ph.D. in Finance from University of Wisconsin, Madison.

Dr. Wasin Siwasait

is an Assistant Professor of Finance at Chulalongkorn Business School. His research is in the areas of Asset Pricing (Skewness in Asset Pricing), and Financial Econometrics (MIDAS: Mixed Data Sampling). He received his PhD. in Economics from University of North Carolina at Chapel Hill, U.S.A. in 2015.

Preface

Wider adoption of technology has made equity markets more fascinating as we see prices changing more rapidly to new information, and winning traders replace losing traders as technology evolve. Under the current investment landscape, the study of market microstructure lays out the foundation to understanding the limitations of efficiency market hypothesis stemming from investor irrationality, information asymmetry, and trading costs.

The book begins by describing how learning about information move prices. The word information is extremely broad, ranging from public fundamental information to variety of alternative datasets that may affect investor sentiment. Designing event studies that can capture responses to scheduled public announcements and unscheduled shocks are important in assessment of market quality that is important concern for law and regulation, not just for academic curiosity. The book ends with a discussion on intraday volatility modeling with python code. Few books try to blend the two disciplines. Finance-trained people may not be keen on programming and computer scientists may not grasp financial economics when they develop trading and prediction models.

This book is part of the part of a project titled “Market Microstructure of the Stock Exchange of Thailand: Data platform and applications” under financial support from the Capital Market Development Fund (CMDF). The content consists of graduate-level information on market microstructure for those who wish to begin empirical studies on the topic. Related to this project are two other books. One is titled “The Stock Exchange of Thailand Intraday Data Handbook”, describes the Stock Exchange of Thailand’s intraday data and how to filter the original data to obtain stock sets and time intervals of interest. The other, written in Thai, is titled, “The market microstructure of equity markets: Understanding liquidity, trading, and volatility.” Readers can find a companion website, https://sites.google.com/tbs.tu.ac.th/market-microstructure, that includes intraday data related material, updates, and teaching materials such as short-videos and examples.

Bin Zhao and Wasin Siwasarit

Chapter 1
Foundations of Market Microstructure

1.1 Understanding the limits of Efficient Market Hypothesis (EMH) with intraday data

One of the foundational theories in finance is the Efficient Market Hypothesis (EMH), which posits that financial markets incorporate all available information quickly and accurately into asset prices. This theory rests on three key assumptions:

1. Investors are rational;
2. If some investors are irrational, their behaviors are random and cancel each other out;
3. If a subset of investors is irrational in systematic ways, rational arbitrageurs will step in to eliminate mispricing.

However, a substantial body of empirical research challenges the notion of consistently rational markets. Numerous studies have documented predictable patterns in investor trading behavior, suggesting that markets may not always be efficient.[1]

---

[1] See Zhao, 2025, for a comprehensive review of related concepts and evidence.

For example, some investors appear to trade on superior or advanced information, raising questions about the distribution and processing of information across investor types. Using intraday level data enables the documentation of many important and interesting behavioral patterns for different types of investors that reveal how their trading intensity, timing, and reactions to news differ across the trading day. The analyses provide valuable insights into how information, sentiment, and market frictions shape trading behavior throughout the day and thus cause the market to depart from efficiency. By examining high-frequency trading data, researchers can also capture the dynamics of order submission, execution, and cancellation in real time, allowing for a more nuanced understanding of market behavior and asset pricing dynamics that traditional daily data cannot provide. In this chapter, we introduce a key sources of market microstructure components that can aid or limit the path to efficient price discovery. These factors include the concept of incomplete information, investor heterogeneity, and behavioral biases that can be measured via trading activity such as turnover.

1.2 Information Asymmetry in Markets

Financial markets are rarely homogeneous in information. Some investors may possess superior knowledge or analytical abilities, while others act on limited or noisy signals. Understanding how different investor types process and act on information is central to empirical market microstructure research. Asymmetric information affects liquidity, spreads, price impact, and the overall efficiency of markets.

1.3 Investor Types on the SET

Intraday data from the Stock Exchange of Thailand (SET) classifies investors into three types:

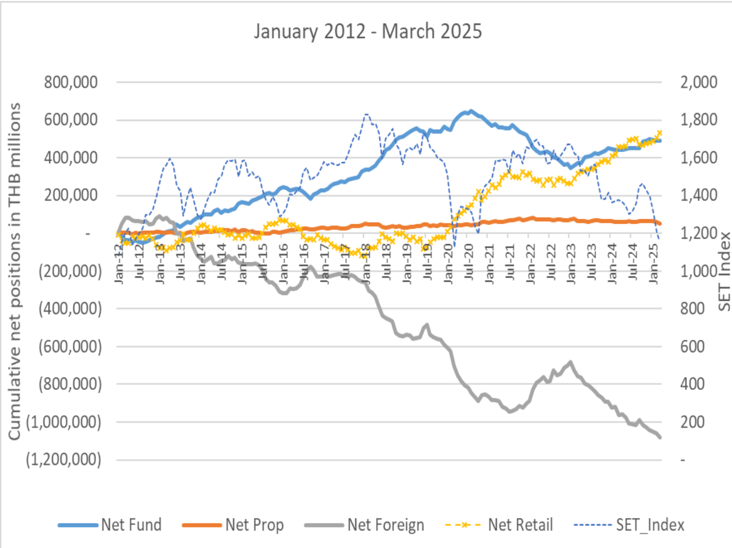
- **Domestic individual investors** – Often referred to as "noise traders," these investors frequently trade excessively and exhibit behavioral biases such as the disposition effect (selling winners too early and holding losers too long) and herding.
- **Domestic institutional investors** – Typically mutual funds or other large domestic investors, they tend to trade less frequently and with more discipline. Their trades are often predictive of future stock returns, suggesting informational advantages or more structured decision-making. Included in this category are

proprietary traders who are market makers and trading activities for own brokerage accounts are also included.

- **Foreign investors** – Particularly in emerging markets, foreign investors may have superior information-processing capabilities. Research using SET data (Bailey et al., 2007, 2012) and similar studies in China (Lundblad et al., 2023) indicate that foreign investors can develop local informational advantages that affect market behavior.

This composition mirrors that of many global markets, allowing findings from Thailand to speak to broader patterns in investor behavior. Over the past 20 years (circa 2012-2025), retail investors, domestic institutional investors, and foreign investors account for 44%, 20%, and 36% of total market trading activities. Figure 1.1 reveals net buying positions by investor type from 2012-2025. As buying and selling needs to net out among the traders, we observe that throughout this period, foreign investors are net sellers whereas retail and local funds have been net buyers.

Figure 1.1 Cumulative net buying positions in THB millions by investor type on SET between 2012-2025.



Source: SET Computations by authors

1.4 Measures of Investor Activity

To study investors' trading behaviors empirically, researchers rely on both **trade data** and **order data**,[2] from which key metrics are constructed to assess **information acquisition and**

[2] The SET provides both trade data and order book data available in "deal" files and "order" files which we discuss in detail in Pavabutr and Zhao (2025).

processing skills. One fundamental measure used in this context is **turnover**, which can be used not only to capture general trading activities, but also to gauge speculative behavior. We describe measures of turnover in detail next.

1.4.1 Turnover

Turnover is a widely used measure of trading intensity. It can be calculated at the individual stock level or at the portfolio level, depending on the research objective.

Stock-Level Turnover

At the stock level, turnover is defined as the ratio of the total number (or value) of shares traded to the total number (or value) of shares outstanding over a specific time period. For example, to calculate the annual turnover rate for stock *i*:

- By quantity of shares:

$$Turnover_{i,j} = \frac{Total\ \#\ of\ shares\ \text{of stock}\ i\ traded\ in\ year\ j}{Average\ daily\ number\ of\ shares\ outstanding_{i,j}}$$

- By value of shares:

$$Turnover_{i,j} = \frac{Value\ of\ shares\ \text{of stock}\ i\ traded\ in\ year\ j}{Average\ daily\ value\ of\ shares\ outstanding_{i,j}}$$

If not all shares are tradable (due to lock-ups or regulatory restrictions), the denominator may be adjusted to include only tradable shares.[3]

Portfolio-Level Turnover

Portfolio turnover reflects how actively a specific investor or fund trades. It is especially useful for analyzing fund manager behavior or individual investor tendencies. High turnover may indicate excessive trading, risk tolerance, or gambling-like behavior (Hoffman et al., 2013; Kumar et al., 2021).

A common approach is to calculate portfolio turnover as the proportion of assets traded over a given period. The denominator typically reflects the average value of the portfolio during the period, while the numerator represents trading activity.

Below are four common definitions of portfolio turnover:

1. Turnover 1 – *Minimum of Buy and Sell Values*:

$$Turnover1 = \frac{\text{Min}(buy\ value, sell\ value)}{Average\ Portfolio\ Holding\ Value}$$

[3] For SET stocks, we can look up the firm's percentage free float from www.set.or.th and search for firm ticker and select shareholder information.

This measure is widely used by the U.S. Securities and Exchange Commission (SEC) and Morningstar. It reflects the proportion of the portfolio that was turned over during the period. For instance, a turnover of 1.0 per year suggests an average holding period of one year, while a turnover of 2.0 implies a six-month holding period.

2. Turnover 2 – *Average of Buy and Sell Values*:

$$Turnover2 = \frac{Average\ (buy\ value, sell\ value)}{Average\ Portfolio\ Holding\ Value}$$

3. Turnover 3 – *Minimum of Buy and Sell Values over Max Holdings*:

$$Turnover3 = \frac{Min\ (buy\ value, sell\ value)}{Maximum\ Portfolio\ Holding\ Value}$$

4. Turnover 4 – *Average of Buy and Sell Values over Max Holdings*:

$$Turnover4 = \frac{Average\ (buy\ value, sell\ value)}{Maximum\ Portfolio\ Holding\ Value}$$

If only trade-level data are available (and not daily holdings), researchers must reconstruct portfolio holdings by cumulating trades over time. The initial holdings at the beginning of the study period may be inferred based on net future trading. In such cases, turnover reflects portfolio adjustment frequency rather than exact holding periods.

1.4.2 Trading Imbalance

Trading Imbalance is another key measure for investor activity. It is normally measured at account type level. For example, we can look at whether mutual funds have the same trading preference for certain type of stocks as individual investors. Do mutual funds tend to sell those stocks whereas individual investors tend to buy those stocks at the same time? By cumulating trading imbalance over a sample period, it can give us very important trend and information about investors' preference and action. In addition, some studies have used this measure to see whether certain type of investors' trade can predict future stock returns, or whether certain investors' trade contain important information. There are different ways to measure it either based on trade or order data.

Trade Based Trading Imbalance

One way to measure this is to look at trading imbalance, or how much more investors are buying or selling. For example, following Kaniel et al. (2008) and Titman et al. (2022), trading imbalance is defined as Net Buy, which is cumulative value of shares bought minus cumulative value of shares sold, normalized by the average daily volume which is the average daily volume over the past trading year. Formally, for each stock i, on each day t, within each investor group j, net buying is defined as

$$Net\ Buy_{i,t,j} = \frac{\sum_{i,t,j} Buy\ Value - \sum_{i,t,j} Sell\ \ Value}{Average\ Daily\ Volume_{i,t-1,j}}$$

This Net Buy can then be cumulated for a certain sample period to demonstrate trading pattern.

Order Based Trading Imbalance

Alternatively, order data can be used to assess how aggressive different investor types are in submitting marketable orders versus limit orders. This complements trade-based measures and can highlight information processing behavior.

More formally, following Chordia and Subrahmanyam (2004), for stock i, day d, and investor group G, order flow imbalance is calculated as

$$Oib_{i,d,G} = \frac{\sum_{j\in G} BuyVol_{i.d.j} - \sum_{j\in G} SellVol_{i,d,j}}{\sum_{j\in G} BuyVol_{i.d.j} + \sum_{j\in G} SellVol_{i,d,j}}$$

1.5 Behavioral Biases and Trading Patterns

While traditional market microstructure focuses on how orders, prices, and liquidity are shaped by rational decision-making and institutional rules, actual trading behavior often deviates from these predictions. Different investor groups—individuals, institutions, and foreign investors—exhibit behavioral biases that influence not only their own performance but also the functioning of financial markets. With access to intraday order and trade data, we can document these biases empirically through observable patterns such as turnover, trading imbalance, and order aggressiveness.

1.5.1 Overconfidence of Individual Investors

Overconfidence is one of the most robust behavioral biases among individual investors. Many retail traders overestimate their knowledge or ability to predict market movements, leading them to believe they have an informational advantage even when they do not. This results in excessive trading, often with larger trade sizes relative to their portfolio, and contributes to high portfolio turnover. Empirical research shows that retail investors' trades are frequently contrarian and fail to generate superior returns, highlighting the costs of overconfidence. In intraday data, overconfidence manifests as frequent small trades, disproportionate market order submissions, and high turnover ratios.

1.5.2 Herding Behavior

Herding occurs when investors mimic the trades of others rather than relying on independent analysis. This behavior is especially prevalent among individuals who may lack the time or resources to evaluate information fully. Herding can create clusters of buying or selling activity, amplifying price swings and sometimes fueling bubbles or crashes. Intraday trading imbalance measures are useful in detecting herding, as they capture simultaneous shifts in net buying or selling across investors in the same group. Retail herding often reflects emotion-driven responses to recent price movements, while institutional herding may be driven by benchmarking or reputational concerns.

1.5.3 The Disposition Effect

The disposition effect refers to investors' tendency to sell winners too quickly and hold losers too long. This bias reflects both regret aversion and the desire to "lock in" gains. Retail investors are most prone to the disposition effect, though institutional investors are not immune. At the intraday level, the disposition effect can be observed in net selling of recently appreciated stocks and continued buying or holding of declining stocks. The effect distorts liquidity provision, reduces portfolio efficiency, and contributes to predictable return patterns.

1.5.4 Order Submission Strategies

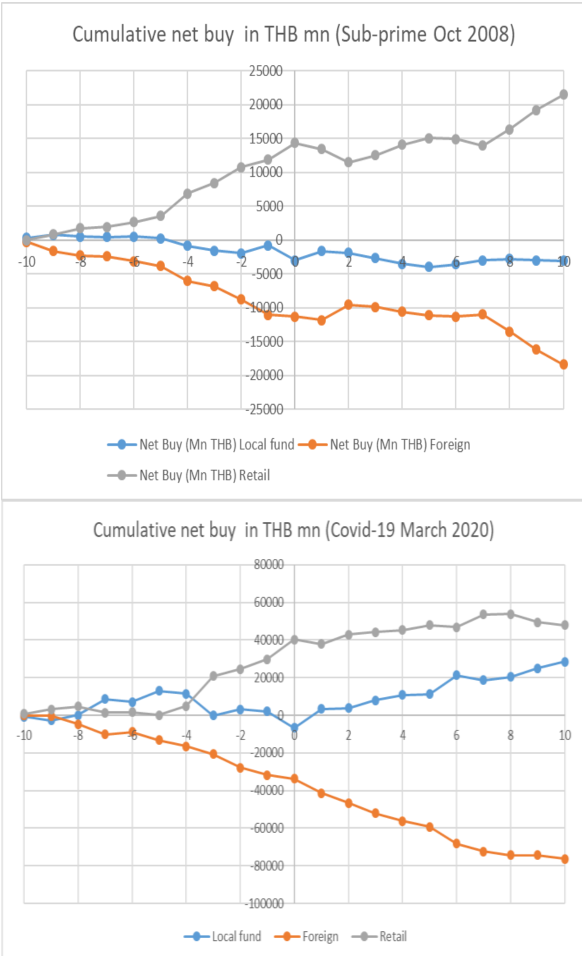
Investor behavior is also evident in how orders are submitted. Market orders reflect impatience or urgency, while limit orders represent a willingness to wait for execution at favorable prices. Retail investors, driven by overconfidence or impatience, are more likely to submit market orders, often at a cost. Institutional investors and foreign investors tend to use limit orders, order splitting, or iceberg orders to minimize market impact and conceal their trading intentions. Intraday order book data can reveal these patterns, allowing researchers to compare aggressiveness across investor types and link order strategies to behavioral biases.

1.5.5 Liquidity Provision and Behavioral Response

Liquidity provision is not purely mechanical; it reflects behavioral responses to risk and uncertainty. Retail investors often withdraw from liquidity provision during volatile periods, either canceling limit orders or switching to aggressive trades out of fear.

Institutions, with more resources and discipline, may step in to provide liquidity when spreads widen, while foreign investors may choose selectively when to supply or demand liquidity depending on their information advantage. Interestingly, Thai intraday data of cumulative net buy 10 days around the announcement of Lehman Brothers' backruptcy in October 2008 and Covid-19 lockdown in March 2020 reveals that retail investors are suppliers of liquidity during volatile times as shown in Figure 1.2. Empirical measures such as bid-ask spreads, order book depth, and cancellation rates provide a window into these behavioral dynamics.

Figure 1.2 Cumulative net trading positions by investor type -10/10 days around event announcement sub-prime (October 2008) and Covid-19 lockdown (March 2020)



Source: SET Computations by authors

1.5.6 Intraday Patterns of Behavior

Trading activity varies systematically over the course of a trading day, and these patterns differ across investor groups. Retail investors are often most active at the opening and closing auctions, reacting emotionally to overnight news or attempting to time the market close. They are also more likely to display behavioral biases such as overreaction to short-term price changes, herding, and a preference for trading at specific times of the day—often near market openings or closings. In contrast, institutional investors, with greater access to information, advanced trading technologies, and lower transaction costs, tend to exhibit more strategic trading patterns. They often trade in larger volumes, time their orders to minimize market impact, and play a leading role in the process of price discovery. Empirical evidence suggests that institutional trades are generally more informed and profitable, while retail trading tends to be driven by sentiment and less related to fundamentals. These distinctions not only reflect varying levels of sophistication and information access but also contribute to the overall structure and dynamics of intraday market activity. Foreign investors may concentrate trading at particular hours—such as around macroeconomic announcements in global markets. These time-of-day patterns can be documented through intraday volume, order imbalance, and spread changes, offering insights into behavioral timing.

1.5.7 Volatility, Uncertainty, and Behavioral Trading

Periods of high volatility and uncertainty provide a natural laboratory for behavioral analysis. Retail investors often herd more aggressively during volatile periods, either chasing momentum or fleeing from risk. Institutions may act as stabilizers, providing liquidity and trading strategically to exploit temporary mispricing. Foreign investors may either retreat due to uncertainty or intensify trading if they perceive an informational advantage. Volatility's effect on behavior can be captured empirically through measures such as conditional turnover, spread widening, and order aggressiveness, helping to differentiate behavioral responses across investor groups.

1.5.8 Behavioral Asymmetries Across Investor Types

Finally, behavioral biases are not uniform across investors—they differ systematically between individuals, institutions, and foreign participants. For example, retail investors often overreact to earnings surprises, while institutions adjust positions more cautiously. Foreign investors may anticipate regulatory announcements or macroeconomic shocks more accurately, reflecting informational advantages. Intraday data from the Stock Exchange of Thailand (SET), which classify trades by investor type, are especially well-suited to studying these asymmetries. By comparing net buying, turnover, and order aggressiveness across groups, researchers can identify which investors drive market reactions and how behavioral differences shape liquidity, volatility, and efficiency.

References

Bailey, W., Mao, C. X., & Sirodom, K. (2007). Investment restrictions and the cross-border flow of information: Some empirical evidence. *Journal of International Money and Finance, 26*(1), 1-25.

Bailey, W., Mao, C. X., & Sirodom, K. (2012). Locals, foreigners, and multi-market trading of equities: Intraday evidence from Thailand. Pacific-Basin Finance Journal, 20, 101-121.

Hoffmann, A. O. I., Post, T., & Pennings, J. M. E. (2013). Individual investor perceptions and behavior during the financial crisis. *Journal of Banking & Finance, 37*(1), 60–74.

Kaniel, R.; Saar, G., & Titman, S. (2008). Individual Investor Trading and Stock Returns. *The Journal of Finance, 63*(1), 273-310.

Kumar, P., Nguyen, T., & Putnins, T. J. (2021). Dark trading and price discovery. *Review of Financial Studies, 34*(2), 792–837.

Lundblad, C. T., Shi, D., Zhang, X., & Zhang, Z. (2023). *Are foreign investors informed? Trading experiences of foreign investors in China.* Working paper. SSRN

Titman, S., Wei, C., & Zhao, B. (2022). Corporate actions and the manipulation of retail investors in China: An analysis of stock splits. *Journal of Financial Economics*. (Forthcoming)

Pavabutr, P., & Zhao, B. (2025). *The Stock Exchange of Thailand Intraday Data Handbook.* Thammasat University Press.

Zhao, B., (2025). *Behavioral Finance: The market, investors, and corporations.* Thammasat University Press.

Chapter 2
Applications for Research

2.1 Linking Theory and Data

Existing empirical evidence underscores the importance of examining intraday trading behavior to better understand how different investor groups interact and how their actions collectively shape market outcomes. Despite the rich body of international research, there remains a need to explore these behavioral patterns in specific market contexts, particularly in emerging or developing markets where trading mechanisms, investor composition, and regulatory environments differ significantly from those in mature markets. By analyzing detailed intraday data, this book demonstrates how such analysis can contribute to the growing literature by identifying the behavioral tendencies of different investor types, assessing their role in price formation, and evaluating the extent to which psychological factors influence trading dynamics within the day. Such analysis not only deepens our understanding of investor behavior but also provides practical insights for market regulators, brokers, and policymakers seeking to enhance market quality and efficiency.

Building on the empirical evidence of intraday behavioral patterns and investor heterogeneity, it becomes essential to connect these observations with theoretical frameworks that explain how information and behavior translate into price dynamics. Market microstructure models, such as the Kyle (1985) and Glosten–Milgrom (1985) frameworks, provide the conceptual tools to interpret how informed and uninformed trading, order flow, and liquidity provision interact within a trading day. By linking theory with high-frequency data, researchers can not only quantify the impact of behavioral biases on prices but also examine the mechanisms through which private information is incorporated into the market. This section explores these models in detail, illustrating how intraday data from the SET can be used to operationalize theoretical constructs and uncover the dynamic relationship between investor behavior and market outcomes.

In brief, the Kyle model provides insights into how a strategically informed trader optimally spreads trades over time, balancing profit opportunities against the price impact of their actions. In contrast, the Glosten–Milgrom (GM) model emphasizes sequential trading and bid–ask spreads, showing how market makers update beliefs and prices in response to potentially informed trades. Together, these models offer complementary perspectives: one focuses on the behavior of an individual informed trader within a continuous market, while the other highlights the information processing role of the market as a whole. By applying these frameworks to SET intraday data, researchers can link theoretical predictions to observable patterns, such as order flow, trading imbalances, and short-term price adjustments, thereby bridging behavioral insights with formal market microstructure analysis.

2.2 Kyle model (1985): Informed trading and price impact

Overview and key insights

The basic setting is as the following:

- **One risky asset** is traded in a competitive auction market.
- **Three types of agents**:

1. **Informed trader (insider):** Knows the true value of the asset before trading.
2. **Noise traders:** Submit random buy/sell orders (exogenous liquidity demand).
3. **Market maker (dealer):** Competitive, sets prices to clear the market. The market maker does not know which orders are informed or noise, only observes the *total* order flow.

$$Q_t = x_t + u_t$$

Where x_t is insider's trade, which is strategic and based on private information; u_t is noise trader demand which is random and uninformed. The core idea is that the informed trader wants to exploit private information (about the true value of the asset). But, if they trade too aggressively, the market maker will infer the information from order flow and adjust the price, reducing profits. If the insider trades too cautiously, the information advantage is wasted. So, the insider trades gradually and strategically to "camouflage" their trades among the noise traders.

This trade-off creates **Kyle's λ** (lambda): a measure of price impact or an inverse measure of market depth.

$$P_t = \mathbb{E}[V|Q_t] = P_0 + \lambda Q_t$$

where:

- P_t is the transaction price at time t.
- V is the true asset value.
- Q_t is the total signed order flow (informed + noise).
- λ is the *Kyle lambda*, the slope of the price impact function.

In equilibrium:

The insider uses a linear strategy:

$$x_t = \beta(V - P_{t-1})$$

where x_t is the insider's trade size.

This means if the asset is underpriced $(V > P_{t-1})$, the insider buys proportionally, the larger the mispricing, the more they trade. But since $\beta < 1$, so they only trade part of their informational advantage each period to disguise themselves among noise traders.

The market maker sets price competitively, conditional on total order flow:

$$P_t = P_{t-1} + \lambda Q_t = P_{t-1} + \lambda(x_t + u_t)$$

where u_t is the noise trade.

λ balances informativeness: enough price sensitivity that information gets incorporated over time, but not so high that no trading occurs. Essentially, λ is the *price impact coefficient*, which measures how much the price moves per unit of order flow. Economically, it reflects market depth: a small λ means the market is deep (orders have little effect), a large λ means the market is thin (orders move the price a lot).

Because of the insider's strategic behavior, prices partially adjust to private information in each round. Over time, as the insider keeps trading, the information gets fully incorporated into prices.

Regarding the interaction between strategy and price impact: the insider knows their trades will move the price by $\boldsymbol{\lambda}$, profit per share is $V - P_t$.

If they buy too much, P_t rises toward V, shrinking the margin. So, they spread trades over time ("drip-feed" strategy).

This creates a gradual revelation of information. In the early periods: the insider earns positive profits because prices are still below V. In the later periods: as the market maker updates beliefs from order flow, the price converges toward V, and the insider's edge disappears.

We can also have a simple numerical example of Kyle's model to make it easier to understand.

Step 1. Setup

- True value of the asset:

$$V = 100$$

(only the informed trader knows this).

- Market maker's prior belief (unconditional mean):

$$E[V] = 90$$

- Noise trader order:

$$u = +10$$

(a random buy order, unknown to market maker in advance).

- Kyle equilibrium says the price impact rule is linear:

$$P = 90 + \lambda Q$$

where $Q = x + u$, insider's order x plus noise trader order.

Let's assume (for illustration) that $\lambda = 0.5$.

Step 2. Insider's problem

The insider wants to buy because $V = 100$ is higher than the current market maker's belief of 90.

- Profit if insider buys x:

  $\pi(x) = (V - P)\,x = (100 - (90 + 0.5(x + u)))x$

  Plug in $u = 10$:

$$\pi(x) = \big(100 - \big(90 + 0.5(x + 10)\big)\big)\,x = (10 - 0.5x - 5)x$$
$$= (5 - 0.5x)x$$

- Maximizing profit:

  $\pi(x) = 5x - 0.5x^2$

  First-order condition:

$$5 - x = 0 \Rightarrow x = 5$$

  So the insider buys 5 shares.

Step 3. Market clearing price

Total order flow:

$$Q = x + u = 5 + 10 = 15$$

Price set by market maker:

$$P = 90 + 0.5 \times 15 = 97.5$$

Step 4. Outcomes

- **Informed trader's profit:**

$$(100 - 97.5) \times 5 = 12.5$$

- **Noise trader:** Bought at 97.5 when the true value is 100 ⟶ small gain (2.5 per share, but they were not strategic).

- **Market maker:** Breaks even in expectation (given they don't know whether order flow was from noise or insider).

Intuition

- The insider does not buy all 10 shares undervalued (100 - 90).
- They buy moderately (5 shares) to hide among the noise order of +10.
- The price moves toward the true value (97.5 vs. 100) but does not fully reveal it in one round.
- If there were multiple rounds, the insider would keep trading gradually, and the price would converge to 100 over time.

The key insights of the Kyle's model are the following:

- Price impact is linear in order flow.
- Liquidity (market depth) is finite: prices do not fully adjust instantly to information, but information gets incorporated gradually as insider trades unfold.
- Informed traders trade smoothly, not all at once, to hide among noise and maximize profits.
- Kyle's λ captures this balance: high noise trading implies low λ, hence insider trades more aggressively. Thin markets means high λ, so insider trades cautiously.
- **Market makers infer information from total order flow**, but cannot distinguish perfectly between informed and uninformed trades.

Applications of Kyle's model

1. Benchmark for market impact and liquidity measurement (Kyle's λ)

Kyle's λ measures how much prices move in response to order flow as the following equation:

$$\Delta P = \lambda Q$$

where Q = net order flow.

Researchers estimate λ by regressing price changes on signed trade volume (order imbalance). This makes λ a practical proxy for price impact and market depth.

In practice, λ can be used in liquidity studies, trading cost analysis, and risk management. For liquidity, asset classes with high λ (illiquid bonds, small-cap stocks) have large price impacts from trading. For trading cost analysis, quant desks and asset managers use λ to model the *slippage* costs of executing large orders. For risk management, regulators use λ-like measures to assess how vulnerable markets are to order-flow shocks (e.g., stress testing).

For example, if λ is high, then even small trades move prices a lot, which implies thin, illiquid market. On the other hand, if λ is low, the market is deep and resilient.

2. Understanding how private information gets incorporated into prices

Kyle's model shows that informed traders don't dump all their information at once. Instead, they trade gradually to hide among noise, so information is revealed to the market *over time*. There are several applications of this insights in finance research such as price drifts after earnings announcements, predictability of order imbalances, and the speed that prices reflect private information. In general, the model formalizes the idea of price discovery, i.e., how order flow from informed traders gradually pushes the price toward fundamental value.

3. Basis for later models of HFT optimal execution, and market design

For High-Frequency Trading (HFT): Kyle's framework has been extended to study how HFTs use speed to exploit short-lived informational advantages. HFT strategies (like "sniping" or "market making with private signals") can be analyzed as variations of informed trading in Kyle's setting.

For Optimal Execution (Trading algorithms): Traders executing large orders face a Kyle-type trade-off: trading too aggressively moves the price against them. Modern execution algorithms (like Value Weighted Average Price (VWAP), Time Weighted Average

Price (TWAP), Almgren-Chriss (2000) models) are directly inspired by Kyle's gradual trading logic.

For Market Design: Regulators use Kyle-type models to study how changing tick size, transparency, or trading rules affect informed vs. uninformed traders. For example, regulators can design dark pools or block trading venues to reduce price impact of large orders.

There are additional applications beyond the classic three applications we discussed above. For example, in empirical market microstructure, Kyle's model is used to estimate how much of order flow is informed vs. liquidity-driven (e.g., Easley & O'Hara's PIN model; discussed later in this Chapter, builds on Kyle). In corporate finance: Kyle's model helps explain insider trading behavior, share repurchases, and how informed shareholders affect stock prices. In macro-finance, Kyle's model provides a framework for how information asymmetry impacts market volatility and systemic risk. To summarize, Kyle's model is not just a theoretical curiosity — it underpins practical measures of liquidity, helps us understand price discovery, and serves as the foundation for modern trading and market design.

4. Testing for informed trading.

We can directly use intraday data to apply Kyle's model in research. Since we can't directly observe who has private information, we rely on empirical proxies like turnover and order/trading imbalance to test hypotheses about informed trading.

Turnover

We can use the formula from Chapter 1 to compute turnover. The idea is that if informed traders are active, we should see unusually high turnover around times when information is released (or expected to be released). Direct applications include earnings announcements and insider trading. Spikes in turnover suggest private information trading particularly before major event announcements. However, since high turnover could also reflect noise trading or liquidity shocks, turnover, by itself is an imprecise but useful signal.

Trading (Order Flow) Imbalance

We can use the formula from Chapter 1 to compute trade or order flow imbalance. Kyle (1985) points out that insider order flow is systematically tilted toward the direction of mispricing, which implies that positive imbalance (excess buys) means possibly informed buying if asset is undervalued, whereas negative imbalance (excess sells) suggests possibly informed selling if asset is overvalued. In applications, studies often find that order imbalances predict short-term returns — consistent with information being gradually incorporated into prices. In addition, institutional investors' imbalances (vs. retail) are especially informative, since institutions are more likely to have research-based informational advantages.

Investor type and informed trading

By combining these proxies with investor-level data (if available), we can test whether institutional investors are better informed than retail investors. If so, then their order imbalances are more predictive of future returns. A similar measure can be imposed in a study to compare superior information between foreign and local investors or between users of HFTs and traditional technology.

However, we also need to note some limitations of these direct test. Since neither turnover nor imbalance perfectly distinguishes informed trading from liquidity trading. Hence, careful research design are needed, such as choosing event windows, conduct cross-sectional tests, and execute robustness checks. Nevertheless, many empirical papers use Kyle's λ itself as a liquidity measure estimated from data (regressing price changes on order flow).

Here are some classic and widely cited empirical papers that can be used as great reference which use turnover or order imbalance as proxies for informed trading or information-based trading activity.

Table 2.1 Summary of selected empirical papers on liquidity and informed trading motivated by Kyle

Theme	Citation	Key Insight
Turnover (Trading Volume) and Information	Campbell, John Y., Sanford J. Grossman, and Jiang Wang (1993). *Trading Volume and Serial Correlation in Stock Returns*. QJE.	Volume relates to information flow and return predictability.
	Lee, C. M., & Swaminathan, B. (2000). Price momentum and trading volume. *the Journal of Finance*, *55*(5), 2017-2069.	High turnover is linked to momentum continuation, consistent with gradual information diffusion.

Table 2.1 Summary of selected empirical papers on liquidity and informed trading motivated by Kyle (Cont.)

Theme	Citation	Key Insight
	Chordia, Tarun, and Avanidhar Subrahmanyam (2004). *Order Imbalance and Individual Stock Returns*. JFE.	Imbalance and volume measures explain short-term stock returns.
	Hong, Harrison, and Jeremy C. Stein (2007). *Disagreement and the Stock Market*. Journal of Economic Perspectives.	Turnover reflects heterogeneous beliefs and information.
Order Imbalance and Informed Trading	Kyle, Albert S. (1985). *Continuous Auctions and Insider Trading*. Econometrica.	Price impact is linear in net order flow — foundational theory.
	Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance, 46*(1), 179–207.	Develops econometric methods linking order flow to price discovery.

Table 2.1 Summary of selected empirical papers on liquidity and informed trading motivated by Kyle (Cont.)

Theme	Citation	Key Insight
	Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman (1996). *Liquidity, Information, and Infrequently Traded Stocks*. JF.	Introduces PIN measure based on order imbalances.
	Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2002). *Order Imbalance, Liquidity, and Market Returns*. JFE.	Aggregate order imbalances predict market returns and affect liquidity.
	Griffin, John M., Jeffrey H. Harris, and Selim Topaloglu (2003). The Dynamics of Institutional and Individual Trading. Journal of Finance.	Institutional imbalance predicts short-term returns more than retail.

Table 2.1 Summary of selected empirical papers on liquidity and informed trading motivated by Kyle (Cont.)

Theme	Citation	Key Insight
Linking Investor Types and Superior Information	Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance, 14*(5), 590–610.	Retail trading imbalances are predictable and not very profitable.
	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.	Retail turnover is high but unprofitable — suggests lack of informational edge.
	Kaniel, R., Saar, G., & Titman, S. (2008). Individual investor trading and stock returns. *Journal of Finance, 63*(1), 273–310.	Retail imbalance sometimes predicts returns, suggesting liquidity provision rather than informed trading.

2.3 Glosten-Milgrom (1985) model: Sequential trade and spread decomposition

Overview and key insights

The Glosten-Milgrom (GM) model (1985) is a microstructure model that explains how bid-ask spreads arise in markets with asymmetric information. It builds on the idea that some traders have private information about the true value of an asset, while others (called noise traders) trade for liquidity or exogenous reasons.

The key points are: the market has dealers (or market makers) who set prices; there are informed traders (who know the true value) and *uninformed/noise traders* (who trade randomly); and prices adjust sequentially after observing trades because each trade conveys information about the likelihood that the trader is informed.

The GM model introduces sequential trade, meaning trades occur one at a time, and the market maker updates beliefs after each trade.

The mechanics are as follows:

1. Dealer starts with a prior belief about the asset's value, v.
2. A trader arrives and either buys or sells:
   - If the trade is from an informed trader, it is likely in the direction of the true value.

- If the trade is from a noise trader, it is random.

3. Dealer observes the trade but not the trader's type, and updates the posterior belief using Bayes' rule.

4. Dealer sets next quotes based on updated beliefs.

This sequential updating leads to price movements that reflect the probability that a trade was informed.

In addition, the GM model explains why there is a bid-ask spread. Essentially, the spread compensates the dealer for the risk of trading with informed traders.

The spread can be decomposed into components:

a. Adverse Selection Component

The part of the spread due to informed traders. Dealers widen the spread to avoid losses when an informed trader buys (driving price up) or sells (driving price down). Mathematically, if $\boldsymbol{v}$ is the true value, the expected loss from trading with an informed trader is:

$$E[loss] = P(trader\ informed) \times | v - P |$$

This is the main driver of spreads in informed markets.

b. Order Processing / Inventory Component

costs, inventory risk, or compensation for providing liquidity. In GM, the spread mostly arises from information asymmetry, so this component is small or zero if ignoring costs.

We can use a simple example to walk through the intuition. Suppose that the true value $v = 100$ with 50% probability of being high (120) or low (80). The probability that a trader is informed is $\pi = 0.2$. Dealer sets a buy (bid) and sell (ask) price.

Step 1: Trader arrives and buys.

Dealer asks: "Was this trade informed?"

Bayesian update:

$$P(trade\ informed \mid buy) = \frac{\pi P(buy \mid informed)}{\pi P(buy \mid informed) + (1-\pi) P(buy \mid noise)}$$

Step 2: Dealer updates quotes.

If dealer thinks the trader might be informed, dealer raises the ask and lowers the bid slightly.

Step 3: Repeat sequentially.

Each subsequent trade conveys incremental information, narrowing the difference between price and true value.

The outcome will be that bid-ask spread exists even in frictionless markets, purely due to information asymmetry. In addition, sequential trades reveal information gradually, so prices converge to true value over time.

The key insights from the GM model are four folds: 1. Bid-ask spreads are endogenous: they arise naturally in response to adverse selection; 2. Sequential trading is informative: each trade updates the market maker's beliefs; 3. Price improves over time: As more trades occur, the market incorporates private information, and the effective spread (impact of information) declines; and 4. Spread is composed of mostly adverse selection (information asymmetry) in the basic GM model, and inventory or processing costs can be added in extensions.

The GM model carries important behavioral and empirical implications for understanding how markets function in practice. A central prediction is that assets subject to a higher likelihood of informed trading will exhibit wider bid-ask spreads, as market makers must protect themselves against potential losses when trading with better-informed counterparties. Conversely, assets that are highly liquid and characterized by low information asymmetry—where trading is driven primarily by uninformed, liquidity-motivated investors—tend to have narrower spreads, reflecting lower adverse selection costs. This prediction resonates with observed market behavior: large-cap, frequently traded stocks generally have tight spreads, while thinly traded or information-sensitive assets often display much wider spreads. From a behavioral perspective, the model also sheds light on how psychological biases influence trading and liquidity. For example, overconfident investors who overestimate the precision of their private signals may act like informed traders, thereby widening spreads even in the absence of genuine information. Similarly, noise traders motivated by sentiment or herding behavior generate trades that mimic informed activity, forcing dealers to adjust quotes cautiously. A striking illustration of these dynamics can be seen in the trading of “meme stocks” such as GameStop and AMC, where a surge of retail traders—driven more by social media sentiment than by fundamentals—created substantial order flow uncertainty. Market makers responded with significantly wider spreads, reflecting the difficulty of distinguishing between informed speculation and pure noise. By contrast, highly liquid blue-chip stocks such as Apple or Microsoft typically display very narrow spreads because trading is dominated by institutional investors and liquidity traders, with relatively little ambiguity about the flow of information. Empirically, the model underpins tests of order flow informativeness, the degree to which trades convey new information, and the price impact of trades, helping researchers and practitioners disentangle whether liquidity costs stem from genuine private information or from behavioral trading patterns. In this way, the framework not only provides a foundation for analyzing spreads and liquidity but also offers a lens through which to interpret the influence of behavioral factors in shaping market outcomes.

The GM model offers a powerful theoretical framework for understanding how order flow conveys information in markets with asymmetric information. Because the model emphasizes the sequential updating of beliefs by market makers, it naturally lends itself to empirical testing using observable trading variables such as turnover, order imbalance, and volume–price dynamics. In what follows, we discuss how these proxies can be used to test GM predictions, and how extensions such as spread decomposition and the Probability of Informed Trading (PIN) model provide additional tools for measuring the role of informed versus uninformed traders.

Turnover and Liquidity

One straightforward implication of the GM model is that the proportion of informed trading is inversely related to market liquidity. When turnover is high and trading activity is dominated by liquidity-driven investors, the adverse selection risk for dealers is lower, which should translate into narrower bid-ask spreads. Conversely, in thinly traded securities, the chance that a given trade is informed is higher, which pushes spreads wider. This intuition can be tested empirically by examining the relationship between spreads and turnover. A typical specification is:

$$Spread_{i,t} = \alpha + \beta \cdot Turnover_{i,t} + \gamma \cdot Controls_{i,t} + \varepsilon_{i,t}$$

where spread may be measured as the quoted spread, effective spread, or realized spread, and **turnover** is defined as trading volume relative to shares outstanding. The GM model predicts a negative coefficient, $\beta < 0$, consistent with the idea that higher turnover reduces the informational component of spreads. From a behavioral perspective, however, this relationship may break down if high turnover reflects speculative sentiment rather than liquidity trading. Episodes such as the trading surges in “meme stocks” illustrate this possibility: even though turnover was extremely high, market makers widened spreads because they faced difficulty distinguishing sentiment-driven trades from genuinely informed trades.

Order Imbalance and Price Impact

A second line of testing focuses on order imbalance. In the GM framework, informed traders consistently buy when the asset is undervalued and sell when it is overvalued, creating directional pressure in order flow. Thus, large buy–sell imbalances should reveal information about fundamentals and predict future returns.

This can be tested by regressing next-period returns on contemporaneous order imbalance:

$$R_{i,t+1} = \alpha + \beta \cdot Order\ Imbalance_{i,t} + \gamma \cdot Controls_{i,t} + \varepsilon_{i,t+1}$$

A positive coefficient, $\beta > 0$, would support the prediction that buyer-initiated pressure leads to upward price adjustments. In practice, imbalances are often measured using the difference between buyer- and seller-initiated trades, sometimes weighted by volume. The behavioral extension again adds nuance: if imbalances arise from herd behavior or sentiment-driven trading, they may still move prices in the short term, but such effects are less likely to persist, leading to reversals. This distinction between permanent and temporary price impacts provides a way to separate informational order flow from behavioral noise.

Volume–Return Relationship

The GM model also predicts that it is not the sheer quantity of trading that conveys information, but rather the direction of the trade. Buyer-initiated trades signal positive information, while seller-initiated trades signal negative information. This idea can be tested by classifying trades into buys and sells—for example, using the Lee–Ready algorithm—and regressing returns on signed volume:

$$R_{i,t} = \alpha + \beta \cdot Signed\,Volume_{i,t} + \varepsilon_{i,t}$$

If the coefficient β is statistically significant, it provides evidence that order flow is informative, in line with GM. However, behavioral finance again complicates the story: overconfident investors may generate excess volume without true informational content. If their trades move prices initially but the effects reverse later, it indicates that the order flow was driven more by psychology than by superior private information.

Spread Decomposition

While the GM model highlights adverse selection as the dominant reason for the existence of bid-ask spreads, subsequent empirical work has shown that spreads can also reflect inventory risk and order-processing costs. Spread decomposition studies empirically separate these components. The methodology uses trade and quote data to estimate how much of the observed spread is attributable to adverse selection. If the GM framework holds strongly, the adverse selection component should be the largest part of the spread, especially in information-sensitive securities or during periods of heightened information asymmetry such as earnings announcements.

Applications of Glosten-Milgrom’s model

The Probability of Informed Trading (PIN)

Perhaps the most direct empirical operationalization of the GM model is the Probability of Informed Trading **(PIN)**, introduced by Easley et al. (1996). This model explicitly estimates the fraction of trades that are informed versus uninformed using maximum likelihood techniques on sequences of buy and sell orders. PIN has become a widely adopted measure of information asymmetry in empirical finance. Researchers have used it to compare information environments across firms, industries, and countries, and to study how informed trading responds to events such as mergers, earnings announcements, and regulatory changes. The PIN framework thus embodies the central insight of the GM model—that order flow reveals information—and turns it into a practical empirical measure.

Event Studies

Finally, event studies provide another natural testing ground for the GM model. Around major information events—such as earnings announcements or takeover rumors—adverse selection risk increases. Dealers, anticipating a higher probability of informed trading, respond by widening spreads. At the same time, order imbalances around these events become more informative, as private information is actively traded. Measuring changes in spreads, imbalances, and PIN during these periods provides evidence consistent with GM predictions.

Taken together, these empirical approaches confirm the central intuition of the Glosten-Milgrom model: order flow conveys information, and market makers adjust their pricing strategies to account for the possibility of informed trading. Testing can be carried out through regressions of spreads on turnover, price changes on order imbalance, or returns on signed volume. Spread decomposition provides further insight into the share of spreads attributable to adverse selection, while the PIN model offers a direct estimate of informed trading activity. Importantly, behavioral considerations complicate the picture by showing that not all order flow that resembles informed trading is genuinely information-based—overconfidence, herding, and sentiment can generate similar patterns. Distinguishing between lasting price adjustments and temporary mispricing is therefore crucial for empirical work inspired by the GM model.

Here is a short list of classical empirical paper that applies GM model directly summarized in Table 2.2.

Table 2.2 Summary of selected empirical papers on liquidity and informed trading motivated by Glosten-Milgrom

Theme	Citation	Key Insight
PIN model and informed trading	Easley, David, and Maureen O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance, 47*(2), 577–605.	Early PIN-style model estimating the probability of informed trading from buy/sell order flow.
	Easley, David, Maureen O'Hara, and P.S. Srinivas (1998). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *Journal of Finance, 53*(2), 431–465.	Shows informed traders sometimes prefer options markets; relates order flow to price impact.
	Easley, David; Engle, Robert F.; O'Hara, Maureen; Wu, Liuren (2008). Time-Varying Arrival Rates of Informed and Uninformed Trades. *Journal of Financial Econometrics, 6*(2), 171–207.	Extends PIN estimation to allow time-variation; links higher informed-trade probability to wider spreads and higher price impact.

Table 2.2 Summary of selected empirical papers on liquidity and informed trading motivated by Glosten-Milgrom (Cont.)

Theme	Citation	Key Insight
Order imbalance as information proxy	Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2002). Order Imbalance, Liquidity, and Market Returns. *Journal of Financial Economics, 65*(1), 111–130.	Signed order imbalance predicts short-term price changes and affects spreads/liquidity, consistent with GM.
	Chan, L. K., & Lakonishok, J. (1995). The behavior of stock prices around institutional trades. *The Journal of Finance*, *50*(4), 1147-1174.	Institutional order flow moves prices; large imbalances are treated as potentially informed.
Turnover and information asymmetry	Amihud, Yakov, and Haim Mendelson (1986). Asset Pricing and the Bid–Ask Spread. *Journal of Financial Economics, 17*(2), 223–249.	Spreads proxy for transaction costs/information risk; illiquid assets earn higher expected returns.

Table 2.2 Summary of selected empirical papers on liquidity and informed trading motivated by Glosten-Milgrom (Cont.)

Theme	Citation	Key Insight
	Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe (1998). Liquidity and Stock Returns: An Alternative Test. *Journal of Financial Markets, 1*(2), 203–219.	Turnover as liquidity proxy; lower turnover linked to higher expected returns, capturing information-related costs.

2.4 Research ideas using SET intraday data

The Kyle (1985) and Glosten–Milgrom (1985) models provide complementary perspectives on the microstructure of financial markets. While Kyle's model focuses on how private information is gradually incorporated into prices through sequential trading and how order flow impacts price formation, the GM model emphasizes how market makers adjust spreads to compensate for adverse selection risk. Together, these models suggest a rich set of empirical questions that can be explored using the intraday trade and order data available from the Stock Exchange of Thailand (SET).

This section presents sample research questions, outlines empirical strategies, details data preparation and key metrics, and provides methodology blueprints for concrete research examples.

2.4.1 Sample Research Questions

Based on the Kyle and GM models, the following research questions illustrate potential empirical applications using SET data:

1. **Do foreign investors exhibit more informed trading around earnings announcements?**

Informed trading should be more intense during information events such as earnings announcements. Comparing foreign investor order flow to other investor types provides a direct test of informed trading intensity, consistent with both Kyle and GM.

2. **How do spreads and liquidity respond to SET regulatory changes?**

The GM model predicts that adverse selection costs are a key driver of bid-ask spreads. Regulatory changes affecting information transparency or market structure should therefore alter spreads and liquidity.

3. **Do domestic institutions provide liquidity when volatility spikes?**

Kyle's model predicts that liquidity providers adjust activity based on perceived information risk. Investigating institutional trading behavior in high volatility periods reveals whether they act as liquidity providers or withdraw from the market.

4. How does order imbalance predict short-term returns across investor types?

Order imbalances (OI) refers to the difference between number of buy share volume and sell share volume.[4] The GM model links the direction of order flow to adverse selection risk and subsequent price changes. Comparing order imbalance effects across investor categories tests the informativeness of different groups.

2.4.2 Suggested Empirical Strategies

Several strategies can be applied to answer these questions:

- **Difference-in-Differences (DID) Analysis:** Compare event vs. non-event periods or pre- and post-regulatory change periods to isolate causal effects.

---

[4] Order imbalances (OI) often computed from (Buy Volume – Sell Volume)/(Buy Volume + Sell Volume). The ratio frames the possible values between -1 and +1.

- **Intraday Regressions:** Model returns or spreads as functions of contemporaneous order flow, e.g.:

$$R_{i,t} = \alpha + \beta \cdot OI_{i,t} + \gamma X_{it,} + \varepsilon_{i,t}$$

- **Comparative Analysis:** Estimate models separately for different investor types to examine heterogeneity in informed trading and liquidity provision.
- **Event-Time vs. Calendar-Time Analysis:** Examine immediate (event-time) vs. gradual (calendar-time) responses to events or changes.

2.4.3 Data Preparation and Key Metrics

Empirical analysis requires meticulous preparation of SET intraday trade and order data:

- **Intraday Data:** Use transaction prices, volumes, timestamps, and investor type identifiers.
- **Aggregate by Investor Type:** Separate trades for foreign investors, domestic institutions, and retail investors.
- **Construct Metrics:**
  - **Quoted Spread:** Difference between best ask and bid prices.
  - **Effective Spread:** Captures execution cost and price impact.

  - **Depth:** Quantity available at the best bid and ask.
  - **Price Impact:** Change in mid-quote after trades.
  - **Turnover:** Volume relative to shares outstanding.
  - **Order Imbalance:** Signed difference between buyer- and seller-initiated volume relative to total volume.
- **Data Cleaning:** Remove anomalies and synchronize trades and orders to ensure precision.

2.4.4 Methodology Blueprints: Concrete Examples

Below are detailed examples showing how to operationalize these research questions using SET data, linking theory, data, and empirical strategies.

Example 1 — Do Foreign Investors Exhibit More Informed Trading Around Earnings Announcements?

This question examines whether foreign investor trades have greater predictive power for returns during earnings announcements, consistent with Kyle and GM models.

Steps:

1. Define event windows around quarterly earnings announcements.
2. Aggregate intraday foreign investor trades.

3. Compute metrics: order imbalance, effective spread, price impact, turnover.

4. Estimate a difference-in-differences regression:

$$R_{i,t+1} = \alpha + \beta_1 \cdot OI_{i,t}^{foreign} + \beta_2 \cdot Event\ Dummy_{i,t} + \beta_3 \cdot (OI_{i,t}^{foreign} \times Event\ Dummy_{i,t}) + \gamma X_{i,t} + \varepsilon_{i,t+1}$$

5. Interpret β_3 as the additional predictive power of foreign investor imbalance during events.

6. Compare with other investor types for relative informativeness.

Expected result: Higher imbalance predictive power during earnings announcements for foreign investors.

Example 2 — How Do Spreads and Liquidity Respond to SET Regulatory Changes?

This question tests whether regulatory reforms alter adverse selection costs and liquidity.

Steps:

1. Identify a regulatory change and define pre/post periods.
2. Aggregate intraday data for affected stocks.
3. Compute spreads, depth, turnover.
4. Estimate a DiD regression:

$$Spread_{i,t} = \alpha + \beta_1 \cdot PostReg_t + \beta_2 \cdot Treatment_i + \beta_3 \cdot (PostReg_t \times Treatment_i) + \gamma X_{i,t} + \varepsilon_{i,t}$$

5. Interpret β_3 as the impact of the regulatory change on spreads.

Expected result: Narrower spreads and higher depth if regulation reduces information asymmetry.

Example 3 — Do Domestic Institutions Provide Liquidity When Volatility Spikes?

This question examines institutional liquidity provision during volatile periods.

Steps:

1. Define volatility spikes using realized volatility thresholds.
2. Aggregate domestic institution trades intraday.
3. Compute volatility, imbalance, liquidity measures.
4. Estimate:

$$Liquidity_{i,t} = \alpha + \beta \cdot VolatilitySpike_{i,t} + \gamma \cdot OI_{i,t}^{domestic} + \delta X_{i,t} + \varepsilon_{i,t}$$

5. Compare behavior during volatility spikes vs. calm periods.

Expected result: Increased liquidity provision by domestic institutions during volatility spikes.

Example 4 — How Does Order Imbalance Predict Short-Term Returns Across Investor Types?

This question tests whether order imbalance contains predictive information for returns and whether this differs across investor types.

Steps:

1. Compute signed order imbalance by investor type.
2. Estimate:

$$R_{i,t+1} = \alpha + \sum_k \beta_k \cdot OI_{i,t}^k + \gamma X_{i,t} + \varepsilon_{i,t+1}$$

3. Compare β_k across investor types and event windows.

Expected result: Foreign and institutional investors' imbalance will have stronger return predictive power than retail investors.

2.5 Integrating Theory and Empirics

These examples demonstrate how market microstructure models provide a rich framework for understanding the informational content of order flow, the dynamics of spreads, and the role of different investor types in liquidity provision. Applying these models to SET intraday data enables researchers to explore nuanced questions about market microstructure in an emerging-market context, offering insights for both academics and practitioners.

References

Almgren, R., & Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk, 3*(2), 5–39.

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid–ask spread. *Journal of Financial Economics, 17*(2), 223–249.

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.

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics, 108*(4), 905–939.

Chan, L. K., & Lakonishok, J. (1995). The behavior of stock prices around institutional trades. *The Journal of Finance*, *50*(4), 1147-1174.

Chordia, T., & Subrahmanyam, A. (2004). Order imbalance and individual stock returns. *Journal of Financial Economics, 72*(3), 485–518.

Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics, 65*(1), 111–130.

Datar, V. T., Naik, N. Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets, 1*(2), 203–219.

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

Easley, D., Kiefer, N. M., O'Hara, M., & Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance, 51*(4), 1405–1436.

Easley, D., O'Hara, M., & Srinivas, P. S. (1998). Option volume and stock prices: Evidence on where informed traders trade. *Journal of Finance, 53*(2), 431–465.

Easley, D., Engle, R. F., O'Hara, M., & Wu, L. (2008). Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics, 6*(2), 171–207.

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

Griffin, J. M., Harris, J. H., & Topaloglu, S. (2003). The dynamics of institutional and individual trading. *Journal of Finance, 58*(6), 2285–2320.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance, 46*(1), 179–207.

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives, 21*(2), 109–128.

Kaniel, R., Saar, G., & Titman, S. (2008). Individual investor trading and stock returns. *Journal of Finance, 63*(1), 273–310.

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

Lee, C. M., & Swaminathan, B. (2000). Price momentum and trading volume. *the Journal of Finance*, *55*(5), 2017-2069.

Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance, 14*(5), 590–610.

Chapter 3
Investor Behavior Around Key Market Events

3.1 Conceptual Framework

Market events are critical moments when private information, public news, and liquidity conditions interact to influence investor behavior. Understanding how investors respond to such events requires a shift beyond traditional event-study methods that focus solely on abnormal returns. Instead, this chapter emphasizes behavioral patterns in trading activity, drawing on microstructure theory and behavioral finance to examine how information asymmetry shapes responses. To keep discussion brief, we describe the overview and process of conducting market events in bullet points to provide readers with quick reference. For details of event study econometrics, the reader can refer to Kothari and Warner (2007) for a very comprehensive review of concepts from design to interpretation. Bhattacharya (2023) provides a review of event studies using high frequency data.

Two key theoretical perspectives discussed in Chapter 2 guide this analysis. First, the Kyle (1985) model explains how informed and uninformed traders interact over time, with informed traders strategically distributing their trades to minimize detection. Second, the Glosten–Milgrom (1985) model highlights how adverse selection risk drives market-maker behavior, particularly bid-ask spreads, and how spreads reflect the presence of informed trading. Together, these models provide a robust conceptual framework for interpreting investor behavior before, during, and aftermarket events. We provide a general framework to design event studies in this area, then discuss three empirical studies that conduct event analysis. The first case examines how the Covid-19 lock down event alters investor trading behavior. The second compares informed trading price discovery process between underlying stocks and their derivatives. The third, evaluates the effectiveness of trading halts. All cases employ SET's intraday dataset.

3.2 Types of Events

Investor behavior varies across different types of market events. Events can be broadly classified into three categories:

Firm-Specific Events

These events relate to information about a particular firm and often involve significant private and public information release such as earnings announcements, dividend declarations, share buybacks, or trading halts.

Market-Wide Events

These events affect the entire market or economy and often involve macroeconomic news release or regulatory changes. Common macroeconomic announcements that often trigger market-wide responses on sentiment an liquidity are disclosure of GDP, employment data, inflation reports, or monetary policy guidance. Regulatory changes that can affect information asymmetry and liquidity conditions, include alterations in disclosure requirements, tick sizes, or short-selling bans.

Microstructure Events

These are events that arise from the trading process itself rather than exogenous news. For example, large block trades or large transactions that can temporarily move prices and impact liquidity. Another example is limit order book shocks from sudden changes in depth, order cancellations, or additions to the order book that signal shifts in liquidity or trader intentions.

3.3 Event Windows and Analysis

To systematically study investor behavior around events, researchers define event windows comprising pre-event, event, and post-event periods.

Defining Windows

- **Pre-event period:** Captures anticipatory behavior, potentially revealing strategic accumulation or liquidation by informed traders.
- **Event period:** The moment of public information release or market reaction.
- **Post-event period:** Captures the adjustment process, including price discovery, liquidity restoration, and reversal effects.

Addressing Overlapping Events

In real markets, events often cluster or overlap, creating challenges in isolating behavioral responses. Researchers must carefully control for overlapping events through filtering criteria, clustering-adjusted regressions, or event-specific subsamples.

Measuring Intraday Responses

Intraday analysis allows researchers to capture the dynamic nature of behavior during events:

- **Volume:** Trading volume spikes often indicate heightened attention or information asymmetry.
- **Spreads:** Bid-ask spreads widen under uncertainty and informed trading.
- **Volatility:** Realized intraday variance increases during information events.
- **Order imbalance:** Net buying or selling pressure reveals the presence and direction of informed trading.

3.4 Metrics for behavior analysis

Understanding behavior around events requires constructing appropriate microstructure measures related to liquidity since events are new information that can either lead to more trading or reduce trading from fear of picking off risks. We list some common measures below:

Spreads

- **Quoted spread:** The difference between best ask and best bid prices, capturing liquidity cost at a given moment.
- **Effective spread:** The actual cost of executing trades, accounting for price impact.

- **Realized spread:** Measures compensation to liquidity providers after accounting for subsequent price movements.

Depth and Order Imbalance

- **Depth:** Quantity available at the best bid and ask, reflecting liquidity and resilience of the order book.
- **Order imbalance:** The signed difference between buyer-initiated and seller-initiated trades. Large imbalances often signal the presence of informed trading or liquidity provision/withdrawal.

Trading Volume by Investor Type

Segmenting volume by investor category (foreign, domestic institutions, retail) reveals heterogeneity in behavioral responses.

Volatility

Events can cause abrupt changes in demand and supply and thus volatility spikes. Basic measures of volatility are realized variance, usually defined by squared intraday returns and intraday range which captures price range within short intervals.

3.5 Linking Behavior to Market Outcomes

Behavioral patterns around events have direct implications for market quality and efficiency. Empirical studies show that turnover and order imbalance often spike before, during, and after events. Such patterns can be used to infer the presence of informed trading and strategic liquidity provision. Next, different investor categories exhibit distinct behavioral responses. Existing empirical research suggests that foreign investors often display higher informed trading around earnings announcements and macroeconomic events. domestic institutions may provide liquidity during volatility spikes, consistent with liquidity-provision models whereas retail investors tend to trade less strategically but may respond strongly to public news and behavioral biases.

All in all, behavioral responses affect how quickly prices incorporate information, the cost of trading, and the stability of markets. High informed-trading intensity can improve price efficiency but increase spreads. Liquidity provision during events supports market stability but may be withdrawn under extreme uncertainty. Finally, regulatory changes that alter information asymmetry have measurable effects on spreads, depth, and turnover.

3.6 Concrete Empirical Research Designs

This section provides practical blueprints for designing empirical studies on investor behavior around key market events. Each example follows a consistent structure: defining the event, data preparation, constructing behavioral metrics, specifying empirical models, and interpreting results.

3.6.1 Firm-Specific Event Example: Earnings Announcements

Research Question:

Do foreign investors exhibit more informed trading around earnings announcements?

Steps:

1. Define Event Window:
   - Use quarterly earnings announcement dates for SET-listed firms.
   - Define pre-event (e.g., −3 days), event day, and post-event (+3 days) windows.
2. Data Aggregation:
   - Extract intraday trades for foreign investors.
   - Aggregate into fixed intervals (e.g., 5-minute bins).

3. Construct Key Metrics:

   - Order imbalance for foreign investors:

$$OI_{i,t}^{foreign} = \frac{Buy\ Volume_{i,t}^{foreign} - Sell\ Volume_{i,t}^{foreign}}{Total\ Volume_{i,t}}$$

   - Effective spread:

$$Eff\ Spread_{i,t} = 2 \times \frac{|\ P_{i,t} - M_{i,t}\ |}{M_{i,t}}$$

   - Price impact and turnover.

4. Empirical Model:

   A difference-in-differences regression:

$$R_{i,t+1} = \alpha + \beta_1 \cdot OI_{i,t}^{foreign} + \beta_2 \cdot Event\ Dummy_{i,t} + \beta_3 \cdot (OI_{i,t}^{foreign} \times Event\ Dummy_{i,t}) + \gamma X_{i,t} + \varepsilon_{i,t+1}$$

5. Interpretation:

A significantly positive β_3 indicates higher predictive power of foreign investor imbalance during earnings announcements, consistent with informed trading.

3.6.2 Market-Wide Event Example: Regulatory Changes

Research Question:

How do spreads and liquidity respond to regulatory changes?

Steps:

1. Identify Regulatory Change:
   - E.g., disclosure requirement changes, tick-size adjustments.
   - Define pre/post regulatory windows (e.g., 30 days before and after).
2. Data Aggregation:
   - Select affected firms and gather intraday order book data.
3. Construct Key Metrics:
   - Quoted and effective spreads.
   - Market depth and turnover.
4. Empirical Model:

   Difference-in-differences regression:

$$Spread_{i,t} = \alpha + \beta_1 \cdot PostReg_t + \beta_2 \cdot Treatment_i + \beta_3 \cdot (PostReg_t \times Treatment_i) + \gamma X_{i,t} + \varepsilon_{i,t}$$

5. Interpretation:

A significant negative β_3 suggests spreads narrow after regulation, consistent with reduced adverse selection risk.

3.6.3 Microstructure Event Example: Volatility Spikes

Research Question:

Do domestic institutions provide liquidity when volatility spikes?

Steps:

1. Define Volatility Spike:
   - Use intraday realized variance to identify high volatility periods (e.g., >90th percentile).
2. Data Aggregation:
   - Extract intraday trades by domestic institutions.
3. Construct Key Metrics:
   - Order imbalance for domestic institutions:

$$OI_{i,t}^{domestic} = \frac{Buy\,Volume_{i,t}^{domestic} - Sell\,Volume_{i,t}^{domestic}}{Total\,Volume_{i,t}}$$

- Quoted spread, depth, turnover.

4. Empirical Model:

$$Liquidity_{i,t} = \alpha + \beta \cdot VolatilitySpike_{i,t} + \gamma \cdot OI_{i,t}^{domestic} + \delta X_{i,t} + \varepsilon_{i,t}$$

5. Interpretation:

A positive γ suggests domestic institutions step in as liquidity providers during volatility spikes.

3.6.4 Investor-Type Behavior Example: Order Imbalance Predictability

Research Question:

How does order imbalance predict short-term returns across investor types?

Steps:

1. Construct Order Imbalance for Each Investor Type:

$$OI_{i,t}^{k} = \frac{Buy\ Volume_{i,t}^{k} - Sell\ Volume_{i,t}^{k}}{Total\ Volume_{i,t}}$$

2. Empirical Model:

$$R_{i,t+1} = \alpha + \sum_{k} \beta_k \cdot OI_{i,t}^{k} + \gamma X_{i,t} + \varepsilon_{i,t+1}$$

Where k indexes investor types (foreign, domestic institutions, retail).

3. Interpretation:

Differences in β_k indicate which investor type's order imbalance contains greater information content for predicting returns.

Comparative and Robustness Analyses

Across all examples, comparative and robustness checks are crucial:

- **Comparative Analysis:** Examine differences across investor types, event windows, or market conditions.
- **Intraday vs. Calendar-Time Analysis:** Compare high-frequency responses to longer-term patterns.
- **Robustness Checks:** Use alternative definitions of order imbalance, different event window lengths, and controls for volatility, turnover, and spread.

Implications of Empirical Findings

These empirical designs, grounded in the Kyle and Glosten–Milgrom models, allow researchers to link observed trading behavior to theoretical predictions about information asymmetry, liquidity provision, and price discovery. The results have direct implications for market efficiency, regulatory policies, and trading strategy.

By combining conceptual frameworks with event-based empirical strategies, this chapter provides a comprehensive toolkit for analyzing investor behavior around key market events. Moving beyond abnormal returns, the emphasis on microstructure metrics

— spreads, depth, imbalance, and turnover — allows a richer understanding of how information asymmetry shapes market dynamics. These insights are not only valuable for academics but also for regulators and market participants seeking to improve market quality and efficiency.

3.7 Case Study: Gambling for Resurrection — Thai Retail Investors During the COVID-19 Lockdown (Adapted from Do retail investors gamble more during lockdown? International Review of Finance, 2024, Pantisa Pavabutr and Bin Zhao)

Background

In early 2020, the COVID-19 pandemic triggered a global market collapse. In Thailand, the Stock Exchange of Thailand (SET) fell sharply, and the government imposed a strict lockdown from March to May 2020. For most households, this was a sudden negative wealth shock — both portfolio values and income expectations declined. According to Prospect Theory (Kahneman & Tversky, 1979), individuals evaluate outcomes relative to a reference point, such as recent wealth or expected standard of living. When actual wealth falls below the reference point, the value function becomes convex in losses, making people risk-seeking. This case examines whether Thai retail investors responded to lockdown-related losses by trading more actively and favoring lottery-type stocks — small, cheap, volatile stocks that offer a small chance of large payoffs.

Data and Institutional Setting

The study uses intraday transaction data from September 2019 – June 2020, containing: Timestamps, prices, and volumes for each trade, stock identifiers (tickers), buyer/seller investor type (Foreign, retail, and local institutions). This structure allows observation of real trading behavior, not just aggregate flows. The lockdown (March – May 2020) serves as a natural experiment contrasting behavior before, during, and after the shock.

Behavioral Framework: Prospect Theory and Reference Dependence

Under prospect theory, investors derive value from gains and losses relative to a reference point $\boldsymbol{R}$:

$$v(x) = \begin{cases} x^{\alpha}, & x \geq 0, \\ -\lambda(-x)^{\beta}, & x < 0, \end{cases} \quad where\ \lambda > 1$$

- **Concave for gains** ⟶ risk aversion.
- **Convex for losses** ⟶ risk-seeking ("gambling for resurrection").
- **Loss aversion** parameter $\lambda \approx 2\text{–}3$ in empirical studies.

During the lockdown, investors likely perceived themselves below reference wealth R, which should increase preference for positively skewed, high-volatility (“**lottery-like**”) assets.

Identifying Lottery-Type Stocks

Following Kumar (2009) and Bali et al. (2011), a stock qualifies as lottery-type if it has low price (bottom quartile of price distribution), high idiosyncratic volatility (IVOL), and High idiosyncratic skewness (ISKEW).

Estimation:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{m,i}(R_{m,t} - R_{f,t}) + \beta_{s,i}SMB_t + \beta_{h,i}HML_t + \varepsilon_{i,t}$$

$$IVOL_i = \sigma(\varepsilon_{i,t}), \qquad ISKEW_i = \frac{E[(\varepsilon_{i,t} - \bar{\varepsilon}_i)^3]}{(Var(\varepsilon_{i,t}))^{3/2}}$$

Stocks in the top quartile of $IVOL$ and $ISKEW$ and bottom quartile of price are labeled as *lottery stocks*.

Key Measures of Investor Behavior

Variable	Definition	Behavioral Interpretation
Turnover Ratio $Turnover_t^g$	Daily trading value by investor group *g* ÷ portfolio market value	High values ⟶ overconfidence or sensation-seeking

Variable	Definition	Behavioral Interpretation
Lottery Trading Share $LotteryTrade_t^g$	Fraction of trading volume in lottery stocks	Preference for skewed payoffs
Portfolio Return Ret_t^g	Daily/weekly weighted average return of group *g*'s holdings	Performance outcome of behavior
Lockdown Dummy $Lockdown_t$	1 = Mar–May 2020; 0 otherwise	Captures shock period

Empirical Design

Model 1 – Preference for Lottery Stocks

$$LotteryTrade_t^g = \alpha + \beta_1 Lockdown_t + \beta_2 Retail_g + \beta_3 (Lockdown_t \times Retail_g) + \gamma' X_t + \mu_t + \varepsilon_t$$

- X_t : control variables (market volatility, market return, trading day dummies)
- $\beta_3 > 0$ expected if retail investors gamble more during lockdown.

Model 2 – Trading Intensity

$$Turnover_t^g = \alpha + \delta_1 Lockdown_t + \delta_2 Retail_g + \delta_3 (Lockdown_t \times Retail_g) + \theta' Controls_t + \varepsilon_t$$

- $\delta_3 > 0$: retail investors increase activity.

Model 3 – Portfolio Performance

$$Ret_t^g = \alpha + \phi_1 Lockdown_t + \phi_2 Retail_g + \phi_3 (Lockdown_t \times Retail_g) + \xi' Z_t + \varepsilon_t$$

- $\phi_3 < 0$: deteriorating performance due to excessive risk-taking.

Empirical Findings

Indicator	Pre-Lockdown	Lockdown	Change	Interpretation
Retail Turnover (per month)	12 ×	29 ×	↑ 140 %	Overtrading surge
Lottery Share of Retail Trades	~ 18 %	~ 45 %	↑ 27 %	Gambling preference
Retail Portfolio Return	–15 %	–90 %	↓ significant	Poor outcomes
Institutional Turnover	1 ×	1 ×	—	Stable behavior
Institutional Portfolio Return	–10 %	–30 %	Moderate decline	More disciplined

Interpretation

Retail investors displayed classic *reference-dependent risk-seeking*: facing losses, they shifted toward high-skew, low-price stocks. Their turnover soared, while portfolio returns collapsed — confirming that emotional and cognitive biases can destroy wealth during crises.

Behavioral and Policy Implications

1. **Reference Dependence in Practice** – Investor risk-taking is shaped by perceived changes in wealth, not absolute wealth.
2. **Sensation-Seeking Substitution** – Lockdown limited entertainment options, increasing speculative trading ("boredom trading").
3. **Overconfidence Amplification** – Digital trading platforms enabled rapid turnover, reinforcing self-attribution bias.
4. **Investor Protection** – Regulators could emphasize behavioral risk disclosures or cooling-off periods during volatile times.

Methodological Insights for Researchers

The use of micro-level transaction data from the Stock Exchange of Thailand (SET), which includes investor-type identifiers, offers researchers a powerful framework for testing behavioral finance theories directly in trading behavior. Such detailed data allow for the observation of how different investor groups—retail, institutional, and foreign—respond to shifts in market conditions, providing clear evidence of psychological mechanisms such as risk-seeking, loss aversion, and lottery preference. The methodological approach used in this case study can be extended to a wide range of other settings, including political or economic shocks, major policy announcements, and sentiment-driven phenomena such as meme stock episodes. To replicate or adapt this methodology, researchers should focus on three key components: first, identifying an exogenous shock that plausibly alters investors' perceived wealth or reference point; second, defining a proxy for gambling or skewness preference—such as portfolios tilted toward lottery-type stocks, high idiosyncratic volatility, or extreme return distributions; and third, comparing behavior either across investor groups or between pre- and post-event periods to isolate the behavioral response. This approach provides a robust empirical design that bridges psychological theory and market microstructure evidence, allowing researchers to study how investor sentiment and reference-dependent preferences manifest in actual trading data.

3.8 Case Study: Determinants of Informed Traders' Market Choice (Adapted from Factors determining where informed traders trade, 2002, Thammasat University. Faculty of Commerce and Accountancy. Business Research Center, by Arnat Leemakdej)

Background

Understanding where informed traders choose to trade — whether in the underlying stock or its derivative market — is central to market microstructure and asset pricing research. This question matters because the migration of informed traders can influence market liquidity, volatility, and price discovery. The study by Leemakdej et al. (2008) investigates these dynamics in the context of the Thai market following the introduction of stock warrants. Warrants are derivative instruments giving the holder the right to purchase stock at a set price before expiry. Their introduction offers new opportunities for leverage and risk management, potentially altering trading behavior.

Research Question

Why do informed traders choose to trade in one market rather than another? Specifically, what factors drive informed traders to migrate between the underlying stock and its warrant?

Conceptual Framework

The paper builds on two strands of theory:

1. **Asymmetric Information Hypotheses** — These include (i) migration of informed traders to derivative markets to exploit leverage, (ii) migration of uninformed traders away from stock markets to avoid adverse selection, and (iii) emergence of new traders drawn by derivatives.

2. **Asymmetric Risk Preference Hypotheses** — These posit that derivatives enable risk sharing, potentially reducing spot market volatility.

The "migration of informed traders" hypothesis predicts that informed traders will shift to derivative markets offering higher leverage and lower transaction costs, thereby reducing volatility in the underlying stock market.

Methodology

The authors develop a structural market microstructure model using intraday transaction data from the Stock Exchange of

Thailand (SET) in 1996 for 29 stock–warrant pairs. The model distinguishes between:

- **Informed traders** — traders with superior knowledge of firm-specific news.
- **Uninformed traders** — traders without private information.

Key model variables:

- Relative price of stock to warrant (leverage incentive).
- Market liquidity (measured by matched orders ratio between stock and warrant).
- Bid–ask spreads.
- Probability of news and bad news events.
- Stock and warrant volatility (risk measures).

The authors estimate the proportion of informed traders in each market and use regression analysis to identify the determinants of their trading location.

Hypotheses Tested

1. **Leverage opportunity:** Informed traders prefer warrants because lower prices and leverage allow bigger gains with less capital. Therefore, more informed traders will trade warrants if leverage opportunity is high.

2. **Liquidity:** Informed traders need sufficient trading volume to hide their trades and thus informed traders would migrate to more liquid markets.

3. **Information concealment:** Informed traders prefer markets with more uninformed traders to disguise their trades. The presence of more uninformed traders attracts informed traders.

4. **Risk sharing:** Derivatives might allow informed traders to hedge risk. If warrants help to reduce risk in the underlying stock market, then trading will migrate for hedging purposes.

Key Findings

- **Migration of informed traders:** Informed traders strongly prefer the warrant market once warrants are listed, confirming the migration hypothesis. On average, nearly all informed traders shift to trading warrants, leaving the underlying stock market dominated by uninformed traders.
- **Liquidity matters:** The proportion of informed traders in the stock market increases with liquidity, but only to a point. When markets are thin, informed traders avoid the stock market to hide their identity and avoid detection.
- **Leverage incentive:** Contrary to theory, relative price differences between stocks and warrants are not significant

determinants of migration, rejecting the pure leverage-motivation hypothesis.

- **Risk sharing:** There is no evidence that informed traders migrate for hedging purposes.

- **Uninformed traders:** These traders tend to trade more in the stock market during news events, particularly when bad news is likely, suggesting a speculative behavior rather than informed trading.

- **Volatility preference:** Both informed and uninformed traders show a preference for trading in markets with higher volatility, challenging some risk-sharing theories.

In sum, the study provides a rare empirical test of the migration-of-informed-traders hypothesis within a detailed market microstructure framework. It highlights the critical role of liquidity and asymmetric information in shaping trading location choices, offering valuable insights for academics, market regulators, and practitioners designing derivative markets.

Implications of findings

The findings suggest that the migration of informed traders is largely driven by market microstructure factors such as liquidity and the ability to conceal trading intentions rather than simple leverage opportunities or hedging motives. This migration reshapes the composition of trading activity, influencing volatility and price discovery in both the underlying stock and warrant markets. The methodological framework can also be adapted to other contexts where derivative markets exist, such as options, futures, or ETFs, and can be extended to investigate other information triggers. For example, how political or macroeconomic shocks affect informed traders' migration, how different derivative structures change trading incentives, or the impact of market regulations such as short-sale restrictions.

3.9 Case Study "The Effectiveness of Trading Halts and Investor Trading Performance", adapted from "The Effectiveness of Trading Halts and Investor Trading Performance: An Intraday Analysis on the Stock Exchange of Thailand. International Research Journal of Finance and Economics, (102), 191-209. By Nareerat Taechapiroontong, Charlie Charoenwong, Chiraphol N. Chiyachantana and Radchda Lurang.

Background

Trading halts are temporary suspensions of trading in a security and are implemented in many markets with the stated goals of ensuring orderly trading, maintaining price stability, and allowing investors sufficient time to process material information. In practice, halts can be triggered by several conditions: significant corporate events or announcements, unusual price fluctuations suggesting information asymmetry, requests from issuers to clarify major developments, or extraordinary events that might affect trading systems. The purpose of halts, in theory, is to reduce volatility and facilitate efficient price discovery. However, there is considerable debate about their efficacy. Proponents argue that halts help markets absorb and incorporate information without disorderly price movements, while critics contend that halts delay price adjustments, impede price discovery, increase costs for investors, and can even lead to higher post-halt volatility. There is also concern that institutional investors, who can process information more quickly than retail traders, might gain an advantage during halts, thereby increasing inequality in trading outcomes.

Research Objectives

This paper investigates the role and effectiveness of trading halts in promoting market efficiency in an emerging market context, specifically the Stock Exchange of Thailand (SET). The authors examine whether trading halts improve the price discovery process, control volatility, and influence trading behavior, as well as how different characteristics of halts — including their timing, duration, and the size of the affected firm — affect these outcomes. They also examine the performance of different investor types, including retail domestic investors, institutional domestic investors, and foreign investors, in the context of trading halts. The study is particularly important because it addresses a gap in the literature: most prior studies have examined intra-day halts only, whereas this paper considers both delayed opening halts and intra-day halts, and analyses a broad range of halt characteristics on an order-driven exchange.

Hypotheses Tested

The authors mainly tested if halts allow new information to be incorporated quickly into price without prolonging volatility and distorting trading behavior. To do so, they classify halts according to whether they involved "good news," "bad news," or "no news" based on price changes immediately after resumption of trading. They also categorize halts as delayed openings or intra-day, and group them by duration (shorter than one hour versus between one and two hours) and by the size of the firm affected. The data are combined with intraday trade-by-trade records that include information on price, volume, and investor type.

Data and Methodology

The paper uses SETSMART database to collect halt announcements, times, and their causes. Trade-by-trade information from January 1999 to December 2007. The authors collected a total of 882 halts. After filtering out multiple halts, samples with insufficient liquidity, and trading prices below THB 1, 228 halts covering 150 firms are used in the study. Investors are classified into retail, foreign, and local institutions. To test the hypothesis the authors compare abnormal returns, volatility, and trading activity in event periods (around halts) to non-halt periods.

Key empirical findings

The findings show that halts generally facilitate a rapid price discovery process, particularly for "good news" halts. On average, abnormal returns spike immediately after halts, with the first 15 minutes showing the most significant price adjustments, before reverting to normal within approximately 30 minutes. For bad news halts, price adjustments take longer — around 75 minutes — and there is evidence of pre-halt price increases, suggesting that some information leakage or anticipatory trading occurs before halts. No-news halts produce relatively minor price changes, consistent with the absence of significant information. The timing and duration of halts also matter: delayed opening halts produce stronger abnormal returns than intra-day halts, while shorter halts are associated with quicker price adjustments. Larger firms tend to experience stronger post-halt abnormal returns. Overall, these results support the conclusion that halts are effective in improving price discovery, though they also indicate that pre-halt information leakage is a concern

Regarding volatility, the study finds a substantial increase immediately after halts, with abnormal volatility rising by as much as 720% in high-low price range measures and 480% in absolute return measures. However, this heightened volatility is short-lived, typically returning to normal within an hour, indicating that halts succeed in controlling volatility rather than spreading it over longer periods. Good news halts produce higher and more prolonged volatility than bad news halts, and delayed opening halts generate greater post-halt volatility than intra-day halts. Longer halts — those exceeding two hours — are particularly associated with greater volatility, a finding with important policy implications, since it suggests that prolonged suspensions may inadvertently increase price uncertainty rather than contain it. Firm size also matters: medium-sized stocks exhibit the highest volatility after halts, although volatility normalizes relatively quickly.

Trading activity around halts is characterized by pronounced surges before and after the halt. Trading volume and the number of trades increase markedly prior to halts, spike immediately after trading resumes, and then decay over a period of about three days. This pattern suggests that halts trigger concentrated trading as investors adjust their portfolios and respond to new information. The surge in activity is most pronounced for good news halts, and it is particularly evident in delayed opening halts, halts involving large firms, and halts of longer duration. The authors' regression analysis confirms that abnormal trading volume is a strong predictor of abnormal volatility, and that longer halts are associated with higher volatility. Halt timing and firm size have a weaker but still noticeable effect.

Investor Trading Performance

A particularly novel part of the study examines the performance of different investor groups around halts. Retail domestic investors are the dominant group, accounting for more than 85% of trading volume and value during halt events. They tend to trade intensively before halts and engage in a contrarian strategy, buying low before halts and selling high afterward. Institutional investors focus more on larger stocks, trade earlier than retail investors, and show more moderate price advantages. Foreign investors concentrate on large-cap stocks but trade at higher prices and tend to perform worse than domestic investors around halts. This finding suggests that domestic investors have an informational advantage, possibly due to greater familiarity with the local market or more timely access to information before halts. Retail investors, in particular, appear to exploit this advantage most effectively, especially in the case of good news halts.

Policy Implications

The research finds halts improve market efficiency by aiding price discovery and limiting excessive volatility. Furthermore, regulators should monitor halt durations (shorter halts limit volatility better), recognize that domestic investors gain more from halts due to information asymmetry, and ensure transparency to minimize insider trading or leakage before halts. The paper concludes that trading halts on the SET generally improve market efficiency by facilitating rapid price adjustments and reducing information asymmetry. They stabilize prices and volatility in the short term while enabling investors to adjust to new developments. However, the benefits depend on the nature of the halt: shorter halts tend to produce better outcomes in terms of volatility control, while longer halts can increase uncertainty. The findings also reveal that domestic investors, particularly retail investors, are able to trade more profitably around halts than foreign investors, highlighting the potential for information asymmetry. Policymakers should therefore carefully consider halt duration and transparency mechanisms to maximize the benefits of trading halts while minimizing unintended consequences.

References

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics, 99*(2), 427–446.

Bhattacharya, R. R. (2023). *Event studies with intraday data* (Working paper). SSRN.

Leemakdej, A. (2002). *Factors determining where informed traders trade* (Working paper). SSRN

Leemakdej, A., Khanthavit, A., & Perrakis, P. (2008). Warrant pricing and listing in an emerging market: The case of Thailand. (Working paper). SSRN.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Kothari, S.P., and Warner, J., 2007, Econometrics of Event Studies, In: Eckbo, E. (Ed.), Handbook of Empirical Corporate Finance, Amsterdam, The Netherlands: Elsevier/North-Holland.

Kumar, A. (2009). Who gambles in the stock market? *Journal of Finance, 64*(4), 1889–1933.

Pavabutr, P., & Zhao, B. (2024). Do retail investors gamble more during lockdown? *International Review of Finance,* 24(4), 572–603

Taechapiroontong, N., Charoenwong, C., Chiyachantana, C. N., & Lurang, R. (2012). The effectiveness of trading halts and investor trading performance: An intraday analysis on the Stock Exchange of Thailand. *International Research Journal of Finance and Economics, 102*, 191–209.

Chapter 4

Advanced Volatility Modelling with Market Microstructure

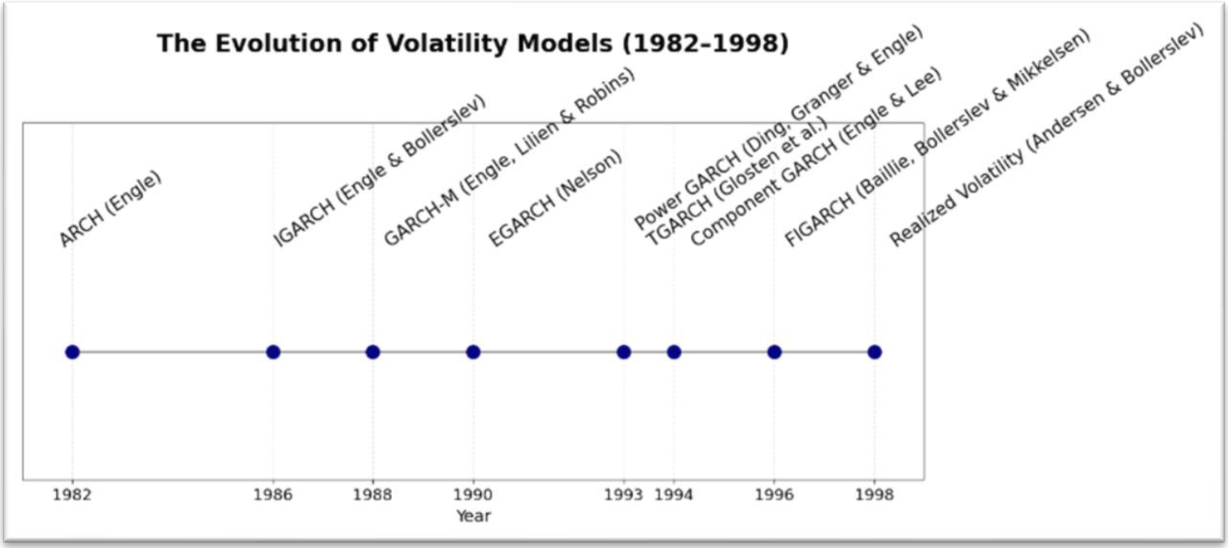
4.1 Evolution of Volatility Models

Modeling and forecasting of stock market return conditional variance are of great importance to both academics and practitioners. Volatility is a key variable in numerous financial applications, including portfolio management, option pricing, and the assessment of value at risk (VaR) for financial institutions, *inter alia*. However, a fundamental challenge arises from the fact that the volatility process is difficult to define and fickle to measure. From finance literature, the operating definition of volatility is dispersed depending on average periods and data frequencies. Visual eyeballing also reveals that volatility exhibits clustering, larger jumps around crises, and fat tails.

Consequently, a vast body of research has emerged proposing econometric models to capture the salient features of volatility. Since the seminal work of Engle (1982), substantial progress has been made in this field as depicted in Figure 4.1. Nevertheless, most of these models exhibit a high degree of intertemporal volatility persistence (Bollerslev et al., 1988; and Ghysels et al., 1996). One notable limitation of standard volatility models is their inability to adequately explain the variability in ex-post squared returns.

Andersen and Bollerslev (1998) address this issue by examining how the choice of a proxy for the latent population measure of volatility influences the evaluation of volatility forecasting models. They demonstrate that when squared daily returns are used as a proxy for daily variance, GARCH models exhibit poor forecasting performance. In contrast, when employing the sum of intraday squared returns—an approach that utilizes more information—the forecasting accuracy of GARCH models improves markedly.

Figure 4.1: The Evolution of Volatility Models during the year 1982-1998



Source: Authors' own illustration

Table 4.1 Summary of the volatility models during the period 1982-1998

No.	Model	Key Contributors	Year	Form of Model
1	ARCH	Engle	1982	$\sigma_t^2 = \alpha_0 + \Sigma_i \alpha_i \varepsilon_{t-i}^2$
2	GARCH	Bollerslev	1986	$\sigma_t^2 = \alpha_0 + \Sigma_i \alpha_i \varepsilon_{t-i}^2 + \Sigma_j \beta_j \sigma_{t-j}^2$
3	IGARCH	Engle & Bollerslev	1986	$\alpha + \beta = 1$ (unit-root like persistence)
4	GARCH-M	Engle, Lilien & Robins	1987	$r_t = \mu + \lambda\sigma_t^2 + \varepsilon_t$
5	Multivariate GARCH (VECH/BEKK)	Bollerslev, Engle & Wooldridge	1988	$H_t = C'C + A'\varepsilon_{t-1}\varepsilon_{t-1}'A + B'H_{t-1}B$
6	EGARCH	Nelson	1991	$\ln(\sigma_t^2) = \omega + \Sigma_i[\alpha_i(\|\varepsilon_{t-i}\|/\sigma_{t-i}) + \gamma_i(\varepsilon_{t-i}/\sigma_{t-i})] + \beta_i \ln(\sigma_{t-i}^2)$
7	TGARCH / GJR-GARCH	Glosten, Jagannathan & Runkle	1993	$\sigma_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
8	Component GARCH	Engle & Lee	1993	$\sigma_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1})$

Table 4.1 Summary of the volatility models during the period 1982-1998 (Cont.)

No.	Model	Key Contributors	Year	Form of Model
9	Power GARCH (PGARCH)	Ding, Granger & Engle	1993	$\sigma_t^{\delta} = \alpha_0 + \Sigma_i \alpha_i(\lvert\varepsilon_{t-i}\rvert - \gamma_i\varepsilon_{t-i})^{\delta} + \Sigma_j \beta_j\sigma_{t-j}^{\delta}$
10	FIGARCH	Baillie, Bollerslev & Mikkelsen	1996	$(1-\beta L)(1-L)^{d}\,\varepsilon_t^{2} = \omega + [1-\alpha L-\beta L](\varepsilon_t^{2}-\sigma_t^{2})$
11	Realized Volatility Framework	Andersen & Bollerslev	1998	$RV_t = \Sigma_i\, r_{t,i}^{2}$ (sum of intraday squared returns)

Source: Authors' own illustration

The use of squared returns as a measure of volatility has long been recognized as problematic, since it provides an extremely noisy estimate of the *ex post* volatility of financial assets. This limitation has motivated researchers to search for more accurate and informative measures of volatility. With the growing availability of high-frequency or intraday financial data, it has become possible to observe price movements in much finer detail and to incorporate this information into empirical analysis. This development has had a profound impact on research in both applied econometrics and financial market microstructure.

In the applied econometrics literature, studies that attempt to describe the characteristics of price processes—particularly the behavior of volatility within a trading day—are generally referred to as *high-frequency* models. These models exploit the rich information contained in intraday returns to provide a more precise depiction of the volatility process.

In this chapter, we aim to introduce a family of extended ARCH-type models that incorporate intraday information directly into the volatility framework. By combining traditional time-series modeling with high-frequency data, these models help us understand short-term market dynamics more clearly and enhance the accuracy of volatility forecasting. The data we used to illustrate the model, comes from the Stock Exchange of Thailand (SET) and has been sampled at 30- minute frequency. The use of a relatively lower frequency than using complete order book record is to reduce microstructure noise.[5]

A major breakthrough in this line of research came with the introduction of the realized volatility framework. The central idea is that by aggregating high-frequency intraday returns, one can construct a nonparametric measure that closely approximates the true underlying volatility over a given period. This approach, pioneered by Andersen and Bollerslev (1998), provided an empirical bridge between the observable data and the latent volatility process that traditional ARCH or GARCH models could only infer indirectly.

The realized volatility measure not only offers a more accurate proxy for latent volatility but also enables researchers to examine important features of financial markets—such as volatility clustering, persistence, and the impact of macroeconomic announcements.

---

[5] See Hansen and Lunde (2006) for more a technical discussion of this issue.

4.2 Heterogeneous Autoregressive Model for Realized Volatility (HAR–RV Model)

Building upon the realized volatility framework, Corsi (2009) proposed the *Heterogeneous Autoregressive (HAR) model* as a simple yet powerful approach to capture the long-memory property of realized volatility. The HAR model is motivated by the idea that market participants operate over different investment horizons—such as daily, weekly, and monthly—which collectively shape the dynamics of observed volatility. By modeling realized volatility as a linear combination of its own past values over these heterogeneous time scales, the HAR model can reproduce the slowly decaying autocorrelation typically observed in volatility series.

One of the main advantages of the HAR model is its intuitive structure and ease of estimation, which make it particularly suitable for both academic research and practical forecasting applications. Despite its simplicity, the model performs remarkably well in explaining and predicting the persistence of volatility, often rivaling more complex specifications. As a result, the HAR framework has become a standard benchmark in the literature on volatility modeling and forecasting, and it serves as the foundation for numerous extensions that incorporate additional features such as jumps, asymmetries, and realized measures of covariance.

In Corsi’s (2009) work, known as the “Boston Group” model, the HAR–RV is recommended to group volatility horizons into daily, weekly, and monthly intervals because different types of market participants (e.g., retail traders, fund managers, institutional investors) operate on distinct investment horizons.

To understand the structure of the HAR–RV model, it is designed to capture and forecast realized volatility by grouping market participants into short-, medium-, and long-term horizons as follows:

Daily Horizon: Represents retail traders who adjust portfolios and manage risk on a day-to-day basis.

$$RV^D = RV_{t-1}$$

Weekly Horizon: Reflects the views of mid-term institutional fund managers who monitor returns and volatility on a weekly basis.

$$RV^w = \frac{1}{5}\sum_{j=1}^{5} RV_{t-j}$$

Monthly Horizon: Represents long-term institutional investors who focus on macro trends and plan investments on a monthly basis.

$$RV^{M} = \frac{1}{22}\sum_{j=1}^{22} RV_{t-j}$$

The HAR–RV model can be expressed as follows:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_t^{(w)} + \beta_3 RV_t^{(m)} + \varepsilon_{t+1}, \varepsilon_{t+1} \sim i.i.d.$$

In general, the HAR–RV model exhibits good forecasting performance even when estimated using a simple OLS approach. This horizon decomposition helps the model capture the long-memory effect of volatility and provides useful interpretation of how different market participants contribute to aggregate market volatility.

Example 4.1 Application to SET Intraday Data Using Python

The following example demonstrates the application of intraday data from the Stock Exchange of Thailand (SET) to estimate the HAR–RV model for the ticker "KBANK." during 2018–2020.

Step 1: From the data frame *df* in Example 4.2, we can construct the HAR–RV model using the following command structure.

```
1. # Filter for KBANK and sort
2. df_k = df[df['Ticker'] == 'KBANK'].copy()
3. df_k.sort_values(['date', 'Unnamed: 0'], inplace=True)
```

```python

# Compute intraday log returns
df_k['log_price'] = np.log(df_k['medprice'])
df_k['return'] = df_k.groupby('date')['log_price'].diff()

# Compute daily realized volatility
rv = df_k.groupby('date')['return'].apply(lambda x: np.sum(x**2)).to_frame(name='RV')

# Construct HAR variables
rv['lag_D'] = rv['RV'].shift(1)
rv['lag_W'] = rv['RV'].rolling(window=5).mean().shift(1)
rv['lag_M'] = rv['RV'].rolling(window=22).mean().shift(1)

# Drop missing values
rv_clean = rv.dropna()

# Prepare regression
y = rv_clean['RV']
X = rv_clean[['lag_D', 'lag_W', 'lag_M']]
X = sm.add_constant(X)

# Fit HAR-RV model
```

```
model = sm.OLS(y, X).fit()

# Predict fitted values
rv_clean['Fitted'] = model.predict(X)

# Plot actual vs fitted
plt.figure()
plt.plot(rv_clean.index, rv_clean['RV'], label='Actual RV')
plt.plot(rv_clean.index, rv_clean['Fitted'], label='Estimated RV')
plt.title('HAR-RV: KBANK Daily Realized Volatility')
plt.xlabel('Date')
plt.ylabel('Realized Volatility')
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Table 4.2 Estimated results of the HAR–RV model for the ticker "KBANK."

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     RV   R-squared:                       0.135
Model:                            OLS   Adj. R-squared:                  0.131
Method:                 Least Squares   F-statistic:                     36.64
Date:                Thu, 26 Jun 2025   Prob (F-statistic):           5.15e-22
Time:                        22:18:04   Log-Likelihood:                 4551.6
No. Observations:                 710   AIC:                            -9095.
Df Residuals:                     706   BIC:                            -9077.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       5.327e-05   2.13e-05      2.501      0.013    1.15e-05    9.51e-05
lag_D         -0.0151      0.044     -0.344      0.731      -0.101       0.071
lag_W          0.2804      0.102      2.739      0.006       0.079       0.481
lag_M          0.4871      0.114      4.261      0.000       0.263       0.712
==============================================================================
Omnibus:                     1446.029   Durbin-Watson:                   2.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2972711.353
Skew:                          15.154   Prob(JB):                         0.00
Kurtosis:                     318.543   Cond. No.                     9.48e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 4.2 Comparison between the actual realized variance (Actual RV) and the realized variance predicted by the HAR–RV model (Predicted RV) for the stock “KBANK” during 2018–2020.

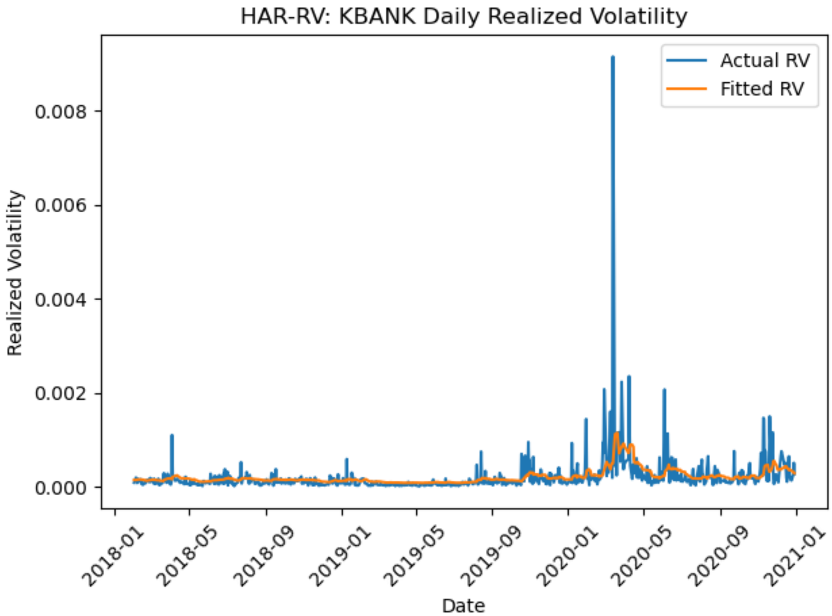


Figure 4.2, it illustrates the comparison between the actual realized variance (Actual RV) and the realized variance predicted by the HAR–RV model (Predicted RV) for the stock "KBANK" during 2018–2020. The blue line (Actual RV) represents the realized volatility computed from real data, while the orange line (Predicted RV) shows the values predicted by the HAR–RV model using 1-day, 5-day, and 22-day lag variables. It can be observed that the model successfully captures the broad trends in market volatility; however, it fails to track sharp volatility spikes accurately, as these sudden surges are often driven by factors not fully explained by the HAR model specification.

4.3 Heterogeneous Autoregressive Model for Realized Volatility Extended with Overnight Returns (HAR–RV–ON² Model)

The *HAR–RV–ON²* model is an extension of the standard HAR–RV model proposed by Corsi (2009), developed to incorporate the effect of *overnight returns*—that is, price changes occurring when the market is closed. Early studies, such as French and Roll (1986), showed that return variances differ significantly between trading and non-trading periods. Andersen and Bollerslev (1997) further demonstrated that volatility follows a distinct intraday pattern, exhibiting a U-shaped cycle with peaks at the market's open and close and a trough at midday. Subsequent studies have emphasized the importance of including this component, as it captures volatility arising from information shocks released during non-trading hours. More recent evidence by Zhang and Zhao (2021) using U.S. equity index futures reveals that over one-third—and in some cases more than 40%—of daily price variation occurs outside trading hours. This overnight volatility arises largely from the release of corporate earnings, macroeconomic data, and global market developments after hours. When trading resumes, markets rapidly assimilate this information, resulting in concentrated bursts of price discovery and heightened volatility (Ahoniemi & Lanne, 2013).

The model can be written as follows:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 RV_t^W + \beta_3 RV_t^M + \gamma\, ON_t^2 + \varepsilon_t$$

Where

$$r_t^{ON} = \ln\left(\frac{\text{Open}_{t+1}}{\text{Close}_t}\right), \quad \textbf{then}\; ON_t^2 = (r_t^{ON})^2$$

The term ON_t^2 represents the squared overnight return, which serves as a proxy for the volatility of an asset arising from *overnight news shocks*—information released while the market is closed. It is generally believed that during non-trading hours, various types of news such as corporate earnings announcements, macroeconomic data releases, and international events are made public. Therefore, ON_t^2 effectively captures the impact of this new information set.

When the opening price of the next trading day deviates significantly from the previous day's closing price, it is often followed by increased intraday volatility. Consequently, ON_t^2 provides an efficient measure of such information-driven movements and helps explain the behavior of realized variance more accurately.

If the estimated coefficient γ is positive and statistically significant, it can be interpreted as evidence that larger overnight information shocks lead to higher intraday volatility on the following trading day—even after controlling for the standard HAR terms

Example 4.2 Application of Intraday Data from the Stock Exchange of Thailand (SET) to Estimate the HAR–RV–ON² Model in Comparison with the HAR–RV Model Using Python

Step 1: Using the dataset

file *"dealprice_2018_2019_2020.csv"*, we can construct the HAR–RV–ON² model with the following command structure in Python.

```
1. import pandas as pd
2. import numpy as np
```

```
import statsmodels.api as sm
from pathlib import Path
# -----------------------------
# Load & prepare data set
# -----------------------------
fp = "dealprice_2018_2019_2020.csv"
df = pd.read_csv(fp)

df['Ticker'] = df['Ticker'].str.replace("b'", "").str.replace("'", "")
df['date'] = pd.to_datetime(df['yymmdd'].astype(int).astype(str), format='%Y%m%d')
df['timeinterval'] = df['timeinterval'].astype(int)
df = df.sort_values(['Ticker', 'date', 'timeinterval'])
df['logp'] = np.log(df['price'])
df['intra_ret'] = df.groupby(['Ticker', 'date'])['logp'].diff()

# Intraday realized variance (open⟶close) per day
rv_daily = (
    df.groupby(['Ticker', 'date'])['intra_ret']
      .apply(lambda x: np.square(x.dropna()).sum())
      .reset_index(name='RV_oc')
)
```

```python
# Overnight r_on aligned to previous date
close_df = df[df['timeinterval'] == 10][['Ticker', 'date', 'price']].rename(columns={'price': 'close'})
open_df = df[df['timeinterval'] == 1][['Ticker', 'date', 'price']].rename(columns={'price': 'open'})
open_next = open_df.copy()
open_next['date'] = open_next['date'] - pd.Timedelta(days=1)
on_merge = pd.merge(close_df, open_next, on=['Ticker', 'date'], how='inner')
on_merge['r_on'] = np.log(on_merge['open'] / on_merge['close'])
on_merge = on_merge[['Ticker', 'date', 'r_on']]
panel = pd.merge(rv_daily, on_merge, on=['Ticker', 'date'], how='left')
panel = panel.sort_values(['Ticker', 'date'])
# -----------------------------
# Pick up one ticker to show (KBANK)
# -----------------------------
```

```
ticker = 'KBANK'
d = panel[panel['Ticker'] == ticker].copy()

d['RV_oc_lead1'] = d['RV_oc'].shift(-1)
d['RV_D'] = d['RV_oc']
d['RV_W'] = d['RV_oc'].rolling(5, min_periods=5).mean()
d['RV_M'] = d['RV_oc'].rolling(22, min_periods=22).mean()
d['ON2'] = (d['r_on'] ** 2)

# We have two models
d_model_on = d.dropna(subset=['RV_oc_lead1', 'RV_D', 'RV_W', 'RV_M', 'ON2']).copy()
d_model_no = d.dropna(subset=['RV_oc_lead1', 'RV_D', 'RV_W', 'RV_M']).copy()
```

Step 1: The ticker “KBANK” is selected as an example for model estimation. The overnight return is then computed, followed by the construction of the variables RV^D, RV^W, RV^M, and ON_t^2.

Step 2: The dataset is divided into two subsets: training data (80%) and testing data **(20%).** Two models are then estimated—the HAR–RV–ON² model and the standard HAR–RV model—for comparison. The forecasting performance of both models is evaluated using the Root Mean Squared Error

(RMSE) and the Mean Absolute Percentage Error (MAPE), following the command structure shown below.

```
common_dates = set(d_model_on['date']).intersection(set(d_model_no['date']))
d_model_on = d_model_on[d_model_on['date'].isin(common_dates)].copy()
d_model_no = d_model_no[d_model_no['date'].isin(common_dates)].copy()

# Re-run train/test split

split_idx = int(len(d_model_on) * 0.8)
train_on = d_model_on.iloc[:split_idx].copy()
test_on  = d_model_on.iloc[split_idx:].copy()

train_no = d_model_no.iloc[:split_idx].copy()
test_no  = d_model_no.iloc[split_idx:].copy()

# Fit models
X_tr_no = sm.add_constant(train_no[['RV_D', 'RV_W', 'RV_M']])
y_tr_no = train_no['RV_oc_lead1']
res_no = sm.OLS(y_tr_no, X_tr_no).fit(cov_type='HAC', cov_kwds={'maxlags':5})
```

```
print(res_no.summary())


X_te_no = sm.add_constant(test_no[['RV_D', 'RV_W', 'RV_M']])
y_te_no = test_no['RV_oc_lead1']
pred_no = res_no.predict(X_te_no)

X_tr_on = sm.add_constant(train_on[['RV_D', 'RV_W', 'RV_M', 'ON2']])
y_tr_on = train_on['RV_oc_lead1']
res_on = sm.OLS(y_tr_on, X_tr_on).fit(cov_type='HAC', cov_kwds={'maxlags':5})

print(res_on.summary())

X_te_on = sm.add_constant(test_on[['RV_D', 'RV_W', 'RV_M', 'ON2']])
y_te_on = test_on['RV_oc_lead1']
pred_on = res_on.predict(X_te_on)

def rmse(a, b):
    return float(np.sqrt(np.mean((a - b) ** 2)))

```

```python
def mape(a, b):
    mask = (a != 0)
    if mask.sum() == 0:
        return np.nan
    return float(np.mean(np.abs((b[mask] - a[mask]) / a[mask])) * 100)

metrics = pd.DataFrame([
    {
        'Ticker': ticker,
        'Model': 'HAR (no ON^2)',
        'Train n': int(len(train_no)),
        'Test n': int(len(test_no)),
        'Adj_R2_train': float(res_no.rsquared_adj),
        'RMSE_test': rmse(y_te_no.values, pred_no.values),
        'MAPE_test_%': mape(y_te_no.values, pred_no.values)
    },
    {
        'Ticker': ticker,
        'Model': 'HAR + ON^2',
        'Train n': int(len(train_on)),
        'Test n': int(len(test_on)),
        'Adj_R2_train': float(res_on.rsquared_adj),
```

```
        'RMSE_test': rmse(y_te_on.values, pred_on.values),
        'MAPE_test_%': mape(y_te_on.values, pred_on.values)
    }
])

# Save outputs
outdir = Path('/Users/wasinsiwasarit/Desktop/All_work/Micro_data')
cmp_path = outdir / f'HAR_RV_compare_{ticker}.csv'
coef_no_path = outdir / f'HAR_RV_NOON_coefficients_{ticker}.csv'
coef_on_path = outdir / f'HAR_RV_ON2_coefficients_{ticker}.csv'
pred_compare_path = outdir / f'HAR_RV_predictions_compare_{ticker}.csv'

metrics.to_csv(cmp_path, index=False)

coef_no = res_no.summary2().tables[1].reset_index().rename(columns={'index':'Term'})
coef_no.to_csv(coef_no_path, index=False)

```

```python
coef_on = res_on.summary2().tables[1].reset_index().rename(columns={'index':'Term'})
coef_on.to_csv(coef_on_path, index=False)

pred_compare = pd.DataFrame({
    'date': test_no['date'].values,
    'RV_true': y_te_no.values,
    'RV_pred_HAR': pred_no.values,
    'RV_pred_HAR_ON2': pred_on.values
})
pred_compare.to_csv(pred_compare_path, index=False)


print("Saved files:")
print(" - Metrics:", cmp_path)
print(" - Coefficients (HAR no ON^2):", coef_no_path)
print(" - Coefficients (HAR + ON^2):", coef_on_path)
print(" - Predictions (test set):", pred_compare_path)
```

Table 4.3 Estimated results of the HAR–RV model for the ticker "KBANK."

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            RV_oc_lead1   R-squared:                       0.234
Model:                            OLS   Adj. R-squared:                  0.228
Method:                 Least Squares   F-statistic:                     4.292
Date:                Mon, 22 Sep 2025   Prob (F-statistic):            0.00533
Time:                        10:35:47   Log-Likelihood:                 3039.0
No. Observations:                 433   AIC:                            -6070.
Df Residuals:                     429   BIC:                            -6054.
Df Model:                           3
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       2.962e-05   1.99e-05      1.486      0.137   -9.44e-06    6.87e-05
RV_D           0.3325      0.188      1.772      0.076      -0.035       0.700
RV_W           0.4687      0.183      2.561      0.010       0.110       0.828
RV_M           0.0052      0.071      0.073      0.941      -0.134       0.145
==============================================================================
Omnibus:                      769.073   Durbin-Watson:                   2.108
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           489027.137
Skew:                          10.744   Prob(JB):                         0.00
Kurtosis:                     166.229   Cond. No.                     1.42e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 5 lags and without small sample c
orrection
[2] The condition number is large, 1.42e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Table 4.4 Estimated results of the HAR–RV–ON² model for the ticker “KBANK.”

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            RV_oc_lead1   R-squared:                       0.631
Model:                            OLS   Adj. R-squared:                  0.628
Method:                 Least Squares   F-statistic:                     54.42
Date:                Mon, 22 Sep 2025   Prob (F-statistic):           4.49e-37
Time:                        10:35:47   Log-Likelihood:                 3197.3
No. Observations:                 433   AIC:                            -6385.
Df Residuals:                     428   BIC:                            -6364.
Df Model:                           4
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.208e-05   1.18e-05      2.722      0.006    8.98e-06    5.52e-05
RV_D           0.0218      0.192      0.114      0.910      -0.355       0.398
RV_W           0.4910      0.257      1.908      0.056      -0.013       0.995
RV_M          -0.0723      0.106     -0.684      0.494      -0.280       0.135
ON2            0.5533      0.124      4.471      0.000       0.311       0.796
==============================================================================
Omnibus:                      321.795   Durbin-Watson:                   2.143
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            20658.247
Skew:                           2.507   Prob(JB):                         0.00
Kurtosis:                      36.465   Cond. No.                     1.43e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 5 lags and without small sample c
orrection
[2] The condition number is large, 1.43e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

From the estimation results of the HAR–RV model for the ticker *"KBANK"* (Table 4.3) and the **HAR–RV–ON² model** (Table 4.4), it is evident that the inclusion of the variable ON_t^2 significantly enhances the model's ability to explain the movement of realized volatility (RV). The coefficient γ is positive and statistically significant, indicating that larger overnight information shocks lead to higher intraday volatility on the following trading day.

Moreover, the coefficient of determination (R^2) increases markedly—from 0.23 in the HAR–RV model to 0.63 in the HAR–RV–ON² model—highlighting the substantial improvement in explanatory power.

When evaluating forecasting performance, the HAR–RV–ON² model also outperforms the standard HAR–RV model based on both Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) criteria, confirming its superior predictive capability.

Ticker	Model	Train n	Test n	Adj_R2_train	RMSE_ test	MAPE_ test_%
KBANK	HAR (without ON^2)	433	109	0.23	7.11E-05	101.27

KBANK	HAR + ON^2	433	109	0.63	7.01E-05	91.86

Step 3: Extract the forecasted values and present them in graphical form.

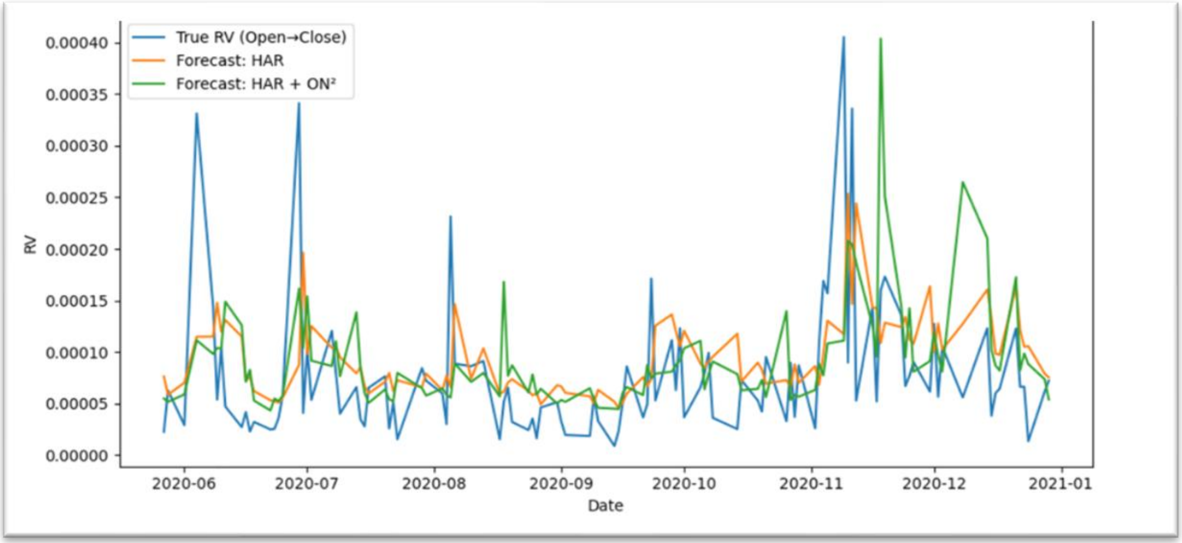
```
import matplotlib.pyplot as plt
pred_df = pd.DataFrame({
    'date': test_no['date'].values,
    'RV_true': y_te_no.values,
    'RV_pred_HAR': pred_no.values,
    'RV_pred_HAR_ON2': pred_on.values})

pred=pred_df
TICKER = "ADVANC"
plt.figure(figsize=(10, 5))
plt.plot(pred['date'], pred['RV_true'], label='True RV (Open⟶Close)')
plt.plot(pred['date'], pred['RV_pred_HAR'], label='Forecast: HAR')
plt.plot(pred['date'], pred['RV_pred_HAR_ON2'], label='Forecast: HAR + ON²')
```

```
plt.title(f'{TICKER} — Realized Variance: True vs Forecasts (Test Set)')
plt.xlabel('Date')
plt.ylabel('RV')
plt.legend()
plt.tight_layout()

out_path = Path(f"/Users/wasinsiwasarit/Desktop/All_work/Micro_data/forecast_compare_{TICKER}.png")
plt.savefig(out_path, dpi=150)
plt.show()

print("Saved plot to:", out_path)

```

It should be noted that Thailand offers a unique setting for studying overnight and intraday volatility. The Stock Exchange of Thailand (SET) closes while U.S. and European markets remain active, allowing global and local news to accumulate outside trading hours. The SET's midday trading break further contributes to distinctive intraday patterns, as prices adjust differently across sessions. Shastri et al. (1995) show that volatility in the SET is particularly high at the opening, yet little research has examined the determinants of overnight and intraday volatility in the Thai equity market.

The effects of such volatility are also heterogeneous across firms. Large firms are more exposed to international shocks, while smaller firms respond primarily to local news. Liquidity influences adjustment speed—thinly traded stocks often exhibit sharper opening jumps, whereas liquid ones adjust more smoothly. Industry linkages further matter: globally connected sectors such as exporters, importers, and financial firms tend to be more sensitive to international events than domestically oriented sectors like utilities or real estate.

Figure 4.3 Comparison between the actual realized variance (Actual RV) and the realized variance predicted by the HAR–RV model (Predicted RV) and the HAR–RV–ON² model (Predicted RV) for the stock "KBANK" during 2020–2021.



4.4 Realized GARCH (Hansen et al., 2012)

Previous volatility models did not incorporate realized volatility (RV) within the GARCH-type framework, despite well-documented evidence of RV's usefulness in improving estimates of future volatility. Recognizing this advantage, Hansen et al. (2012) introduced the Realized GARCH model, marking a major advancement in volatility modeling.[6] The model integrates the strengths of traditional conditional volatility frameworks (the GARCH family) with information from observed realized measures derived from high-frequency data. This hybrid approach enhances both the estimation accuracy and forecasting performance of volatility models by directly linking latent volatility to empirically observed market variation.

The Realized GARCH model consists of three key components:

1. **Return Equation** – describes the behavior of asset returns;

2. **Measurement Equation** – links the realized volatility to the latent conditional variance; and

---

[6] As a related extension, Hansen & Huang incorporated RV into EGARCH.

3. GARCH-type Recursion – defines the evolution of conditional variance over time.

The system can be represented as follows:

$$\begin{cases} r_t = \mu + \sqrt{h_t} z_t, \quad z_t \sim N(0,1) & (Return\ equation) \\ \log RV_t = \xi + \phi \log h_t + \tau(z_t) + u_t, \quad u_t \sim N(0, \sigma_u^2) & (Measurement\ equation: TV) \\ h_t = \omega + \beta h_{t-1} + \gamma r_{t-1}^2 & (GARCH - type\ recursion) \end{cases}$$

where

- r_t denotes the asset return,
- h_t is the latent conditional variance, and
- RV_t is the realized volatility obtained from high-frequency intraday returns.

Parameter estimation is typically performed using **Quasi-Maximum Likelihood Estimation (QMLE)**, which accommodates non-normality and model misspecification commonly observed in financial data.

Example 4.3: Application of Intraday Data from the Stock Exchange of Thailand to Estimate the Realized GARCH Model Using Python

Step 1: Using the *df* dataset from Example 2, we can construct the Realized GARCH model with the following command structure. Since there is currently no direct package that can estimate this model, we must manually create the negative log-likelihood function for the Realized GARCH(1,1) model and minimize it using the L-BFGS-B algorithm.

The L-BFGS-B method (Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound constraints) is an optimization algorithm designed to solve large-scale nonlinear optimization problems subject to boundary constraints on parameters.

```
import pandas as pd
import numpy as np
from scipy.optimize import minimize

# 1) Load intraday data & compute intraday returns
df = pd.read_csv('KBANK.csv', parse_dates=['datetime']).set_index('datetime')
df['ret_intraday'] = np.log(df['price']).diff()

# 2) Aggregate to daily realized variance (RV)
rv = df['ret_intraday'].pow(2).resample('D').sum().dropna()

# 3) Compute daily log-returns
daily_close = df['price'].resample('D').last()
daily_ret   = np.log(daily_close).diff().dropna()

```

```
# 4) Align series
data = pd.concat([daily_ret, rv], axis=1).dropna()
data.columns = ['ret', 'rv']
y = data['ret'].values
x = data['rv'].values
T = len(y)

# 5) Define negative log-likelihood for Realized GARCH(1,1)
def neg_loglik(theta):
    omega, beta, gamma, xi, phi, sigma_u = theta
    # impose positivity constraints
    if omega <= 0 or beta < 0 or gamma < 0 or phi < 0 or sigma_u <= 0:
        return 1e12
    # initialize latent h
    h = np.zeros(T)
    h[0] = x[0]
    ll = 0.0
    for t in range(1, T):
        h[t] = omega + beta * h[t-1] + gamma * x[t-1]
        if h[t] <= 0:
            return 1e12
        # return contribution
```

```
        ll += 0.5 * (np.log(2*np.pi) + np.log(h[t]) + (y[t]**2)/h[t])
        # measurement contribution
        ll += 0.5 * (np.log(2*np.pi*sigma_u**2) +
                     ((x[t] - xi - phi*h[t])**2)/(sigma_u**2))
    return ll

# 6) Initial guesses
init = np.array([1e-6, 0.9, 0.05, x.mean(), 0.8, rv.std()])

# 7) Optimize
bounds = [(1e-12, None), (0, 1), (0, 1), (None, None), (0, 10), (1e-12, None)]
res = minimize(neg_loglik, init, bounds=bounds, method='L-BFGS-B')

# 8) Output results
params = res.x
param_names = ['omega', 'beta', 'gamma', 'xi', 'phi', 'sigma_u']
for name, val in zip(param_names, params):
    print(f"{name:8s} = {val:.6f}")
print(f"\nOptimized negative log-likelihood: {res.fun:.2f}")
```

Step 2: Extract the estimated parameter values for presentation and reporting.

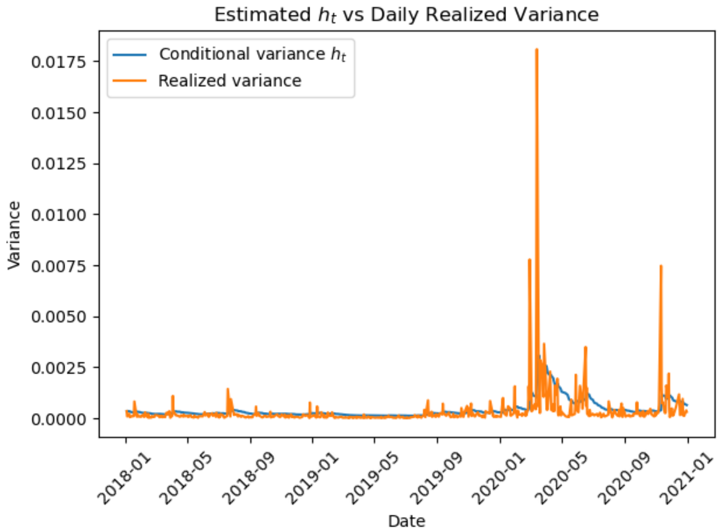
```
# 1) Extract parameters and metadata
param_names = ['omega', 'beta', 'gamma', 'xi', 'phi', 'sigma_u']
params     = res.x
neg_ll     = res.fun
k          = len(param_names)
T          = len(data)

# 2) Build a reporting DataFrame
report_df = pd.DataFrame({
    'Parameter': param_names,
    'Estimate': np.round(params, 6)
})

# 3) Display the parameter table
print(report_df)
# 4) Compute information criteria
AIC = 2 * k + 2 * neg_ll
BIC = k * np.log(T) + 2 * neg_ll

# 5) Print summary metrics
print(f"Negative log-likelihood: {neg_ll:.2f}")
```

```
22. print(f"AIC: {AIC:.2f}")
23. print(f"BIC: {BIC:.2f}")
```

```
  Parameter  Estimate
0     omega  0.000007
1      beta  0.915382
2     gamma  0.094502
3        xi -0.000024
4       phi  0.801106
5   sigma_u  0.000939
Negative log-likelihood: -4539.75
AIC: -9067.51
BIC: -9041.55
```

From the estimation results, the parameters associated with realized volatility (γ and ϕ) play an important role in explaining and predicting market volatility. The estimated correlation between realized volatility and conditional variance ($\phi = 0.80$) indicates that realized volatility effectively tracks the dynamics of the volatility process captured by the model. Meanwhile, the parameter $\gamma = 0.09$ suggests that realized volatility contributes positively to explaining the conditional variance within the model, moving in the same direction as the latent volatility component.

Figure 4.4 Comparison between the actual realized variance (Actual RV) and the conditional variance predicted by the Realized GARCH(1,1) model for the stock "KBANK" during 2018–2020.



4.5 GARCH- Mixed data sampling (GARCH-MIDAS with RV)

Although the Realized GARCH model proposed by Hansen et al. (2012) integrates realized volatility into the standard GARCH framework through a measurement equation that links latent and observed volatility—thereby improving short-horizon volatility estimation by using realized measures as observable proxies for latent variance—it remains limited in capturing long-term volatility persistence and structural shifts. As a result, while the model performs exceptionally well in short-horizon forecasting (typically 1–5 days ahead) and provides a strong contemporaneous fit with realized volatility data, it fails to account for long-memory behavior in volatility dynamics.

To address this limitation, Engle et al. (2013); Conrad and Loch (2015) introduced a new class of models known as the GARCH–MIDAS with realized volatility (GARCH-MIDAS-RV) framework. The key intuition behind this model is to utilize the trend of realized volatility over previous weeks or months to explain persistent shifts in volatility. By smoothing realized volatility through the MIDAS (Mixed Data Sampling) mechanism, this approach effectively captures long-term persistence and multi-horizon dynamics in financial market volatility.

GARCH-MIDAS-RV model betters capture the autocorrelation structure of latent and realized volatility relative to the original Realized GARCH, which is only able to capture the dependency over the very short term. This leads to substantial improvements in empirical fit (log-likelihood and information criteria) and predictive ability, particularly beyond shorter horizons, when benchmarked to the original Realized GARCH. Table 4.5 compares the key conceptual and empirical differences between the Realized GARCH model (Hansen et al., 2012) and the GARCH–MIDAS with Realized Volatility model (Engle et al., 2013; Conrad & Loch, 2015).

Table 4.5 Comparison between Realized GARCH and GARCH–MIDAS with Realized Volatility

Aspect	Realized GARCH (Hansen et al., 2012)	GARCH–MIDAS with Realized Volatility (Engle et al., 2013; Conrad & Loch, 2015)
Core Idea	Integrates realized volatility into the standard GARCH framework through a measurement equation linking latent and observed volatility.	Decomposes volatility into short- and long-run components using MIDAS weights, with realized volatility as the long-term driver.
Objective	Improve short-horizon volatility estimation and forecasting using realized measures as proxies for latent volatility.	Capture long-term persistence and multi-horizon dynamics by smoothing realized volatility through the MIDAS mechanism.
Data Frequency	Requires high-frequency intraday data to compute daily realized volatility.	Uses aggregated realized volatility (e.g., weekly or monthly averages) as the long-term component.

Table 4.5 Comparison between Realized GARCH and GARCH–MIDAS with Realized Volatility (Cont.)

Aspect	Realized GARCH (Hansen et al., 2012)	GARCH–MIDAS with Realized Volatility (Engle et al., 2013; Conrad & Loch, 2015)
Model Structure	Includes return, variance, and measurement equations estimated jointly (QMLE).	Combines daily GARCH for short-run volatility with a MIDAS regression for long-run volatility trends.
Forecast Horizon	Best for short-term forecasts (1–5 days ahead).	Better for medium- to long-term forecasts (weeks to months ahead).
Key Strength	Captures contemporaneous link between returns and realized volatility with high precision.	Captures persistence and long memory in volatility using realized volatility trends.
Limitation	Fails to explain long-term volatility shifts or structural changes.	Requires careful selection of MIDAS lag length and can over-smooth short-term dynamics.

Table 4.5 Comparison between Realized GARCH and GARCH–MIDAS with Realized Volatility (Cont.)

Aspect	Realized GARCH (Hansen et al., 2012)	GARCH–MIDAS with Realized Volatility (Engle et al., 2013; Conrad & Loch, 2015)
Empirical Focus	Financial market microstructure and short-term volatility modeling.	Macro-finance and long-term volatility risk assessment.

Source: Authors' own illustration

The GARCH–MIDAS with realized volatility (GARCH-MIDAS-RV) framework can be represented as follows:

The mean equation of univariate GARCH-MIDAS from Engle et al. (2006) can be expressed as;

$$r_{it} = \mu + \sqrt{\tau_t \, g_{it}}\, \varepsilon_{it}, \forall i = 1, \dots, N_t, \;\; \varepsilon_{it} | \Phi_{(i-1)t} \sim N(0,1)$$

From equation above, univariate log-return of each asset i is determined by (g_{it}) and (τ_t).

Where

g_{it} is short-run component

τ_t is long-run component

N_t is the number of trading days in the month or quarter t and

$\Phi_{(i-1)t}$ is the information set up to day $(i-1)$ of month or quarter t

To see pure movement from volatility, the previous equation must be demeaned.

$$r_{it} - \mu = \sqrt{\tau_t \, g_{it}}\,\varepsilon_{it}, \forall i = 1, \dots, N_t, \;\; \varepsilon_{it} | \Phi_{(i-1)t} \sim N(0,1)$$

Univariate volatility consists of the short-run component, g_{it}, and long-run component, τ_t. For simplicity, the GARCH (1,1) process is assumed to specify g_{it}.

$$g_{it} = (1 - a - b) + a\frac{(r_{it-1} - \mu)^2}{\tau_t} + bg_{i-1,t}$$

The short-run, g_{it}, component is also determined by long-run component, τ_t. In this case, the long-run component would be weighted all covariate information, V_t;

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) V_{t-k}$$

The long-run component value depends on realized volatility. We can calculate its realized volatility by using intraday data such as 30-min returns as shown in the below equation.

$$V_t = \sum_{n=1}^{N} r_{it}^2$$

The long-run component, τ_t, also needs weighting scheme, $\varphi_k(\omega_1, \omega_2)$, to weight information. The Beta weight allows to weighting scheme to weight information in most recent and old in the estimating window, following Ghysels et al. (2006)

$$\varphi_k(\omega_1, \omega_2) = \frac{\left(1 - \frac{k}{K}\right)^{\omega_1 - 1} \left(\frac{k}{K}\right)^{\omega_2 - 1}}{\sum_{j=1}^{K} \left(1 - \frac{k}{K}\right)^{\omega_1 - 1} \left(\frac{k}{K}\right)^{\omega_2 - 1}}, k = 1, \ldots . K.$$

We can achieve a decaying pattern by allowing $\omega_1 > 1$ and $\omega_2 = 1$, while the lower ω_1 can smooth realized volatility more. The k is an optimal lag of each asset.

The optimal lag in GARCH-MIDAS are need to compare log-likelihoods values. GARCH-MIDAS's log likelihood with beta weight is

$$LL = -\frac{1}{2} \sum_{t=1}^{T} [\log(\tau_t(\Phi) g_{it}(\Phi)) - \frac{(r_{it} - \mu)^2}{\tau_t(\Phi) g_{it}(\Phi)}]$$

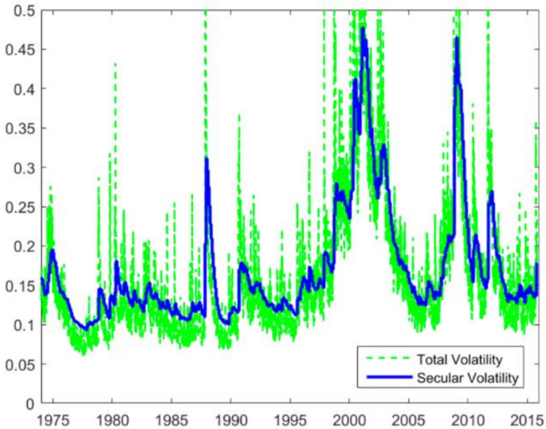
To estimate the GARCH-MIDAS model using realized volatility, researchers can employ the MIDAS MATLAB Toolbox[7]. To the best of our knowledge, a corresponding package for estimating this model is not yet available in Python.

Figure 4.5 illustrates an example of the GARCH-MIDAS model estimated using daily returns of the NASDAQ Composite Index from 1971 to 2015. The model is fitted via maximum likelihood estimation with MIDAS Beta weights. The dashed line depicts the conditional variance series, while the solid line represents the long-run volatility component. Notably, the long-run component exhibits distinct spikes corresponding to major market downturns. The total volatility increases sharply during recessionary periods, confirming the well-documented countercyclical nature of stock market volatility.

---

[7] It is available for download at https://www.mathworks.com/matlabcentral/fileexchange/45150-midas-matlab-toolbox.

Figure 4.5 A GARCH-MIDAS example using the NASDAQ Composite Index daily return data (1971 - 2015).



Source: MIDAS MATLAB Toolbox

4.6 Conclusion

The goal of this chapter has been to characterize the evolution of volatility models and to provide examples of the estimation of more recent models. In sum, a good model is recognized by its ability to generate accurate forecasts and to capture key stylized facts of volatility. The most salient features of volatility—such as clustering, persistence, and asymmetry in response to bad news—were introduced in earlier models of conditional volatility, such as GARCH and E-GARCH. These models are known to perform well over short-term horizons.

To account for longer-term expectations, we introduce models that incorporate time-varying shifts in the volatility process. Using data provided by the SET, sampled at 30-minute intervals to mitigate microstructure noise, we model the volatility of Kasikorn Bank (KBANK) over the period 2018–2020, which includes the shock from COVID-19. We also share snippets of our estimation process implemented in Python, reflecting our view that modern financial econometrics increasingly rests on the integration of market microstructure, econometric modeling, and programming.

References

Ahoniemi, K., & Lanne, M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting, 29*(4), 592–604.

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*(2–3), 115–158.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*(1), 116–131.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174-196.

Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, *30*(7), 1090–1114.

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, 5, 1–50.

Engle, R. F., & Lee, G.G.J. (1993). *A permanent and transitory component model of stock return volatility.* USCD Discussion Paper.

Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, *95*(3), 776–797

Engle, R. F., Lilien, D. M., & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, *55*(2), 391–407.

French, K. R., & Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, *17*(1), 5–26.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance, 48*(5), 1779–1801.

Ghysels, E., Harvey, A. C., & Renault, E. (1996). Stochastic volatility. In G. S. Maddala & C. R. Rao (Eds.), Handbook of Statistics: Vol. 14. Statistical Methods in Finance (pp. 119–191). Amsterdam: North-Holland.

Ghysels, E., Sinko, A., & Valkanov, R. (2006). *MIDAS regressions: Further results and new directions. Working Paper, Department of Economics*. University of North Carolina at Chapel Hill.

Hansen, P. R., & Huang, Z. (2016). Exponential GARCH modeling with realized measures of volatility. *Journal of Business & Economic Statistics, 34*(2), 269-287.

Hansen, P.R. & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics, 24*(2).

Hansen, P. R., Huang, Z., & Shek, H.H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics, 27*(6), 877-906.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

Shastri, K., Sirodom, K., & Kamaek, N. (1995). The behavior of stock return volatility in the Stock Exchange of Thailand. *Pacific-Basin Finance Journal, 3*(2–3), 367–379.

Zhang, Z., & Zhao, R. (2021). Is overnight volatility overlooked? *Journal of Futures Markets, 41*(12), 1845–1868.

Chapter 5
Conclusion and direction for future studies

5.1 Key insights

This book bridges the gap between financial theory, behavioral insights, and empirical evidence drawn from high-frequency trading data. By focusing on the Stock Exchange of Thailand (SET), it has demonstrated how intraday information can be used to study the dynamics of market participation, the impact of investor behavior, and the evolution of price formation in real time.

Chapter 1 established the theoretical and empirical foundation of market microstructure, beginning with the Efficient Market Hypothesis and the role of information asymmetry. It then introduced key indicators of investor activity—turnover and trading imbalance—which serve as behavioral and informational proxies for market dynamics. Through these measures, we can observe how different investor groups, from retail traders to institutions, display distinct behavioral traits such as overconfidence, herding, and the disposition effect, and how these traits shape liquidity provision and order submission patterns.

Chapter 2 translated theoretical models into empirical applications. The Kyle (1985) and Glosten–Milgrom (1985) frameworks were presented as foundational tools to understand how information is revealed through trading and how prices adjust in response to order flow. Applying these models to SET intraday data enables researchers to quantify information asymmetry, estimate market impact (Kyle's λ), and separate informed from liquidity-driven trades. The chapter also stressed methodological discipline—highlighting the importance of event design, data cleaning, and robustness testing—to ensure that empirical findings genuinely reflect behavioral mechanisms rather than data artifacts.

Chapter 3 brought these tools into the context of market events, where behavioral forces often intensify. Intraday analysis around earnings announcements, regulatory changes, and volatility shocks provided insight into how different investor types process new information. The case studies—such as retail investors' speculative behavior during the COVID-19 lockdown and the effectiveness of trading halts—illustrate how emotion and cognition interact with market structure to shape trading outcomes. Together, these examples highlight that market efficiency and behavioral dynamics coexist within the same institutional environment, each influencing short-term price discovery and long-term market quality.

Chapter 4 extended the discussion into advanced volatility modeling, demonstrating how intraday and realized measures can enrich our understanding of market risk. Models such as HAR–RV, HAR–RV–ON², and Realized GARCH capture the persistence and clustering of volatility across multiple horizons, while incorporating the informational value embedded in high-frequency trading. Applying these models to SET data shows how micro-level trading patterns aggregate into macro-level volatility outcomes—linking market behavior, information flow, and risk dynamics within a unified framework.

Across all chapters, three overarching themes emerge:

1. Intraday data illuminate behavioral dynamics in real time. They allow researchers to capture the timing, intensity, and persistence of investor reactions that are invisible in daily or lower-frequency data.

2. Market microstructure models serve as the bridge between theory and behavior. They provide the analytical tools to interpret psychological biases and strategic actions as measurable effects on liquidity, spreads, and price impact.

3. The SET represents a valuable and underexplored laboratory for behavioral finance. With detailed investor-type data and transparent trading mechanisms, it offers rich opportunities to

study the interplay between information asymmetry, regulation, and investor sentiment in an emerging market context.

5.2 Directions for future studies

The integration of behavioral finance, market microstructure, and intraday econometrics is only the beginning of a growing research frontier. Future studies using SET intraday data can explore several promising directions:

- Behavioral predictability of order flow: Investigate whether trading patterns of certain investor types can forecast short-term returns or volatility, and how these predictive relations evolve during periods of stress.
- Information diffusion and trading networks: Examine how information spreads across investor groups and sectors within the same trading day, revealing the pathways of market learning and contagion.
- Impact of regulation and technology: Assess how structural changes—such as algorithmic trading, transaction taxes, or disclosure rules—affect market efficiency, liquidity, and behavioral asymmetries.

- Integration with machine learning methods: Combine traditional models like Kyle and Glosten–Milgrom with modern data-driven techniques to classify trading behavior, detect informed trades, or forecast volatility using microstructure features.
- Cross-market behavioral comparisons: Extend intraday behavioral analyses beyond Thailand, comparing SET dynamics with those of developed and regional markets to highlight cultural, structural, and psychological differences in trading behavior.

By pursuing these directions, future researchers can deepen our understanding of how human behavior, information processing, and institutional design collectively shape market outcomes. The continuous development of intraday data analytics thus offers not only methodological advancement but also an enduring opportunity—to see markets not just as systems of prices, but as dynamic reflections of human judgment, belief, and interaction.

NOTE

Master in Finance –
International Program (MIF)

Thammasat Business School, Thammasat University
Room F-310, 3rd Floor, Multipurpose-II Building, 2 Prachan Rd,
Phra Nakhon, Bangkok 10200 Thailand.

https://mif.tbs.tu.ac.th/

Thailand Capital Market Development Fund

93 Ratchadaphisek Rd, Din Daeng, Bangkok 10400

https://www.cmdf.or.th/

9 786166 298628