IMPERIAL

Imperial College London

Department of Mathematics

Calibrating PDE Limit Order Book Model to Market Data

Author: Waner Chen (CID: 01854827)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2024-2025

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Philipp Jettkant, for his invaluable guidance and support throughout this project. He not only proposed the original research topic, but also provided me with professional insight and expertise that greatly enriched the practical dimension of my work. At the same time, his feedback and advice on the theoretical aspects were essential in shaping the direction and quality of this thesis.

I am also deeply grateful to my family for their unwavering encouragement and support during my studies. Their understanding and care have been a constant source of strength.

Finally, I would like to thank my friends for their companionship and encouragement, which have made this journey both enjoyable and memorable.

Abstract

This thesis investigates the structural properties and dynamic behavior of limit order books (LOBs) by combining high-frequency empirical analysis with partial differential equation (PDE) modeling. Using level-20 data from the LOBSTER database across multiple U.S. equities and industry sectors, the study provides a long-horizon perspective on both local boundary dynamics and global depth curve stability. A key contribution is the empirical examination of boundary conditions near the mid-price: by estimating liquidity gradients and order book slopes, the analysis evaluates whether Dirichlet, Neumann, or Robin-type conditions are most consistent with observed market microstructure, thereby grounding PDE and stochastic PDE formulations in empirical reality. Beyond the boundary, the global shape of depth curves is studied in terms of its persistence across different time scales and assets, offering new insights into price resilience and market stability. Finally, by linking discrete order-event dynamics with continuous PDE/SPDE frameworks and calibrating parameters such as diffusion, cancellation, and replenishment rates directly to tick-level data, the thesis demonstrates how microscopic order flow drives macroscopic liquidity profiles, thereby contributing to a more empirically validated framework for continuous modeling of limit order books.

Contents

1 Limit Order Book Structure and Dynamics **9**
1.1 Structure of the Limit Order Book . . . 9
1.2 Events Driving LOB Dynamics . . . 10
1.3 Example: Mid-Price Dynamics . . . 10
1.4 From Discrete to Continuous Representation . . . 11
1.5 Outlook: Modeling Shape Dynamics . . . 11

2 Data Source and Preprocessing **12**
2.1 Data Source: LOBSTER Database . . . 12
2.2 Sample Selection Criteria . . . 12
2.3 Data Preprocessing . . . 13

3 Empirical Depth and Flow Analysis **15**
3.1 Depth Curve Construction and Analysis . . . 15
3.1.1 Depth Curve Construction from Order Book Snapshots . . . 15
3.1.2 Effect of Window Length . . . 16
3.1.3 Cross-Asset Comparison . . . 16
3.2 Curve Merging for Depth Alignment . . . 19
3.3 Flow Curves from Message Data . . . 19
3.4 Gradient Estimation . . . 22
3.5 Measures and Descriptive Statistics . . . 25
3.5.1 Empirical Features . . . 25
3.5.2 Empirical Patterns . . . 27

4 PDE Modeling Framework for Limit Order Books **29**
4.1 Model Assumptions . . . 29
4.2 Mathematical Formulation . . . 29
4.3 Motivation for Relative Coordinates . . . 31
4.4 Relative Coordinate System and Redefined Functions . . . 31
4.5 PDE Transformation to Relative Coordinates . . . 31
4.6 Parameter Definitions and Economic Interpretation . . . 32
4.7 Boundary Condition Selection . . . 32

5 Numerical Discretization and Parameter Estimation **33**
5.1 Spatial and Temporal Discretization . . . 33
5.2 Parameter Estimation . . . 34
5.2.1 Data-Driven Initial Values . . . 34
5.2.2 Joint Calibration via Curve Matching . . . 34
5.2.3 Extension: Position-Dependent Cancellation Intensity . . . 36

6 Price Impact of Market Orders **40**
6.1 Mechanics of Price Impact under a Linear Depth Profile 40
6.2 Numerical Illustration: Constant-Speed Execution 41
6.3 Joint System and Finite-Difference Scheme with Trading Impact 43
6.4 Price Impact: Joint System vs. Approximation 45

List of Figures

2.1 Example of an order book snapshot at level-20 depth. The bid side (left) and ask side (right) illustrate the distribution of standing volumes across price levels relative to the mid-price. 14

3.1 Example of average depth curve with a 30-minute window length. 16
3.2 Depth curves for AAPL on 5 March 2025 under different window lengths. . 17
3.3 30-minute average depth heatmap for AAPL on 5 March 2025. 17
3.4 Depth curves for AMZN, JPM, TSLA, and WMT on 5 March 2025 under full-day window. 18
3.5 Depth curves for TSLA on different days. 18
3.6 Depth curve plots for AAPL on 18 Jun 2018. 20
3.7 New-order flow curves for AAPL, AMZN, JPM, and TSLA on 5 March 2025. 21
3.8 Deleted-order flow curves for AAPL, AMZN, JPM, and TSLA on 5 March 2025. 22
3.9 Flow Curve Overlay for AAPL. 22
3.10 Flow Curve Overlay for JPM. 23
3.11 Merged Depth Curve Gradients for AAPL (2025-03-05). 24
3.12 Merged Depth Curve Gradients for TSLA (2025-03-05). 24
3.13 Near-mid Gradient (Ask) across days for AAPL. 25
3.14 Near-mid Gradient (Bid) across days for AAPL. 25
3.15 Correlation matrices in 2018 and 2022. 28

5.1 Zoomed-in view of the order book density near the mid-price (AAPL, Dirichlet, Uniform deletion). 35
5.2 Zoomed-in view of the order book density near the mid-price (AAPL, Robin, Uniform deletion). 36
5.3 Empirical depth curve for AAPL on 6 March 2025. The histogram shows the average standing volume across relative price levels with respect to the mid-price. 36
5.4 Zoomed-in view of the order book density near the mid-price (TSLA, Neumann, Uniform deletion). 37
5.5 Empirical depth curve for TSLA on 6 March 2025. The histogram shows the average standing volume across relative price levels with respect to the mid-price. 37
5.6 Zoomed-in view of the order book density near the mid-price (AAPL, Dirichlet, Position-dependent deletion). 38
5.7 Zoomed-in view of the order book density near the mid-price (AAPL, Robin, Position-dependent deletion). 38
5.8 Zoomed-in view of the order book density near the mid-price (TSLA, Neumann, Position-dependent deletion). 39
5.9 Zoomed-in view of the order book density near the mid-price (TSLA, Dirichlet, Position-dependent deletion). 39

6.1 Price impact vs time under constant trading speed (stationary linear profile, Neumann boundary conditions). Curves correspond to $q \in \{0.5, 1.0, 2.0\}$. The steeper time paths for larger q reflect the impact–speed relation in equation (6.1.1). 42
6.2 Price impact vs cumulative volume Q_t under constant trading speed. The curves for different q collapse onto the closed-form square-root law implied by (6.1.2)–(6.1.3). 43
6.3 Joint system: final u with Neumann boundary conditions. 44
6.4 Price impact vs time: Approximation (solid) vs Joint mean (dashed). 45

Introduction

In modern electronic financial markets, the limit order book (LOB) is the central mechanism for price formation and liquidity provision, recording all outstanding limit orders and matching them through market orders. Understanding the dynamics of the LOB— including the evolution of bid and ask quotes, the distribution of order volume across price levels, and the impact of order flow on prices — is crucial for market participants, regulators, and researchers alike. High-frequency LOB data captures millisecond-level submissions, cancellations, and executions of orders, enabling rigorous market microstructure analysis and revealing complex patterns in liquidity distribution, price fluctuations, and order flow imbalance (Biais et al. [2]; Cont et al. [4]).

However, due to the high dimensionality of the limit order book, its complex state dependence, and the interaction between discrete order events and continuous price dynamics, modeling it remains a significant challenge. Traditional approaches often simplify the problem by focusing on the "top of the book" (best bid/ask prices and volumes) or by employing low-dimensional Markov processes, which may overlook important features in the deeper levels of the book (Cont et al. [8]; Cont and de Larrard [6]). In recent years, researchers have increasingly applied tools from stochastic analysis, in particular stochastic partial differential equations (SPDEs), to the modeling of the full structure of the limit order book. These approaches build on well-established theoretical foundations, but their application to market microstructure is relatively recent, providing a framework that bridges microscopic order flow dynamics with macroscopic emergent properties (Bayer et al. [1]; Cont and Müller [7]).

This paper aims to contribute to the literature by analyzing the dynamics of high-depth limit order books (e.g., level = 20) across multiple time scales, combining empirical analysis with mathematical tools such as stochastic process modeling and gradient-based diagnostics. By integrating multi-scale statistical analysis and boundary-sensitive curve characterization, we seek to address key gaps in the existing research: the lack of systematic analysis of high-depth data, particularly the boundary behavior and the global shape of the LOB, and the empirical evaluation of model assumptions using long-term order book data.

Empirical research on limit order books can be traced back to Biais et al. [2], one of the first studies to systematically characterize the order book structure and order flow of the Paris Bourse. They found that orders tend to cluster near the best quotes and that the bid–ask spread is negatively correlated with depth. Hasbrouck [10] provided a comprehensive theoretical framework, incorporating the role of the limit order book in price discovery and liquidity provision, while Bouchaud et al. [3] introduced a physics-inspired perspective, highlighting power-law distributions in order flow and price impact. Early modeling efforts focused on simplified assumptions; for example, Cont et al. [8] proposed a Poisson-process-based stochastic model for limit order book dynamics to describe order arrivals and cancellations, showing that the order book state (the quantity at each price level) follows an ergodic Markov process. This work laid the foundation for analyzing price transitions and liquidity risk, but its discrete-state framework and short-horizon focus left open questions regarding the global shape of the order book and its boundary behavior.

A key development in limit order book modeling is the derivation of scaling limits, which link microscopic order events to macroscopic dynamics. Horst and Paulsen [11] established a law of large numbers for limit order books, showing that under specific scalings of tick size, order arrival rates, and order impact, the best bid/ask prices and the volume density converge to a deterministic system consisting of ordinary differential equations (ODEs) coupled with linear hyperbolic partial differential equations (PDEs). This result provides a theoretical foundation for approximating discrete order book dynamics by continuous models, particularly for long-term analysis. Building on this, Bayer et al. [1] incorporated state-dependent price dynamics, where price changes depend on the volume of outstanding orders. They derived a functional limit theorem, demonstrating that the order book converges to a coupled system of stochastic differential equations (SDEs) and SPDEs, in which prices follow a diffusion process while the volume density is governed by an infinite-dimensional SPDE. This framework captures both the deterministic trends and random fluctuations in the deeper levels of the order book, bridging microscopic randomness and macroscopic continuity.

Cont and Müller [7] developed a tractable SPDE model for the entire limit order book, centered around the mid-price, in which multiplicative noise captures the high-frequency fluctuations of order flow. Their framework explicitly links mid-price dynamics to order book depth and imbalance, and provides a finite-dimensional implementation that facilitates parameter estimation and numerical computation. By incorporating a mean-reverting volume density, the model reproduces empirical features such as volatility clustering and liquidity persistence, making it suitable for practical applications like optimal execution. Complementary work on diffusion approximations, such as Hambly et al. [9], focuses on reflected SPDEs to model boundary constraints—such as the non-negativity of volumes—ensuring consistency with real-market conditions.

Order book events, such as market orders and cancellations, have been a central topic in empirical market microstructure research. Cont et al. [5] analyzed NYSE stock data and showed that price changes are driven by the order flow imbalance, with a robust linear relationship between imbalance and price impact. They further demonstrated that the slope of this relationship is inversely related to market depth, emphasizing the importance of modeling deeper levels of the limit order book, as liquidity away from the best quotes affects price elasticity.

High-quality limit order book data are essential for validating theoretical models and conducting fine-grained empirical research. Huang and Polak [12] developed LOBSTER (Limit Order Book System – The Efficient Reconstructor), which reconstructs the full-depth limit order book from NASDAQ TotalView-ITCH message data, providing researchers with high-frequency order flow information across multiple price levels (e.g., level = 20). With LOBSTER, detailed analyses can be conducted on the precise time series of order submissions, cancellations, and executions.

However, most empirical studies based on LOBSTER focus on relatively short time windows (e.g., several days to weeks) or on lower depth data. Long-term analyses of high-depth order books remain scarce, which limits our understanding of how order book stability and deep liquidity distribution evolve across different market cycles.

In summary, while significant progress has been made in modeling and empirical analysis of limit order books, important gaps remain in our understanding of their structural properties. A recurring limitation in the existing literature is the treatment of boundary behavior: most PDE and SPDE models impose boundary conditions at the best bid and ask in an ad hoc manner, without strong empirical justification. This practice risks misrepresenting the true dynamics of liquidity provision and consumption near the mid-price, precisely where market impact is most acute.

This thesis contributes to filling this gap by conducting a systematic empirical study of the boundary behavior of the limit order book, using level-20 data from the LOBSTER database. By estimating gradients and liquidity slopes near the mid-price, we provide direct evidence on which boundary conditions—Dirichlet, Neumann, or Robin-type—are most consistent with observed market microstructure. This boundary-sensitive analysis goes beyond the conventional "top-of-book" focus and grounds the modeling framework in empirical reality.

A second contribution of this thesis is the study of the global shape of the LOB, and in particular its stability across time and across assets. While prior work has mostly concentrated on shallow depths or short horizons, we examine the long-term evolution of depth curves, comparing their stability at multiple time scales (from seconds to days) and across sectors. This provides new insights into how deep-layer liquidity contributes to price resilience and market stability.

Finally, by linking discrete order-event dynamics with continuous PDE/SPDE formulations, we illustrate how microscopic order flows can provide insights into macroscopic depth profiles. This connection allows us to test model assumptions empirically, and to calibrate parameters such as diffusion, cancellation, and replenishment rates directly to tick-by-tick data. The empirical analysis focuses on key features including (i) liquidity gradients near the mid-price; (ii) the overall depth curve shape and its cross-sectional heterogeneity; (iii) the persistence of these shapes across time scales; (iv) the statistical distribution of order arrivals and cancellations and (v) the contribution of numerical simulations in validating dynamics.

Together, these contributions provide a more empirically grounded framework for continuous modeling of limit order books, addressing both the local boundary behavior and the global shape of liquidity, which have so far received insufficient attention in the literature.

Chapter 1

Limit Order Book Structure and Dynamics

This chapter provides a preliminary overview of the structure and dynamics of a LOB. Since the LOB is the central mechanism of modern electronic financial markets, it is important to describe its functioning before turning to empirical datasets and mathematical modeling. We introduce the discrete state representation of the LOB, the types of events that modify it, and how these events affect fundamental quantities such as the best bid, best ask, and mid-price. Finally, we discuss the continuous approximation of the order book, which motivates the later use of PDEs.

1.1 Structure of the Limit Order Book

At any point in time t, the LOB records all outstanding *limit orders* submitted by market participants. Each limit order specifies a price p, a quantity q, and a side (buy or sell).

To describe the LOB, let $\{p_i\}_{i=1}^{M}$ denote the set of all possible price ticks, ordered increasingly ($p_1 < p_2 < \cdots < p_M$). For each tick, we record the standing buy and sell volumes separately:

$$q_i^b(t) = \text{total buy-side volume posted at price } p_i^b,$$

$$q_i^a(t) = \text{total sell-side volume posted at price } p_i^a.$$

The best bid and best ask at time t are then defined as

$$B_t = \max\{p_i : q_i^b(t) > 0\}, \qquad A_t = \min\{p_i : q_i^a(t) > 0\}.$$

The mid-price is given by their arithmetic average:

$$S_t = \frac{A_t + B_t}{2}.$$

In practice, the set of bid and ask ticks evolves with the mid-price unless the distribution of standing orders remains unchanged. Although the notation allows both $q_i^b(t)$ and $q_i^a(t)$ to be positive at the same tick, in real-world markets this situation does not occur: one should not observe buy orders at prices above the best ask, nor sell orders below the best bid. Thus the inequality $B_t < A_t$ always holds.

1.2 Events Driving LOB Dynamics

The state of the LOB evolves in continuous time as new orders are submitted, existing orders are cancelled, or standing orders are executed by market orders. These mechanisms determine the dynamics of the best bid, best ask, and consequently the mid-price.

The first type of event is the limit order submission. When a trader submits a new order of price p and quantity q, the standing volume at the corresponding price level is updated. If the price p lies strictly within the bid–ask spread, i.e. $B_t < p < A_t$, the spread is narrowed. This can alter either the best bid or the best ask depending on the side of the book, and thereby has an immediate impact on the mid-price.

The second type of event is the market order arrival. A buy market order of size q consumes volume starting from the best available ask price, while a sell market order consumes standing volume on the bid side. If the size of the incoming market order exceeds the available liquidity at the best quote, the entire volume at that level is removed from the book and the next available tick becomes the new best ask (in the case of a buy market order) or best bid (in the case of a sell market order). This change in the best quote leads directly to a discrete jump in the mid-price.

The third type of event is the order cancellation or deletion. Traders often withdraw their limit orders before execution, either to manage risk exposure or to re-post at more favorable prices. When a cancellation occurs at a price level that is not at the best quote, the event does not affect the mid-price but does modify the local liquidity profile of the book. However, if the cancelled order lies at the best bid or best ask and its removal exhausts the standing volume at that level, the corresponding best price moves outward to the next available tick. In this case, the cancellation produces a shift in the bid–ask spread and alters the mid-price in a manner similar to a market order.

Together, these three types of events—submissions, executions via market orders, and cancellations—constitute the fundamental drivers of LOB dynamics. Their interplay generates the microstructural price formation process observed in modern electronic markets.

1.3 Example: Mid-Price Dynamics

Let $\delta = 0.01$ denote the tick size. At time t suppose the best bid and best ask are $B_t = 100.00$ and $A_t = 100.01$, so the mid-price is $S_t = \frac{A_t+B_t}{2} = 100.005$.

Consider a buy market order arriving at time t. A market buy consumes standing liquidity on the ask side beginning at A_t. If its size q is smaller than the available quantity at the best ask, the order is partially filled at A_t and the best quotes remain unchanged; hence the mid-price stays at S_t. If, however, q is at least the resting quantity at the best ask, the level A_t is fully depleted and the next available ask, say $p_2^a = 100.02$, becomes the new best ask. In this case the mid-price jumps to

$$S_{t^+} = \frac{B_t + p_2^a}{2} = \frac{100.00 + 100.02}{2} = 100.01,$$

illustrating how executions that exhaust top-of-book liquidity induce discrete mid-price moves.

A different mechanism arises with limit orders posted inside the spread. To avoid crossing the ask, the spread must be at least two ticks. Suppose the book is in a state with $B_t = 100.00$ and $A_t = 100.02$, so the spread is 2δ, and the mid-price S_t is 100.01. A newly submitted buy limit order at 100.01 improves the bid without interacting with the ask, becomes the new best bid by price-time priority, and narrows the spread to δ. The updated mid-price is

$$S_{t^+} = \frac{100.01 + 100.02}{2} = 100.015.$$

By contrast, if a buy limit is posted at a price $\geq A_t$ (e.g. 100.02), it is marketable and executes immediately against the ask rather than resting in the book.

1.4 From Discrete to Continuous Representation

The limit order book is inherently a discrete object: at each price tick p_i, one records the total outstanding volume $q_i(t)$ at time t. On NASDAQ, the tick size is $\delta = 0.01$ USD, so price levels are separated only by one cent. Given the large number of active participants and the high frequency of order arrivals, the volume profile across ticks can fluctuate rapidly. For example, in highly liquid equities, one can routinely observe several million order book events within a single trading day, corresponding to hundreds of updates per second on average. For mathematical modeling, it is therefore advantageous to approximate this discrete profile by a smooth volume density function.

To formalize this, define the relative price coordinate $x = p - S_t$, where S_t denotes the prevailing mid-price. Under this convention, negative values of x correspond to the bid side of the book, while positive values correspond to the ask side. The collection of discrete volumes $\{q_i(t)\}$ can then be embedded into a continuous representation by defining a function $u_t(x) \approx$ volume density at distance x from the mid-price.

In other words, instead of viewing the LOB as a staircase of volumes indexed by ticks, we regard it as a continuous curve $x \mapsto u_t(x)$. This curve captures the overall shape of supply and demand around the mid-price, rather than the microscopic discreteness of individual ticks. Within this framework, the dynamics of the LOB can be interpreted as the evolution of the density $u_t(x)$ over time, driven by the cumulative effect of order submissions, cancellations, and executions.

This continuous approximation will serve as the foundation for subsequent analysis. It allows one to replace the combinatorial complexity of individual ticks with analytic tools from the study of PDEs, while still retaining the essential features of order-book dynamics at the macroscopic scale.

1.5 Outlook: Modeling Shape Dynamics

The continuous perspective is particularly valuable because it enables the use of analytical tools such as stochastic processes and PDEs. In later chapters, we will develop a PDE-based framework in which diffusion terms capture the reallocation of orders across nearby price levels, decay terms represent cancellations, and source terms describe new submissions as a function of the distance to the mid-price. By grounding the model in the empirical structure outlined above, we aim to connect the discrete microstructure of the LOB with its continuous large-scale dynamics.

Chapter 2

Data Source and Preprocessing

This chapter outlines the dataset and preparatory steps forming the foundation for the analysis. It first introduces the LOBSTER database as the source of high-frequency limit order book data, then specifies the criteria for selecting a representative sample of stocks. Finally, it describes the basic preprocessing applied to transform raw data into a standardized format ready for subsequent depth curve construction.

2.1 Data Source: LOBSTER Database

The dataset employed in this study is obtained from LOBSTER, an academic data service that reconstructs high-frequency LOB data from NASDAQ's Historical TotalView-ITCH message feed [13]. TotalView-ITCH provides a complete record of all visible order submissions, cancellations, deletions, and executions, with time stamps of millisecond precision or better. Using these raw messages, LOBSTER reconstructs the full order book for a specified security and time period, producing both a message file and an order book file for each trading day.

The message file contains a chronologically ordered list of all LOB events within the requested time period. Each record includes the event time (seconds after midnight with sub-second precision), the event type (submission, cancellation, deletion, visible execution, hidden execution, or trading halt), the order ID, the order size, the price (quoted in integer multiples of 0.0001 USD), and the trade direction (buy or sell). The order book file contains snapshots immediately after each event, reporting price and standing volume at each bid and ask level from level 1 (best bid/ask) up to the requested maximum depth.

For this study, the maximum depth is set to 20 levels on each side, allowing the capture of both near and far liquidity. The millisecond-level precision of LOBSTER data, combined with its comprehensive coverage of order submissions and executions, makes it particularly suitable for analyzing market microstructure and order book shape. In particular, the depth information at multiple levels provides the necessary granularity for constructing and comparing depth curves, which form a central component of the present analysis.

Based on this dataset, the next section outlines the sample selection criteria adopted in this study, before proceeding to the subsequent data preprocessing and analysis stages.

2.2 Sample Selection Criteria

To ensure that the dataset supports stable and robust estimation, the analysis concentrates on highly liquid, large-cap stocks whose order books display both sufficient depth and a high frequency of updates. Such market conditions provide a rich sequence of observations that mitigates idiosyncratic microstructure noise and facilitates more reliable estimation of limit order book depth curves.

To mitigate sector-specific biases and enhance the generalizability of the results, the sample encompasses firms across multiple industries. It includes large-cap technology companies, where intense algorithmic trading generates rapid order book dynamics; financial institutions such as banks and brokerages, whose trading activity is closely tied to macroeconomic conditions; consumer discretionary firms, which reflect shifts in broader economic cycles; and healthcare companies, often distinguished by unique volatility patterns. This cross-sector design captures variations in order book behavior driven by industry-specific trading dynamics, from the rapid reactions of technology stocks to market signals to the more stable, institutionally dominated flows in financials, thereby improving the robustness of model calibration.

The analysis in this study begins in 2015 in order to avoid distortions caused by major corporate actions such as Apple's 7-for-1 stock split in 2014, which altered the price scale and affected historical comparability. Exceptional market episodes such as the COVID-19 pandemic in 2020 are included for specific analyses but are excluded when examining baseline structural regularities, so as to reduce the influence of extraordinary volatility and policy-driven market conditions.

The dataset covers regular NASDAQ trading days. To reduce the influence of opening and closing auction imbalances, as well as the heightened volatility and irregular order flow in these periods, the first and last thirty minutes of each trading day are excluded from the baseline analysis. These window parameters are implemented in the code and can be adjusted for robustness checks or alternative specifications.

2.3 Data Preprocessing

Before conducting the main analysis, the raw LOBSTER order book data are transformed into a standardized format to ensure comparability across different assets and trading days. This step involves basic but essential preprocessing operations, which prepare the dataset for subsequent computations and analyses.

For each order book snapshot, the mid-price S_t is computed and all quoted levels on both bid and ask sides are expressed in relative price coordinates $x = p - S_t$. Prices in the LOBSTER dataset are reported as integers scaled by 10^4; accordingly, all price fields are divided by 10^4 to obtain values in USD.

To capture the full depth of the book, the twenty best bid levels and twenty best ask levels are extracted for each trading second and reshaped into a long format (x, q), where x is the relative price and q is the standing volume at that price. This format facilitates aggregation of volume at a given distance from the mid-price across both levels and time within a specified window, forming the basis for the construction of depth curves in later sections.

To illustrate the structure of the dataset, Figure 2.1 provides an example of a limit order book snapshot. The left panel shows the bid side of the book, while the right panel shows the ask side. Each row corresponds to a price level, with its associated standing volume. This visualization clarifies how market liquidity is distributed across different depths and how the snapshot representation forms the foundation for constructing depth curves in subsequent analyses.

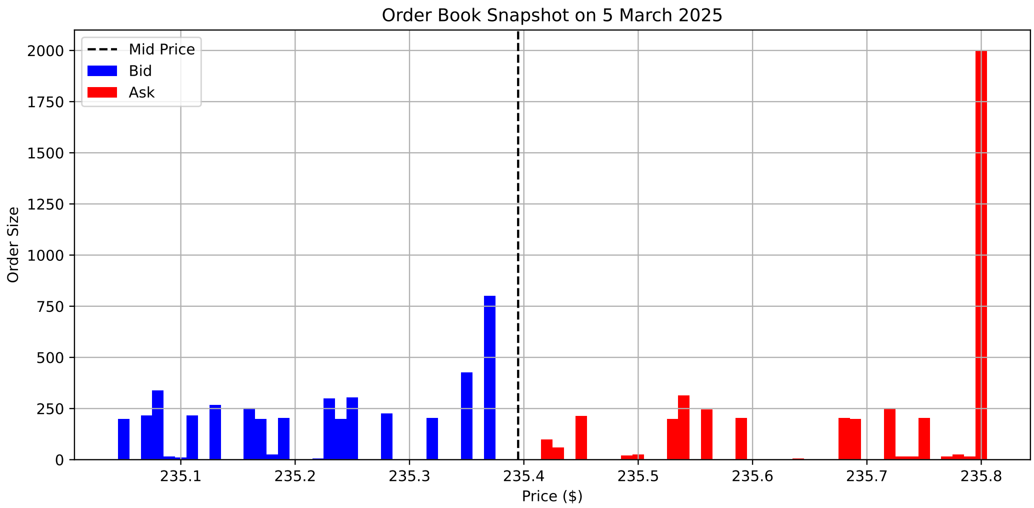


Figure 2.1: Example of an order book snapshot at level-20 depth. The bid side (left) and ask side (right) illustrate the distribution of standing volumes across price levels relative to the mid-price.

Chapter 3

Empirical Depth and Flow Analysis

3.1 Depth Curve Construction and Analysis

3.1.1 Depth Curve Construction from Order Book Snapshots

To construct depth curves, the order book data is segmented into non-overlapping windows of fixed length W minutes, with the step size equal to the window length to ensure no overlap. The first and last thirty minutes of each trading day are excluded before windowing to avoid opening and closing imbalances. Within a window, relative prices are discretized on a 0.005-USD grid. Although the exchange tick size is 0.01 USD, transforming quotes to mid-price–centered coordinates implies that the natural spacing is half a tick: when the bid–ask spread equals one tick, the mid-price lies exactly midway between the best bid and best ask, so these quotes sit at ± 0.005 USD relative to the mid. Using a 0.005-USD grid therefore captures both odd-tick spreads (e.g., 0.01) and even-tick spreads (e.g., 0.02) without aliasing and avoids systematic rounding asymmetries around $x = 0$.

Let $\{(p_i^b, q_i^b)\}_{i=1}^{20}$ and $\{(p_i^a, q_i^a)\}_{i=1}^{20}$ denote the bid- and ask-side prices and volumes at levels 1–20 for each snapshot. After stacking buy and sell sides, each (x, q) pair is assigned to its price bin and aggregated. For each bin x, the depth curve records the window-average standing volume, obtained by summing quantities mapped to x at every snapshot and dividing by the number of snapshots in the window. The resulting series $x \mapsto \bar{q}(x)$ is the depth curve for that window.

Window-end "anchors" mark the time indices used as keys for downstream analysis and plotting. Curves are stored as a dictionary indexed by these window anchors and are used for zoomed views, static plots, and animations with fixed axes, enabling visual diagnostics of intraday stability. This construction method allows consistent computation across different window lengths while preserving the flexibility to examine both microstructural details and aggregated liquidity patterns.

Figure 3.1 illustrates an example of the aggregated depth curve. It can be observed that volumes drop sharply outside the interval $[-0.2, 0.2]$. This pattern is a direct consequence of restricting the order book to 20 levels. As shown earlier in Figure 2.1, not every price level close to the mid-price is populated in each snapshot. Since only 20 levels are retained, quotes beyond ± 0.2 (calculated as 0.01×20 based on the even-tick spread) are truncated. Consequently, when stacking and averaging across snapshots, the volume outside this range is underestimated due to the varying effective depth across time. Therefore, in all subsequent figures, we truncate the depth curves to the interval $[-0.2, 0.2]$ for consistency.

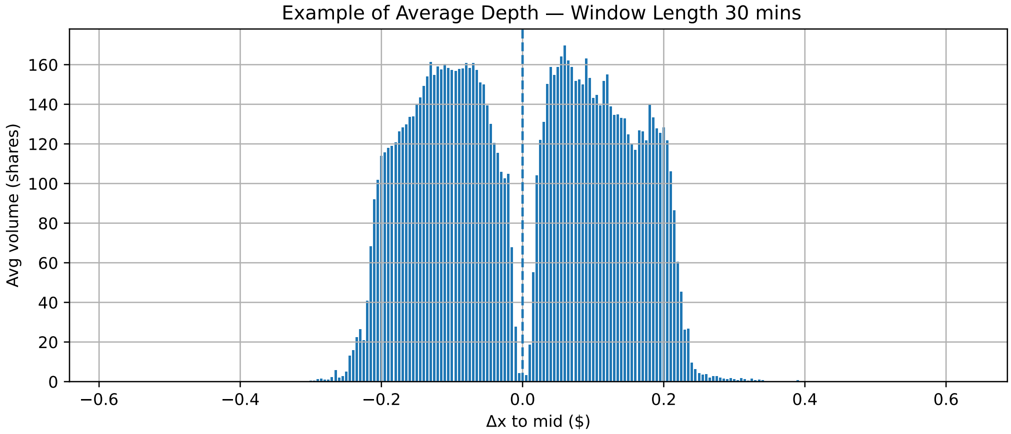


Figure 3.1: Example of average depth curve with a 30-minute window length.

3.1.2 Effect of Window Length

In depth curve construction, the choice of time window length is a key parameter that significantly influences the resulting curve shape and the market features it captures. To illustrate these methodological differences, this study considers four representative time scales: ultra-short (0.01 minutes), short (1 minute), medium (30 minutes), and full-day (approximately 300 minutes).

Ultra-short and short windows preserve the highest level of microstructural detail, enabling the detection of transient liquidity changes and rapid order book dynamics. However, this high temporal resolution comes at the cost of increased noise, as short-term fluctuations may obscure persistent structural patterns.

Medium-length windows achieve a balance between detail retention and stability. The resulting curves are smoother and better suited for statistical analysis, while still reflecting meaningful intraday variations. Full-day aggregation produces highly stable and noise-resistant curves, ideal for modeling the overall liquidity environment. This aggregation inevitably masks intraday changes, as most micro-level fluctuations are averaged out.

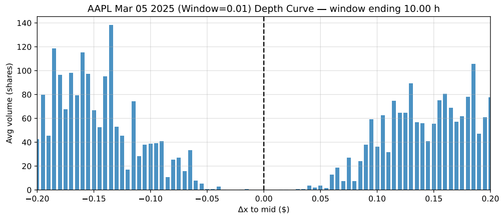
In practice, all curves are generated using the same computational procedure, with the window length being the only variable. Figure 3.2 presents a visual comparison of depth curves under the four window configurations.

In addition, heatmap visualizations, such as Figure 3.3, show that the overall shape of the depth curve remains broadly consistent throughout the trading day, with only minor fluctuations and noise at certain times. This day-level stability suggests that transient deviations are limited in scope. Consequently, to reduce the influence of short-term microstructural noise, subsequent analyses primarily utilise full-day aggregated curves.

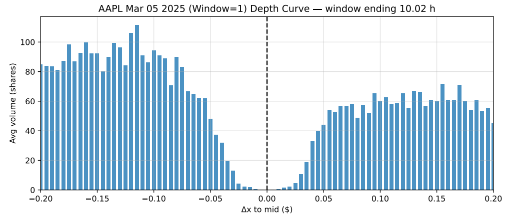
3.1.3 Cross-Asset Comparison

The cross-asset comparison examines depth curve characteristics across four representative stocks from different sectors (Figure 3.4). Based on a broader set of downloaded samples over multiple periods, two dominant liquidity profiles emerge. The first, exemplified by AMZN (Figure 3.4(a)), features relatively high liquidity concentrated near the mid-price, forming a smooth and symmetric depth profile. In contrast, the second profile, observed in TSLA (Figure 3.4(c)) on the date shown, exhibits relatively sparse liquidity across price levels, with shallower volumes and a more convex shape near the mid-price, indicating higher sensitivity of prices to incoming orders.

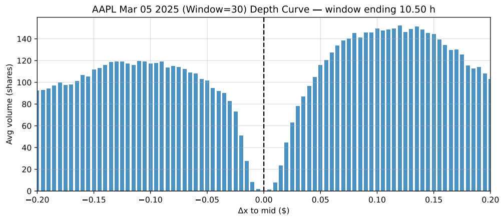
It should be emphasized, however, that such shapes are not universal. Depth curve



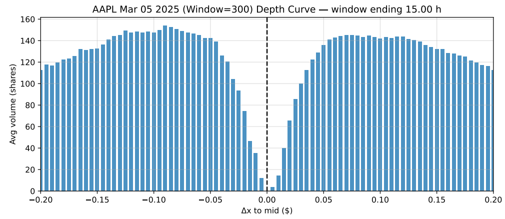
(a) 0.01-minute window



(b) 1-minute window



(c) 30-minute window



(d) Full-day window

Figure 3.2: Depth curves for AAPL on 5 March 2025 under different window lengths.

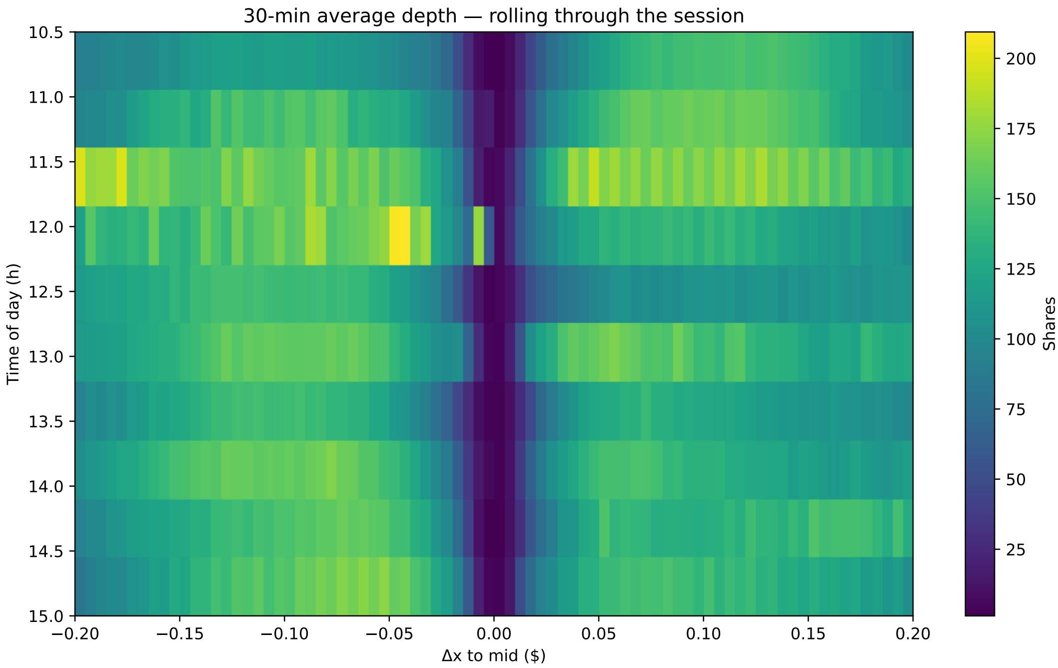
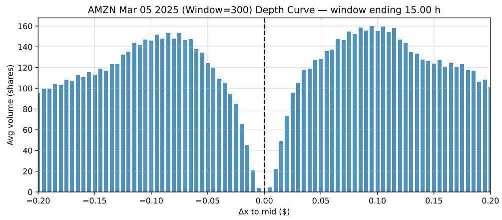


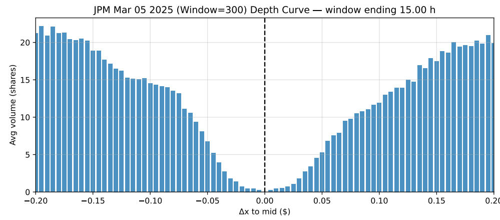
Figure 3.3: 30-minute average depth heatmap for AAPL on 5 March 2025.

profiles often remain relatively stable across adjacent trading days, yet over longer horizons the convexity or concavity near the mid-price may shift, resulting in markedly different shapes on other dates (Figure 3.5). This temporal variation suggests that the observed profiles capture prevailing market conditions rather than immutable structural characteristics of the asset. Several factors may contribute to these changes. First, variations in trading activity or algorithmic order placement across days can shift liquidity concentration toward or away from the mid-price. Second, macroeconomic announcements or firm-specific news may alter traders' willingness to provide liquidity at tighter spreads. Third, the overall level of average trading volume appears to correlate with the depth curve shape: on days with relatively high volume, such as 15 Feb 2023 (Figure 3.5(a)), TSLA's profile tends to show a denser liquidity concentration across price levels, whereas on lower-volume days, such as 14 Jan 2025 (Figure 3.5(c)), the profile exhibits sparser liquidity and a more convex form near the mid-price. The case of 25 Mar 2025 (Figure 3.5(d)) represents an intermediate state, with liquidity distribution lying between these two extremes. Finally, structural features such as periodic volatility regimes can influence whether depth curves appear more convex or concave around the mid-price.

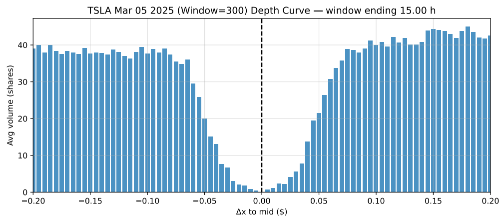
Additionally, certain stocks, such as WMT (Figure 3.4(d)), display depth profiles with alternating "high-low" volumes across adjacent price bins. This pattern arises mechanically from tick-size conventions—specifically, shifts between half-tick and full-tick quoting—rather than reflecting genuine side-to-side liquidity migration. To address these tick-alignment artifacts, the following section introduces a curve-merging approach that consolidates the alternating bins into a smoother and more stable depth representation.



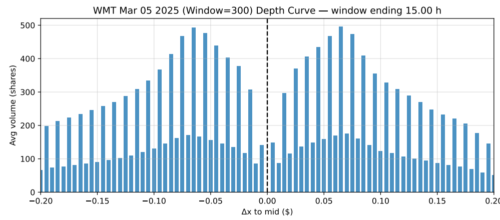
(a) AMZN



(b) JPM

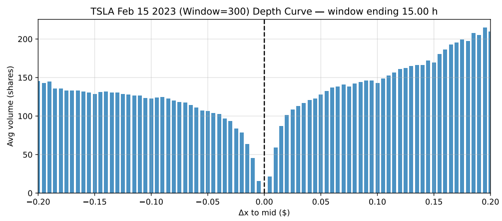


(c) TSLA

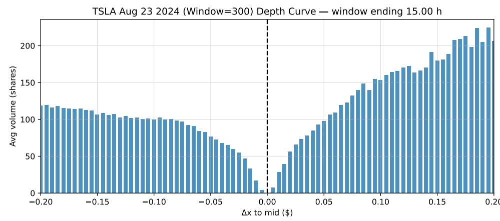


(d) WMT

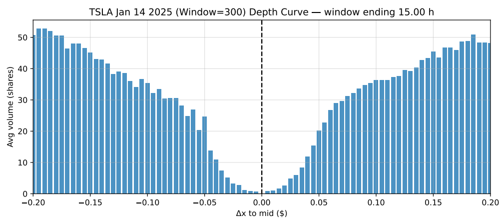
Figure 3.4: Depth curves for AMZN, JPM, TSLA, and WMT on 5 March 2025 under full-day window.



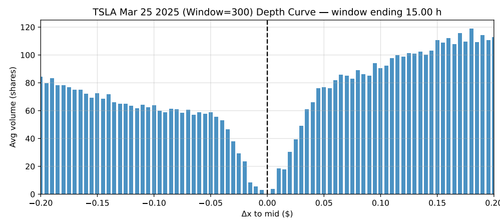
(a) 15 Feb 2023



(b) 23 Aug 2024



(c) 14 Jan 2025



(d) 25 Mar 2025

Figure 3.5: Depth curves for TSLA on different days.

3.2 Curve Merging for Depth Alignment

The alternating "high–low" patterns observed in stocks such as WMT arise from the discrete nature of quoting under a fixed tick size. Since the bid–ask spread can only take integer multiples of the tick size, two distinct quoting regimes appear in mid-price coordinates. When the spread equals an odd multiple of the tick size, quotes are aligned with the grid $\{0.005 + 0.01k\}_{k\in\mathbb{Z}}$. When the spread equals an even multiple, they are instead aligned with the interleaved grid $\{0.010 + 0.01k\}_{k\in\mathbb{Z}}$. Crucially, the spread does not remain fixed but alternates between odd and even multiples. As soon as one regime becomes more frequent than the other, liquidity accumulates preferentially on one grid while the other grid appears depleted. This imbalance creates the characteristic "high–low" oscillation in the depth profile: bins associated with the dominant regime appear elevated, while bins from the less frequent regime appear suppressed. The resulting saw-tooth pattern is therefore not due to genuine liquidity migration, but rather an artifact of alternating half-tick versus full-tick quoting combined with unequal probabilities of odd versus even spreads.

To obtain a more stable and market-aligned representation, we merge half-tick bins outward to the nearest full-tick levels:

$$x \mapsto \operatorname{sign}(x) \left\lceil \frac{|x|}{0.01} \right\rceil \times 0.01.$$

For example, $+0.005$ and $+0.010$ are combined at $+0.010$, $+0.015$ and $+0.020$ at $+0.020$, and symmetrically on the bid side (-0.005 with -0.010, -0.015 with -0.020, etc.). This preserves the buy/sell side and avoids mixing signs.

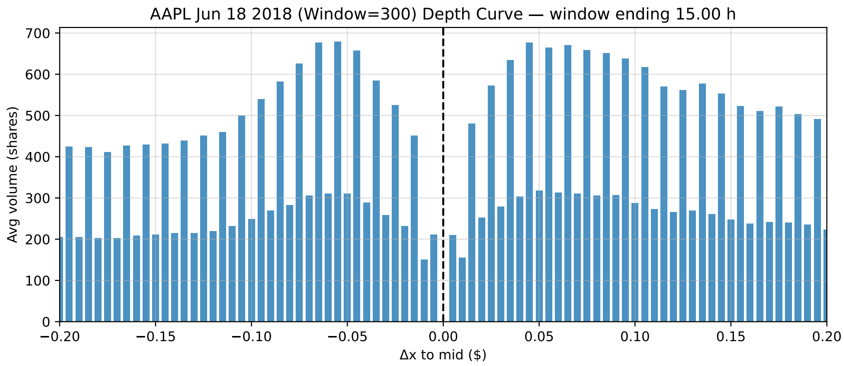
The effect is illustrated in Figure 3.6, which compares the AAPL 18 June 2018 depth curve from a raw window (showing alternating "high–low" oscillations due to half-/full-tick interleaving) with the corresponding outward-merged curve over the same interval. The merged curve exhibits a smoother shape and consistent spacing, reflecting the intended grid alignment.

While this outward merge introduces a slight conservative bias by pushing volume marginally away from the mid-price, it avoids the drawbacks of alternative schemes. For instance, midpoint-collapsing pairs (e.g., 0.0075, 0.0175, …) place bin centers off the quoted grid and create an uneven gap at the origin: the distance between -0.0075 and $+0.0075$ is 0.015, whereas adjacent mid-points on the same side are 0.01 apart. Conversely, inward-collapsing (e.g., $0.015 \to 0.010$) would over-concentrate near-mid liquidity, and in particular, both $+0.005$ and -0.005 would collapse to 0.000, where no volume should exist at the mid-price. By contrast, the outward merge keeps bin centers aligned with the exchange's quoting increments, reducing spurious oscillations and improving cross-sample comparability.

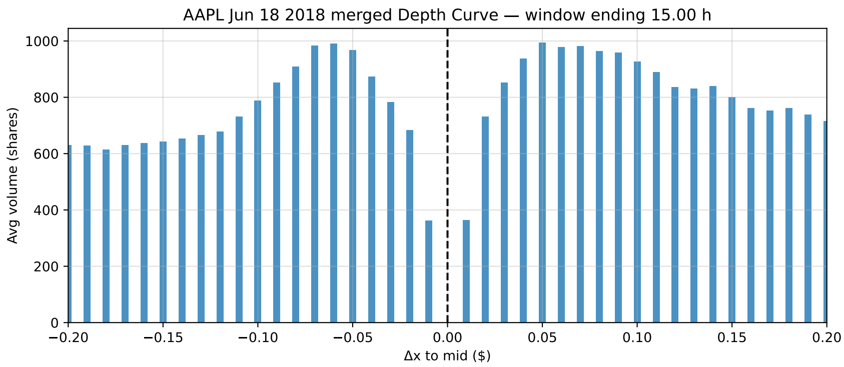
3.3 Flow Curves from Message Data

Message files are used to construct simple flow curves for two event types: new-order events (Type 1) and cancellation/deletion events (Types 2–3). The relative price is computed as $x = p - S_t$, where p is the quoted order price and S_t is the prevailing mid-price. For new orders, a basic sign-consistency check is applied: buy limits are posted below the mid ($x < 0$, Direction $= +1$), and sell limits above the mid ($x > 0$, Direction $= -1$). No sign constraint applies to cancellations or deletions.

Figures 3.7 and 3.8 present flow curves for four representative stocks, constructed in mid-price coordinates with bids in blue and asks in red. Figure 3.7 corresponds to new-order events, while Figure 3.8 shows cancellation and deletion events. While the absolute



(a) Unmerged depth curve.



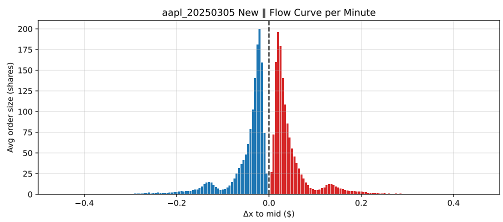
(b) Merged depth curve.

Figure 3.6: Depth curve plots for AAPL on 18 Jun 2018.

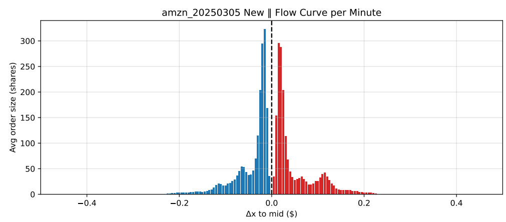
shapes differ across names, a common pattern is relatively stable across stocks once the same mid-price–centered conversion and normalization are applied: the flow curves on both sides of the book generally exhibit a pronounced peak in liquidity close to the mid-price, accompanied by thinner volumes further away.

A closer examination highlights stock-specific characteristics. AAPL exhibits a pronounced peak near the mid-price, accompanied by additional minor shoulders at larger price offsets on both event types (Figures 3.7(a), 3.8(a)). AMZN shows a sharply concentrated distribution around the mid-price with little evidence of secondary structures (Figures 3.7(b), 3.8(b)). In contrast, JPM appears largely unimodal in our sample, with a broad peak centered close to the mid-price (Figures 3.7(c), 3.8(c)). TSLA displays more complex multi-modal patterns, with pronounced peaks emerging at different relative price levels, particularly on the deletion side where a strong secondary cluster is visible (Figures 3.7(d), 3.8(d)).

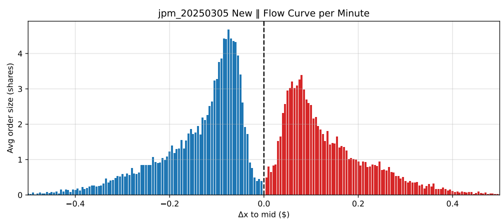
A plausible interpretation is that the major mid-price peak reflects systematic liquidity provision by market makers who refresh quotes close to the mid-price to capture the spread. The additional shoulders or secondary peaks, as observed in AAPL, AMZN and TSLA, may be driven by more conservative participants or block-like trading activity that clusters liquidity at specific offsets from the mid-price. Such secondary features are more likely to arise when the average order size is high, consistent with block-like trading activity that concentrates liquidity at particular relative price levels. These layered peak structures—when they occur—highlight heterogeneity in trading styles and risk tolerance.



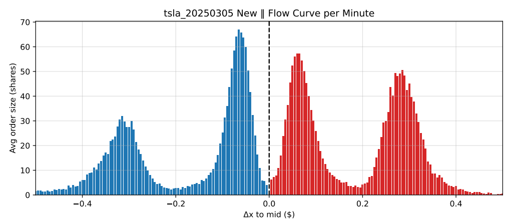
(a) AAPL



(b) AMZN



(c) JPM



(d) TSLA

Figure 3.7: New-order flow curves for AAPL, AMZN, JPM, and TSLA on 5 March 2025.

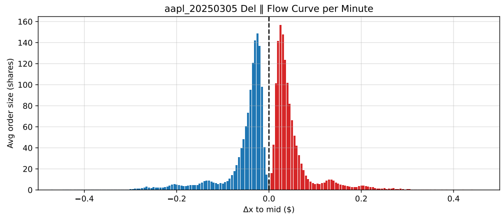
Motivated by the visual similarity between the new- and deleted-order flow curves, we test a simple *constant-fraction* hypothesis: there exists a scalar $c \in (0, 1]$ such that the deletion curve $D(x)$ is well approximated by a scaled version of the insertion curve $I(x)$,

$$D(x) \approx c\, I(x) \quad \text{for } x \in \mathcal{X},$$

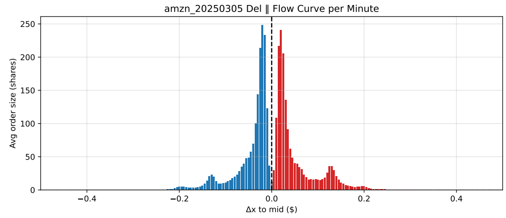
with $\mathcal{X}$ the set of relative-price bins. The parameter c is estimated by least squares,

$$\hat{c} \;=\; \arg\min_{c>0} \sum_{x \in \mathcal{X}} \big(D(x) - c\, I(x)\big)^2 \;=\; \frac{\sum_x I(x) D(x)}{\sum_x I(x)^2},$$

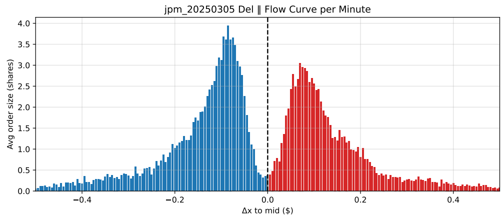
and a root-mean-squared error (RMSE) is reported as a goodness-of-fit summary.



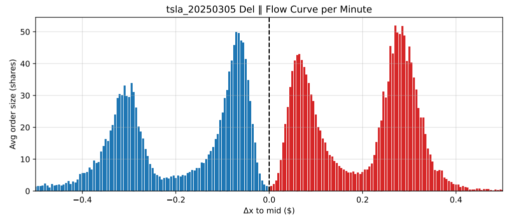
(a) AAPL



(b) AMZN



(c) JPM



(d) TSLA

Figure 3.8: Deleted-order flow curves for AAPL, AMZN, JPM, and TSLA on 5 March 2025.

For visualization, we convert the binned histograms to smooth curves: the scaled insertion curve is plotted in blue and the deletion curve in red. We then "shrink" the insertion curve by the estimated factor $\hat{c}$, making the two shapes directly comparable. Figures 3.9 and 3.10 show two representative examples (AAPL and JPM). In both cases, a *single* scale factor delivers a striking visual fit over a wide range of x: peaks, shoulders, and decay rates align closely after scaling, and the RMSE remains small. This suggests that, at the chosen aggregation level, deletions behave approximately like a constant fraction of insertions.

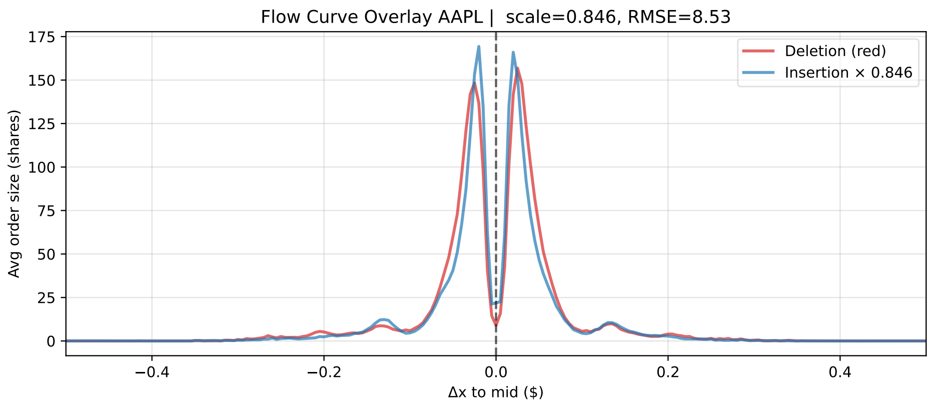


Figure 3.9: Flow Curve Overlay for AAPL.

3.4 Gradient Estimation

The slope of the depth curve at the mid price encapsulates how aggressively liquidity concentrates around the best quotes. A steep one-sided slope indicates a sharp peak close to the mid and a rapid decay of standing volume as one moves one tick away; a flat slope reflects a wide plateau with relatively even liquidity near the mid. In the qualitative taxonomy of Section 3.1.3, names that exhibit a tight, symmetric hump (e.g., AMZN on

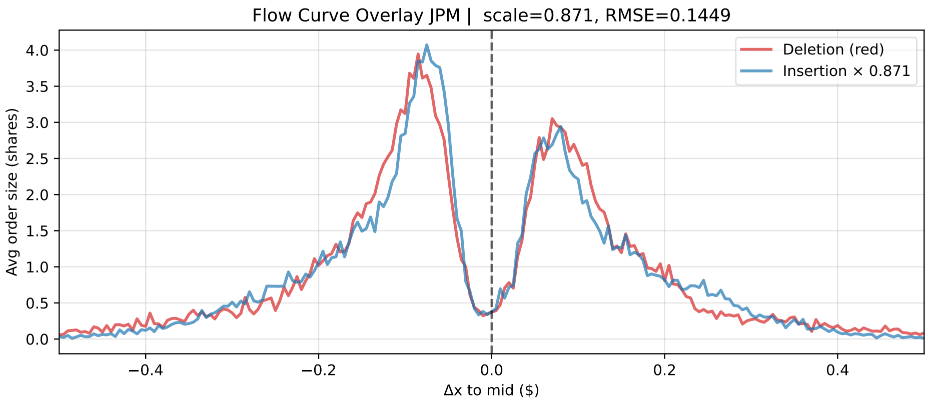


Figure 3.10: Flow Curve Overlay for JPM.

our sample date) correspond to large near-mid slopes in absolute value, whereas broader, shallower profiles (e.g., some TSLA dates) correspond to small near-mid slopes. From the modeling perspective, large slopes are consistent with Dirichlet-like boundary behavior (little residual volume at the best quotes), while small slopes are closer to a Neumann-type specification (a flat profile at the interface). Intermediate cases can be interpreted through a Robin coefficient that links the boundary level to its slope.

The first approach applies a naive one-sided finite difference on the bid side using only the two levels closest to the mid. With tick size Δx,

$$\nabla^{-}u(0) \;\approx\; \frac{u(-\Delta x) - u(-2\Delta x)}{\Delta x},$$

this produces a straightforward estimate of the local slope but is highly sensitive to bin-level noise.

A more accurate scheme uses a second-order one-sided stencil with three neighbouring points:

$$\nabla^{-}u(0) \;\approx\; \frac{3u(-\Delta x) - 4u(-2\Delta x) + u(-3\Delta x)}{2\,\Delta x},$$

which reduces bias for smoother profiles, though it can magnify irregularities if the depth curve oscillates.

To further stabilize estimates, we also fit a low-order polynomial to the nearest bins on the bid side (Savitzky–Golay style) and take its derivative at $x = 0$. The same procedures are applied symmetrically on the ask side.

Overall, the gradient estimates highlight a key structural feature of the order book: assets with liquidity concentrated near the best quotes exhibit steep slopes at the mid, whereas those with more dispersed liquidity show flatter slopes. This observation directly informs the boundary condition choice in the PDE framework, linking empirical microstructure patterns to continuous modeling. At the same time, gradient estimation is inevitably affected by discretization, spread alignment, and local fluctuations. Techniques such as polynomial smoothing and the merging scheme of Section 3.2 help mitigate these effects and produce more robust estimates.

To illustrate more clearly what the gradient estimates capture, Figure 3.11 shows the depth curve of AAPL with the estimated slopes overlaid on both sides of the mid-price. The lines indicate the local tangents we approximate, and the picture makes it evident that AAPL exhibits a steep, peaked profile where the gradient magnitudes are large. By contrast, Figure 3.12 presents the same visualization for TSLA, which is characterized by a much flatter shape around the mid-price and gradients that are orders of magnitude smaller. The near-mid profile is correspondingly shallow, with liquidity distributed more evenly across ticks and less tightly concentrated at the top-of-book.

When we move from single snapshots to cross-day evidence, the contrast becomes even clearer. Figures 3.13 and 3.14 show daily gradients for AAPL across multiple trading days. The results confirm that AAPL remains consistently in the steep, peaked regime, with sizeable slopes on both sides of the book, although the overall level of steepness drifts across dates. Taken together, these visuals highlight the existence of two distinct regimes: a compact, concave liquidity profile with large near-mid gradients, as in AAPL, and a shallow, convex profile with small gradients, as in TSLA.

Although the three estimators deliver gradients of broadly comparable magnitude, their behavior differs in important ways. The naive estimator is highly sensitive to bin-level noise and often exaggerates local irregularities. The polynomial smoother (SG) can in some cases even return misleading signs for the gradient, as the fitted curve does not always reflect the discrete book structure near the mid. By contrast, the second-order estimator provides more stable and interpretable slopes while remaining directly tied to the observed bins. For this reason, the subsequent analysis primarily relies on the second-order specification. Nevertheless, it should be noted that the second-order scheme is not free from limitations and occasional inaccuracies, which will become evident in the subsequent correlation figures, so the SG results are retained as a robustness check.

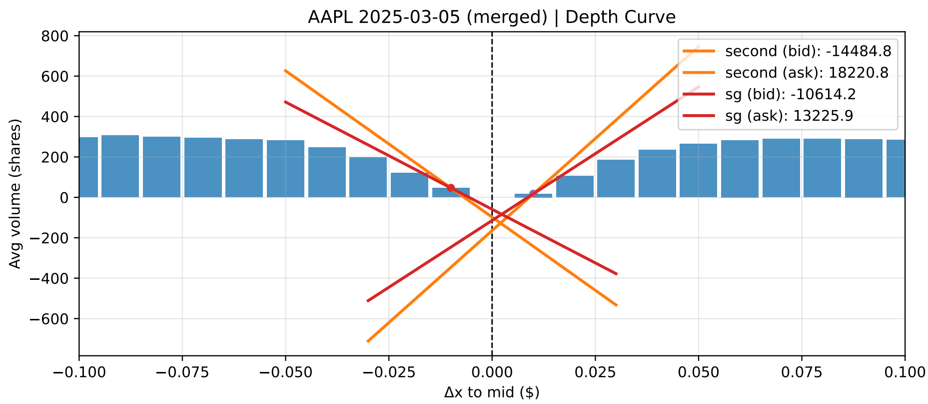


Figure 3.11: Merged Depth Curve Gradients for AAPL (2025-03-05).

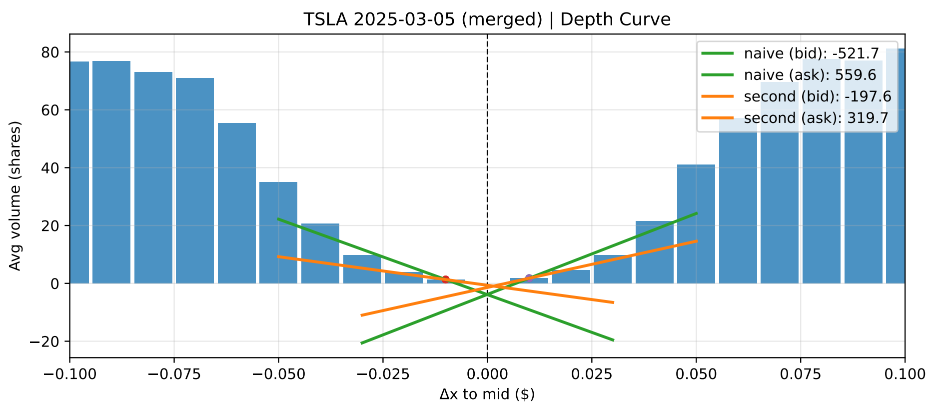


Figure 3.12: Merged Depth Curve Gradients for TSLA (2025-03-05).

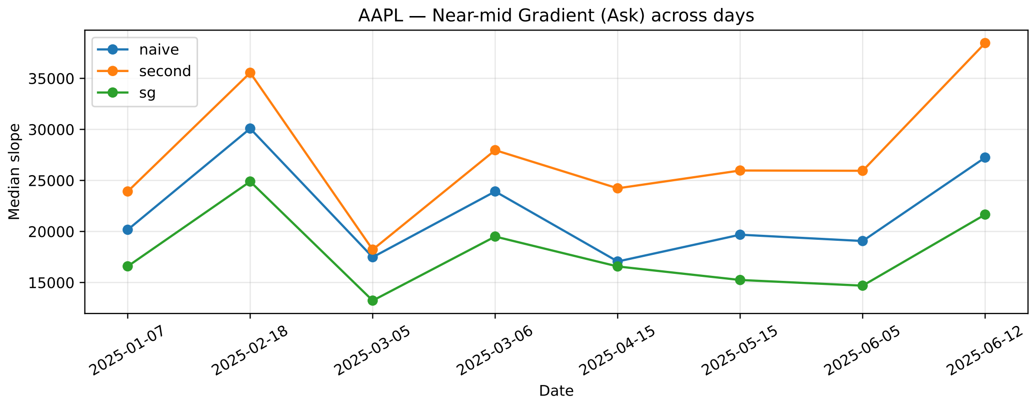


Figure 3.13: Near-mid Gradient (Ask) across days for AAPL.

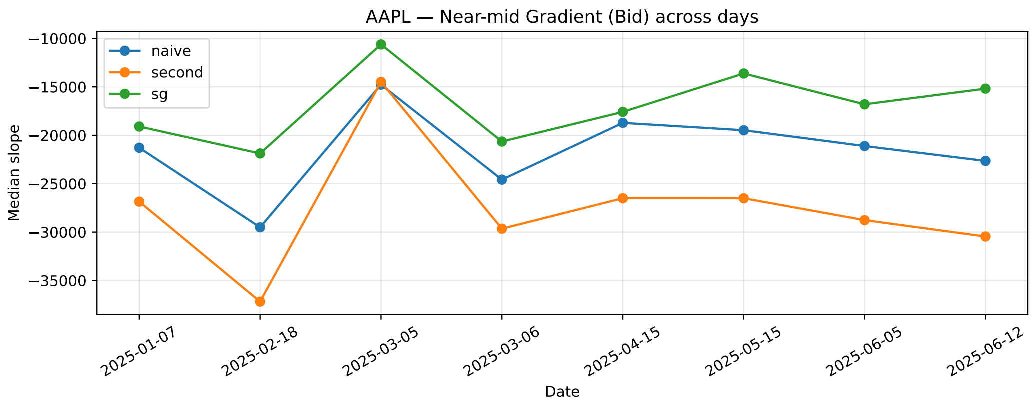


Figure 3.14: Near-mid Gradient (Bid) across days for AAPL.

3.5 Measures and Descriptive Statistics

The aim of this section is to understand how the profile of the limit order book in the vicinity of the mid-price influences subsequent price changes, as captured for example by the realized mid-price volatility. Building on the construction of depth-based gradients, we introduce a set of empirical measures that summarize volatility, spread characteristics, and local liquidity around the mid-price, and examine their joint behavior. This consolidation step bridges the methodological choices of Sections 3.1–3.4 with the subsequent modeling and estimation.

3.5.1 Empirical Features

We begin by introducing the key features extracted from the order book that will later be related to the PDE model dynamics. These features are designed to capture volatility, spread characteristics, and local liquidity around the mid-price.

Definition 3.5.1 (Realized volatility)**.** Let M_t denote the mid-price and let $t_0 < \cdots < t_K$ be observation times in a window $W = [t_0, t_K]$. Define log-returns $r_k = \log M_{t_k} - \log M_{t_{k-1}}$. The realized variance and realized volatility on W are

$$\mathrm{RV}^2(W) = \sum_{k=1}^{K} r_k^2, \qquad \mathrm{RV}(W) = \sqrt{\mathrm{RV}^2(W)}.$$

This provides a natural measure of intraday uncertainty by quantifying short-horizon price variability based on mid-price log returns.

Definition 3.5.2 (Mean spread and scaled mean spread)**.** On a window $W = [t_0, t_K]$, let A_t and B_t denote the best ask and best bid, and define the point-in-time spread $s_t := A_t - B_t$. Let M_t be the mid-price. We define

$$\bar{s}(W) := \frac{1}{K}\sum_{k=1}^{K} s_{t_k}, \qquad \bar{s}_{\text{mid}}(W) := \frac{1}{K}\sum_{k=1}^{K} \frac{s_{t_k}}{M_{t_k}}.$$

We refer to these as Mean Spread and Scaled Mean Spread, and note that a wider spread signals thinner liquidity near the mid price.

Definition 3.5.3 (Spread composition)**.** To capture quoting conventions, we record the empirical distribution of discrete spread values. For instance, spreads of the form $0.005 + 0.01k$ versus $0.01 + 0.01k$ are distinguished, and we also document the frequency of discrete spread levels (e.g. 100, 200, 300, 400, 500 in lobster units).

We define two measures:

1. *Share100 prop.* For assets such as AAPL where spreads of 100 and 200 dominate,
$$\text{Share100Prop} := \frac{\#\{\text{Spread} = 100\}}{\#\{\text{Spread} = 100 \text{ or } 200\}},$$
representing the proportion of one-tick spreads relative to the two most frequent spread values.

2. *Odd-tick prop.* To quantify the prevalence of odd-tick spreads, we define
$$\text{OddTickProp} := \frac{\#\{\text{Spread} \in \{100, 300, 500, \ldots\}\}}{\#\{\text{All spreads}\}},$$
representing the share of spreads with odd multiples of the base tick size.

Together, these measures provide a concise description of spread composition, capturing both fine-scale quoting conventions and broader odd–even tick patterns.

Definition 3.5.4 (Liquidity near the mid)**.** Liquidity near the mid-price is computed in two complementary ways:

1. *Curve-based liquidity* is obtained after constructing the depth curve within a window, by aggregating standing volumes around the mid on the smoothed grid. This highlights the overall curve shape close to the mid.

2. *Frame-based liquidity* is obtained before curve formation, at each snapshot of the book within the window, by averaging volumes at the first three levels on each side. This retains intra-window fluctuations and is more sensitive to fleeting liquidity.

We denote these measures by *NearMidLiq(Curve)* and *NearMidLiq(Frame)*, respectively.

Definition 3.5.5 (Depth gradient)**.** We define the depth gradient as the slope of the bid and ask depth curves, computed using the second-order method described in Section 3.4. The gradient summarizes how liquidity decays away from the mid-price on each side of the book, serving as a complementary measure of market resilience.

Unlike realized volatility 3.5.1, which is computed directly from the raw mid-price series, the depth gradient is estimated from the smoothed depth curves constructed over a window. In other words, gradient estimates reflect the average shape of the order book around the mid-price, rather than snapshot-by-snapshot fluctuations.

3.5.2 Empirical Patterns

We examine the empirical relationships among the constructed microstructure variables across multiple years of data. Rather than using the full daily record for each year, we randomly sample around 30–40 trading days per year to obtain a representative but computationally manageable dataset. This sampling frequency strikes a balance between capturing broad time variation and ensuring tractability of the analysis. For illustration, the correlation matrices for 2018 and 2022 as shown in Figures 3.15(a) and 3.15(b) are presented, with similar structures observed in the other sample years.

A few key relationships stand out. First, realized volatility, the mean spread, and the scaled mean spread are all highly correlated, which is consistent with the idea that periods of high volatility are typically accompanied by wider spreads and less favorable trading conditions. This suggests that these three measures largely capture a common underlying dimension of market activity, namely the level of uncertainty and trading frictions.

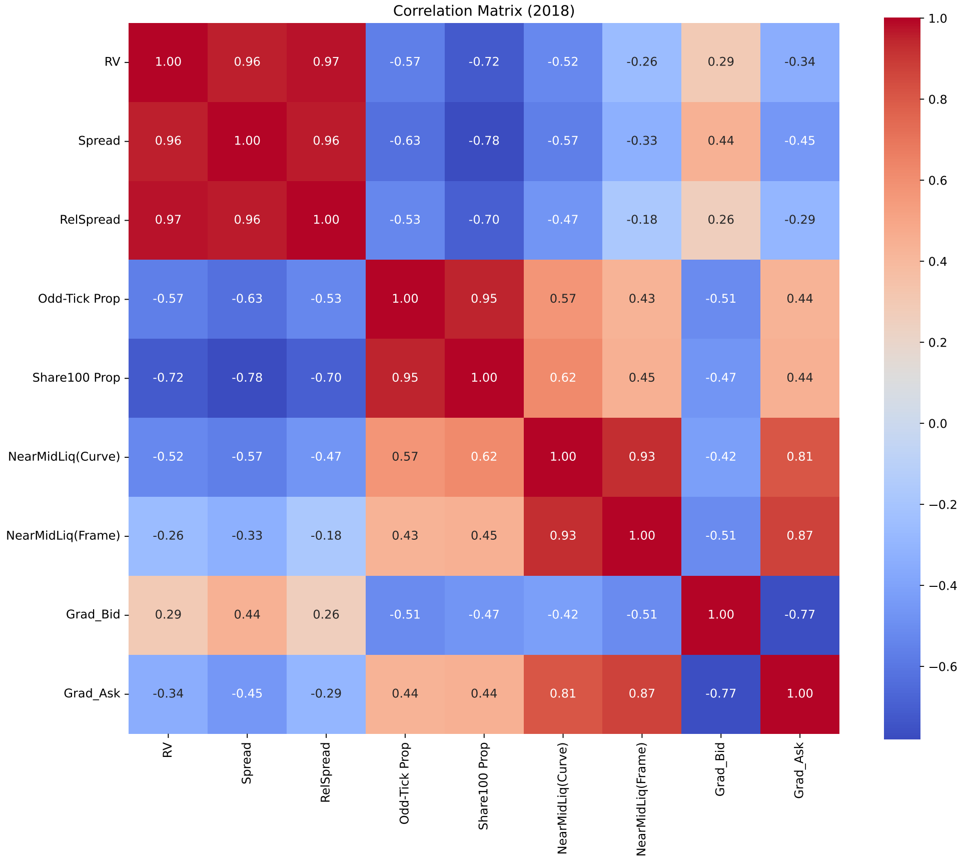
Second, the measures related to order book depth exhibit a distinct structure. The proportion of volume posted at the best quotes (Share100 prop and Odd-tick prop) is strongly negatively correlated with volatility and spreads. This aligns with the intuition that a thinner order book near the best quotes corresponds to higher trading costs and greater price variability. Conversely, when there is more volume concentrated at the top of the book, markets tend to be more stable and liquid.

Third, the liquidity measures around the mid-quote show a strong negative association with volatility and spreads. Markets where a larger fraction of volume is concentrated near the mid tend to display tighter spreads and more stable trading conditions. This pattern reflects the intuition that when liquidity is available close to the mid-price, the cost of immediate execution decreases and prices are less prone to sharp fluctuations. Conversely, when liquidity near the mid is sparse, execution becomes more costly and the market is more vulnerable to transient shocks.

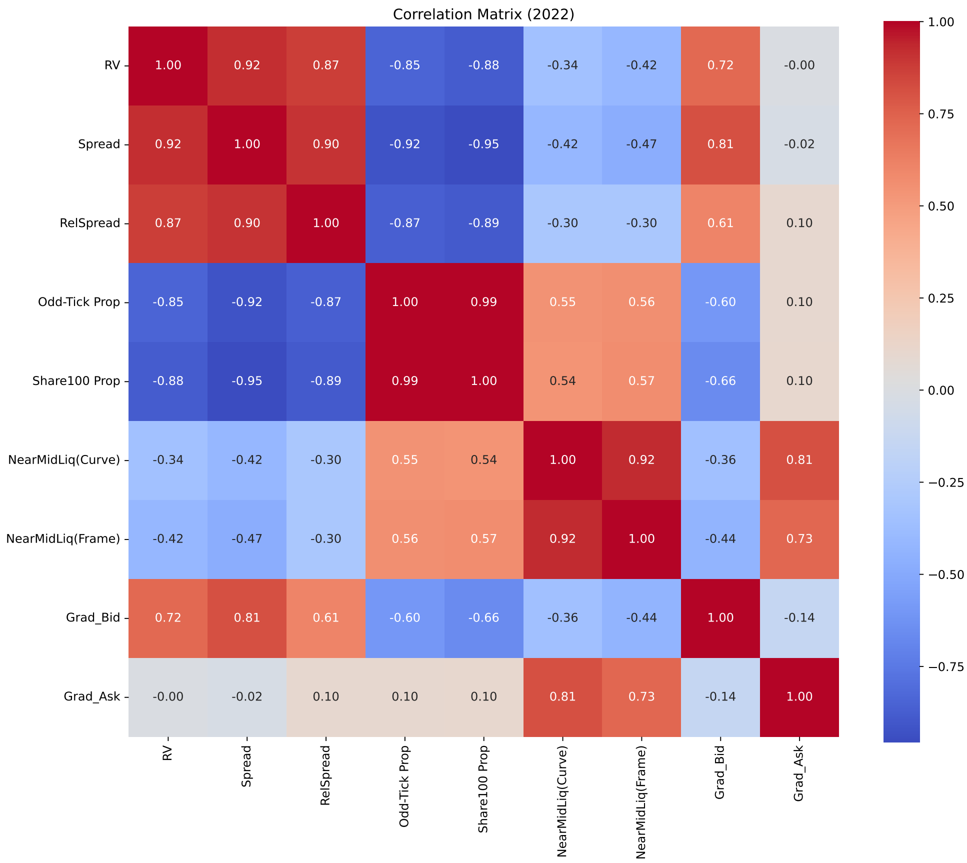
Fourth, the gradient of the order book, which captures the rate at which available volume accumulates as prices move away from the best quotes, is positively related to market stability and negatively associated with volatility. A steeper gradient implies that additional liquidity becomes available quickly as prices shift, creating stronger resistance against large price moves and reducing trading frictions. In contrast, a flatter gradient corresponds to a thinner liquidity profile, under which small trades may induce outsized price changes, resulting in higher volatility and instability. Moreover, this empirical finding is consistent with theoretical results from PDE models of limit order books, which imply that price impact is inversely related to the available volume at the mid-price and to the steepness of the depth profile in its vicinity. In other words, larger gradients around the mid-price correspond to smaller price impacts and lower realized volatility, whereas flatter gradients imply thinner liquidity and heightened volatility.

It is worth noting, however, that the ask-side gradient in Figure 3.15(b) exhibits some anomalies. In principle, one would expect the bid and ask gradients to be strongly negatively correlated, since they should move in opposite directions. The near-zero correlation observed here arises from the way gradients were estimated: as discussed in Section 3.4, we employ a second-order scheme on unmerged curves, and the alternating high–low oscillations in the raw profiles significantly distorted the estimates.

Taken together, these findings complement the earlier patterns by highlighting that both the immediate availability of liquidity near the mid-price and the overall slope of the order book are central to understanding market resilience. They demonstrate that microstructure variables jointly capture different but interconnected dimensions of liquidity provision, and price volatility.



(a) Correlation matrix in 2018.



(b) Correlation matrix in 2022.

Figure 3.15: Correlation matrices in 2018 and 2022.

Chapter 4

PDE Modeling Framework for Limit Order Books

This chapter constructs a continuous PDE model to describe the dynamic evolution of LOB. The model uses the order volume density as the core variable, and the diffusion term, mean-reversion term, and source term are used to capture order repositioning, cancellation, and submission behaviors, respectively.

4.1 Model Assumptions

We consider a continuous-time model for the LOB with infinitesimal tick size and order size, incorporating both limit and market orders. The state of the LOB at time $t \geq 0$ is represented by the real-valued function $u_t(x)$, defined on $\mathbb{R}$, together with the mid-price S_t. For $x < S_t$, $u_t(x)$ measures the buy-side order volume at price x, while for $x > S_t$ it measures the sell-side order volume. This continuous formulation can be viewed as the scaling limit of a discrete order book: as the tick size tends to zero and the frequency of order arrivals increases, the discrete volumes at ticks $x_1, x_2, \dots$ converge to a continuous volume profile. Within this framework, the volume posted at a discrete tick x_i corresponds to the area under $u_t(x)$ between x_i and x_{i+1}. The modeling framework focuses on deep-order interactions up to level 20 in the LOB, as reconstructed from the LOBSTER database. This depth range captures both near-term and far-term liquidity, enabling an analysis of the stability of the order book shape and its sensitivity to different market conditions.

It should be emphasized that this type of PDE model is best interpreted as describing the average shape of the order book over a given time interval. To capture dynamics on shorter time scales, one would need to include an additional stochastic noise term in the PDE, as in the SPDE approaches of Cont and Müller [7] and Hambly et al.[9]. However, incorporating such fluctuations is mathematically more involved, so in this work we restrict attention to the deterministic PDE formulation.

4.2 Mathematical Formulation

In the absolute price coordinate system, the function $u_t(x)$ is assumed to satisfy the following PDE for $x \in \mathbb{R} \setminus \{S_t\}$:

$$\partial_t u_t(x) = \alpha \Delta u_t(x) - \theta u_t(x) + \lambda\big(x - S_t\big), \qquad x \in \mathbb{R} \setminus \{S_t\}. \tag{4.2.1}$$

Here, S_t denotes the mid-price. Importantly, since the mid-price evolves randomly over time and in general follows a semimartingale path, the boundary $\{S_t\}$ is itself stochastic.

As a result, equation (4.2.1) should be interpreted not as a purely deterministic PDE, but as a stochastic PDE driven by the moving "interface" S_t.

This continuous formulation can be understood as the scaling limit of the discrete order book: as the tick size tends to zero and the frequency of order arrivals grows, the stepwise volume profile of discrete ticks converges to a continuous curve. In this formulation, the volume posted at a discrete tick x_i corresponds to the area under $u_t(x)$ between x_i and x_{i+1}.

The diffusion term $\alpha\Delta u_t(x)$, where Δ is the Laplacian, can be motivated from the discrete order book picture. In the discrete setting, the Laplacian corresponds to the second-order finite difference across neighbouring price ticks. Economically, this term captures the effect of traders cancelling their existing orders at one price level and repositioning them at nearby price ticks. In other words, it reflects the natural diffusion of liquidity across price levels: orders posted at a given tick are not permanently fixed there, but can migrate to slightly more competitive or conservative price levels depending on market conditions.

At the same time, one should also recognize that the Laplacian is often included in continuous LOB models partly for technical and smoothing purposes. From a purely empirical perspective, the precise mechanism of order relocation may be more complex than simple diffusion. However, the Laplacian provides a mathematically tractable way to smooth the volume profile and to ensure regularity of solutions, which greatly facilitates the analysis of PDE or SPDE formulations. This is one reason why essentially all continuous LOB models in the literature—such as those of Cont and Müller [7] and Hambly et al.[9] feature a diffusion term of this form.

The mean-reversion term $-\theta u_t(x)$ captures the effects of order cancellations, pulling the LOB towards a stationary profile. The source term $\lambda(x - S_t)$ represents the arrival rate of new limit orders as a function of the distance from the mid-price, linking liquidity provision to the underlying supply–demand structure.

The dynamics of the mid-price S_t require careful specification, since the PDE formulation for $u_t(x)$ only applies away from the mid-price. There are several possible modeling choices. One option is to treat S_t as an exogenous process, either of finite variation (for example, a piecewise deterministic path) or as a diffusion such as $S_t = S_0 + \mu t + \sigma W_t$, where W_t is a standard Brownian motion. This corresponds to modeling the order book conditional on externally specified price dynamics. While such a specification implies that the price could in principle become negative, this outcome is unlikely over short timescales if σ is sufficiently small. A common variation is to instead adopt a geometric Brownian motion specification, which preserves the positivity of the price process.

An alternative is to endogenize, at least partially, the dynamics of S_t, tying price evolution to the imbalance of order flow at the boundary. In such price impact models, a market order of size q shifts the mid-price until the consumed volume under the order book profile matches q.

Our modeling choice follows the former strand: we treat S_t as exogenous, which simplifies the PDE formulation and makes the analysis more tractable. By contrast, models that derive S_t endogenously, while more realistic, require delicate treatment of the order flow–price feedback loop and significantly complicate the analysis. In the related literature, an alternative approach is to derive S_t from market orders, where shifts in the relative buy and sell volumes induce movements in the mid-price. In the continuous limit, this imbalance-based representation naturally leads to a semimartingale dynamic for S_t, as discussed for instance in Cont and Müller [7]. Yet another strand of work considers deterministic formulations: Horst and Paulsen [11] establish a law of large numbers limit in which the discrete order book converges to a coupled ODE–PDE system, and the mid-price (or best bid/ask) evolves according to deterministic boundary conditions.

4.3 Motivation for Relative Coordinates

In the absolute coordinate system, the mid-price S_t moves over time, so any grid defined in absolute prices effectively sits on a shifting computational domain. As the domain drifts, the grid ceases to be aligned with the instantaneous mid, the local resolution in the vicinity of S_t becomes inconsistent, and maintaining numerical stability near the mid is more delicate. In addition, when prices are expressed absolutely, the bid and ask sides are not represented symmetrically, which complicates the design and analysis of discretization schemes that should treat both sides consistently. Re-centering the problem in a coordinate system anchored at the mid-price addresses these issues: the domain becomes fixed in the moving frame, the discretization near S_t remains stable and uniform, and the bid/ask distinction is explicit and symmetric. This change simplifies both the analysis and the numerical implementation.

4.4 Relative Coordinate System and Redefined Functions

We introduce the relative coordinate

$$y = x - S_t,$$

which centers the frame at the mid-price. In this coordinate, $y < 0$ denotes prices below the mid (bid side) and $y > 0$ denotes prices above the mid (ask side). We write two side-specific depth functions as $u_t^b(y)$ for $y < 0$ and $u_t^a(y)$ for $y > 0$, with domains $y \in [-L_b, 0]$ and $y \in (0, L_a]$, respectively. For analysis and numerics it is convenient to combine them into a single field in the moving frame,

$$\tilde{u}_t(y) \equiv u_t(x) = u_t(y + S_t),$$

so that $\tilde{u}_t(y) = u_t^b(y)$ for $y < 0$ and $\tilde{u}_t(y) = u_t^a(y)$ for $y > 0$; the interface at $y = 0$ will be handled by boundary conditions specified later.

4.5 PDE Transformation to Relative Coordinates

Applying Itô's formula to $\tilde{u}_t(y) = u_t(y + S_t)$ gives

$$d\tilde{u}_t(y) = \partial_t u_t(y + S_t)\, dt + \partial_x u_t(y + S_t)\, dS_t + \tfrac{1}{2}\partial_{xx} u_t(y + S_t)\, d\langle S\rangle_t.$$

Substituting the original PDE in absolute coordinates,

$$\partial_t u_t(x) = \alpha\, \partial_{xx} u_t(x) - \theta\, u_t(x) + \lambda(x - S_t),$$

and noting that $\lambda(x - S_t) = \lambda(y)$ in the moving frame, we obtain

$$d\tilde{u}_t(y) = \Big[\alpha\, \partial_{yy}\tilde{u}_t(y) - \theta\tilde{u}_t(y) + \lambda(y)\Big] dt + \partial_y \tilde{u}_t(y)\, dS_t + \tfrac{1}{2}\partial_{yy}\tilde{u}_t(y)\, d\langle S\rangle_t.$$

Assuming the mid-price follows an Itô process $dS_t = \mu\, dt + \sigma\, dW_t$ with quadratic variation $d\langle S\rangle_t = \sigma^2\, dt$, the dynamics in relative coordinates become

$$d\tilde{u}_t(y) = \Big[\alpha\, \partial_{yy}\tilde{u}_t(y) - \theta\tilde{u}_t(y) + \lambda(y) + \mu\, \partial_y\tilde{u}_t(y) + \tfrac{1}{2}\sigma^2\partial_{yy}\tilde{u}_t(y)\Big] dt + \sigma\, \partial_y\tilde{u}_t(y)\, dW_t,$$

often written compactly as

$$d\tilde{u}_t(y) = \Big[\big(\alpha + \tfrac{1}{2}\sigma^2\big)\partial_{yy}\tilde{u}_t(y) - \theta\tilde{u}_t(y) + \lambda(y) + \mu\, \partial_y\tilde{u}_t(y)\Big] dt + \sigma\, \partial_y\tilde{u}_t(y)\, dW_t, \tag{4.5.1}$$

4.6 Parameter Definitions and Economic Interpretation

The coefficient α determines the speed at which the order book profile smooths out due to order reallocation. The parameter θ governs the rate of order removal. The function $\lambda(\cdot)$ specifies the rate at which new orders arrive at a given relative price, thereby shaping the replenishment of liquidity across price levels. Together, these parameters control the dynamic behavior of the LOB and determine its response to market events.

4.7 Boundary Condition Selection

To properly motivate these boundary conditions, it is helpful to recall their discrete counterparts in the approximation of the Laplacian. In a finite-difference scheme for the order book, the second derivative $\Delta u_t(x)$ is approximated by differences of neighboring volumes. Imposing a Dirichlet condition $u_t(S_t^{\pm}) = 0$ corresponds to forcing the boundary nodes at the best bid and ask to be set to zero, i.e. all volume at the best prices is immediately consumed, leaving no residual liquidity. By contrast, a Neumann condition $\partial_x u_t(S_t^{\pm}) = 0$ corresponds in the discrete picture to setting the finite-difference gradient at the boundary to zero, so that the profile flattens out near the mid-price, representing a stable liquidity distribution with balanced supply and demand. More generally, a Robin condition $u_t(S_t^{\pm}) = \mu \partial_x u_t(S_t^{\pm})$ interpolates between these two extremes, modeling a proportional relationship between boundary volume and its slope, and can be interpreted as a market in which liquidity supply responds smoothly to marginal changes in price. The empirical gradient patterns extracted from LOBSTER level-20 data provide the basis for determining which boundary condition best represents the observed market behavior.

Chapter 5

Numerical Discretization and Parameter Estimation

Building on the stochastic PDE formulation developed in Chapter 4, this chapter focuses on the numerical discretization and parameter estimation of the order-volume dynamics. In particular, the relative-coordinate transformation introduced in Section 4.4–4.5 yields the SPDE (4.5.1), where the moving mid-price induces both deterministic drift and stochastic fluctuations in the order book density. This representation highlights the intrinsic randomness of liquidity evolution around the mid-price and provides a natural starting point for simulation and calibration.

We now operationalize this SPDE framework by discretizing it in both space and time, aligning the computational grid with the quoting conventions used in the empirical construction, and incorporating boundary conditions informed by observed near-mid gradients. The resulting scheme enables us to link the model parameters directly to statistics computed from LOBSTER data.

5.1 Spatial and Temporal Discretization

To simulate the dynamics implied by the relative-coordinate SPDE (4.5.1), we discretize the moving-frame coordinate y on a uniform one-dimensional grid with spacing h. On the bid side we set $y_i = -L_b + ih$ for $i = 0, 1, \ldots, N_b$, and on the ask side we set $y_i = ih$ for $i = 1, \ldots, N_a$. Time is sampled on nodes t_n with step Δt, and the numerical state is $u_i^n \approx \tilde{u}_{t_n}(y_i)$. Spatial derivatives are approximated by centered differences,

$$\partial_{yy}\tilde{u}_t(y_i) \approx \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2}, \qquad \partial_y \tilde{u}_t(y_i) \approx \frac{u_{i+1}^n - u_{i-1}^n}{2h}.$$

For the stochastic time evolution corresponding to the moving-frame SPDE, we use an explicit Euler update,

$$\begin{aligned} u_i^{n+1} = u_i^n + \Delta t \left[\left(\alpha + \tfrac{1}{2}\sigma^2\right) \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2} - \theta u_i^n + \lambda(y_i) + \mu \frac{u_{i+1}^n - u_{i-1}^n}{2h} \right] \\ + \sigma \frac{u_{i+1}^n - u_{i-1}^n}{2h} \Delta W_n, \end{aligned} \tag{5.1.1}$$

where $\Delta W_n \sim \mathcal{N}(0, \Delta t)$ represents the Brownian increment over $[t_n, t_{n+1}]$. Boundary nodes at the mid-interface and at the far ends of the domain are imposed according to the chosen boundary conditions.

5.2 Parameter Estimation

We estimate the parameters

$$\Theta = \{\alpha, \theta, \sigma, \mu, \lambda(\cdot), L_b, L_a, N_b, N_a, \Delta t\}$$

in two stages. Stage I produces data-driven initial values for $(\sigma, \mu, \lambda(\cdot), \theta)$. Stage II jointly refines (α, θ) (optionally also μ, σ) by calibrating the PDE-implied depth profiles against their empirical counterparts, using a curve-matching criterion that aligns the simulated and observed profiles. The spatial domains $([-L_b, 0] \cup [0, L_a])$, grid sizes (N_b, N_a), and time step Δt are chosen to satisfy explicit-scheme stability constraints.

5.2.1 Data-Driven Initial Values

The volatility parameter σ could in principle be set using the quadratic variation of the mid-price process. In practice, however, this choice leads to instability in the discretization scheme, as it effectively increases the coefficient in front of the Laplacian term. To ensure stability, we therefore adopt a relatively small fixed value of σ, chosen for tractability rather than exact calibration.

For the drift parameter μ, direct estimation from data is generally not feasible, as reliable inference of drift requires far longer time horizons than are available in our intraday setting. Consequently, we do not attempt to estimate μ from data. Instead, we adopt the simplifying choice of setting $\mu = 0$ in the baseline specification, and, where appropriate, explore the impact of nonzero μ values in sensitivity checks. This approach allows us to isolate the role of drift in shaping the volume curve without over-interpreting noisy empirical estimates.

The construction of the submission intensity function $\lambda(y)$ has already been introduced in Section 3.3, where flow curves are derived from message data. For calibration purposes, we adopt the flow curves from the previous trading day. While this is a simplified choice, in practical applications more refined estimation methods could be implemented.

In our calibration procedure, we approximate $\lambda(y)$ by connecting points of the empirical flow curve with straight lines. We truncate the curve at -2.0 and 2.0 in relative price units, since beyond this range the empirical data contain almost no additional information. We then enforce the boundary conditions $\lambda(\pm L_{a/b}) = 0$ by linear extrapolation to the endpoints. This ensures a smooth, non-negative intensity profile on the PDE grid while remaining computationally tractable, and the truncation does not materially affect the calibration outcome.

For the cancellation or decay parameter, we make use of the deleted-order flow curve introduced in Section 3.3. This curve provides a natural scale for order removal intensity. By comparing the magnitude of deleted flows with the average standing volume $u(y)$, we obtain an effective estimate of θ. This approach delivers a practical calibration scheme that captures the rate of order cancellation consistent with the observed data.

5.2.2 Joint Calibration via Curve Matching

Having fixed the data-driven parameters $(\sigma, \mu, \lambda(\cdot), \theta)$ as described in the previous subsection, we now turn to the calibration of the remaining free parameter α. In this step, the objective is to adjust α such that the simulated depth curves generated by the SPDE scheme exhibit a reasonable match to the empirical profiles observed in the data.

We begin by examining the case of AAPL, where the empirical evidence points to a compact liquidity profile with steep gradients around the mid-price. In line with the discussion of Section 4.7, such profiles are best captured by Dirichlet- or Robin-type boundary conditions, which enforce either vanishing residual liquidity at the best quotes or a proportional relationship between boundary slope and volume. Under these specifications, and with a candidate value of $\alpha = 0.0002$, the simulated order book reproduces the strongly peaked shape of the empirical curves. Figures 5.1–5.3 illustrate this comparison: although the simulated and empirical curves are not identical in detail, they agree in both order of magnitude and overall structure. Such discrepancies are expected, since the model necessarily incorporates randomness through the Brownian increment, as well as simplifications in the specification of the dynamics. Nevertheless, the comparison illustrates that the calibration procedure yields a plausible representation of the order book density.

A more careful inspection reveals that the simulated profiles tend to be more sharply peaked, whereas the empirical depth curves are somewhat more dispersed around the mid-price. This indicates that the current specification, while effective at reproducing near-mid liquidity concentration, may oversimplify the mechanisms that govern volume distribution at wider price levels. One possible refinement is to relax the assumption of a uniform cancellation term $-\theta u$, and instead allow the deletion intensity to vary with distance from the mid-price, $\theta(x - S_t)$. This position-dependent specification parallels the treatment of new order arrivals and may better capture the heterogeneous persistence of orders across price levels. We explore this extension in the following section.

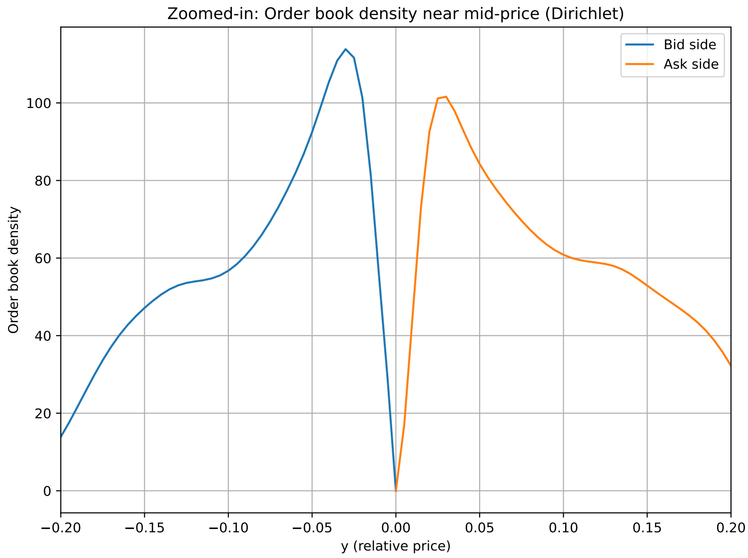


Figure 5.1: Zoomed-in view of the order book density near the mid-price (AAPL, Dirichlet, Uniform deletion).

By contrast, when we turn to TSLA, the empirical curves show a much flatter distribution around the mid-price, with liquidity spread more evenly across ticks and gradients that are close to zero. Such behavior is more naturally represented by a Neumann-type boundary condition, which corresponds to a vanishing slope at the mid. Figures 5.4–5.5 demonstrate that under this specification, the model generates order book shapes broadly consistent with the observed shallow profile. In the empirical depth curve, some unusually high spikes appear, which can be attributed to noise in the data rather than genuine concentration of liquidity. At the same time, Figure 5.4 highlights a limitation of the Neumann specification: while the gradient at the boundary is correctly set to zero, the absolute level of the boundary value is still relatively large, which does not fully capture the weak liquidity typically observed at the best quotes for TSLA.

Figure 5.2: Zoomed-in view of the order book density near the mid-price (AAPL, Robin, Uniform deletion).

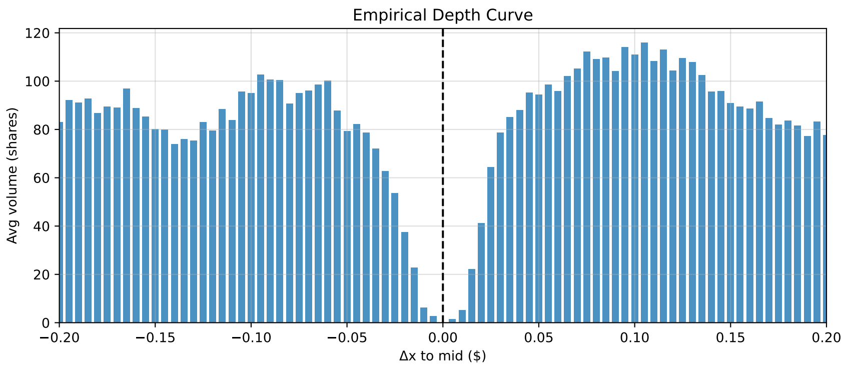


Figure 5.3: Empirical depth curve for AAPL on 6 March 2025. The histogram shows the average standing volume across relative price levels with respect to the mid-price.

5.2.3 Extension: Position-Dependent Cancellation Intensity

In Chapter 4, the cancellation mechanism was modeled through a uniform decay term of the form $-\theta u$, implying a constant proportional rate of order removal across all price levels. While this specification provides a convenient baseline, it does not account for the empirical observation that cancellation activity tends to vary with distance from the mid-price. To address this limitation, we now consider an alternative formulation in which the deletion intensity is allowed to depend on the relative price, in close analogy to the specification of new order arrivals. Concretely, we replace the uniform decay term by a position-dependent function $\theta(x - S_t)$, so that cancellations are no longer homogeneous across the book. This modification preserves the tractability of the PDE system while providing additional flexibility to capture the spatial structure of order cancellations observed in the data. Although in most cases the net insertion–deletion rate is positive, it can in principle take negative values. To ensure that the volume curve remains non-negative, we take the positive part of this term.

In the following, we illustrate the impact of this modification by comparing simulated

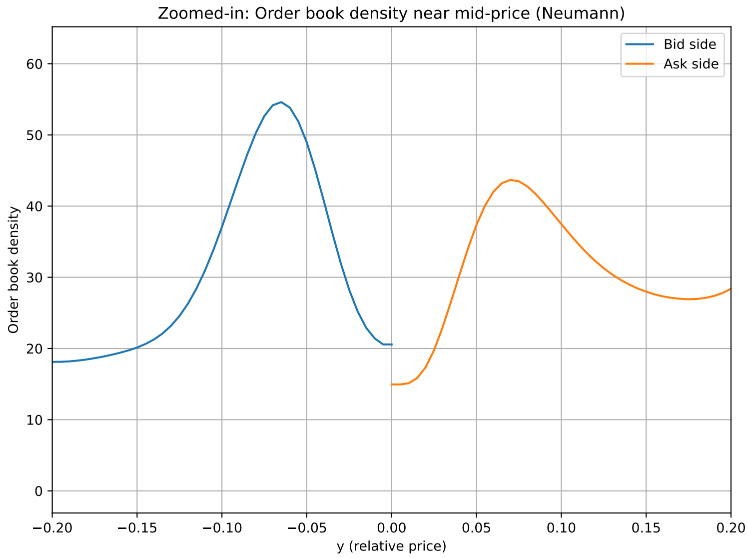


Figure 5.4: Zoomed-in view of the order book density near the mid-price (TSLA, Neumann, Uniform deletion).

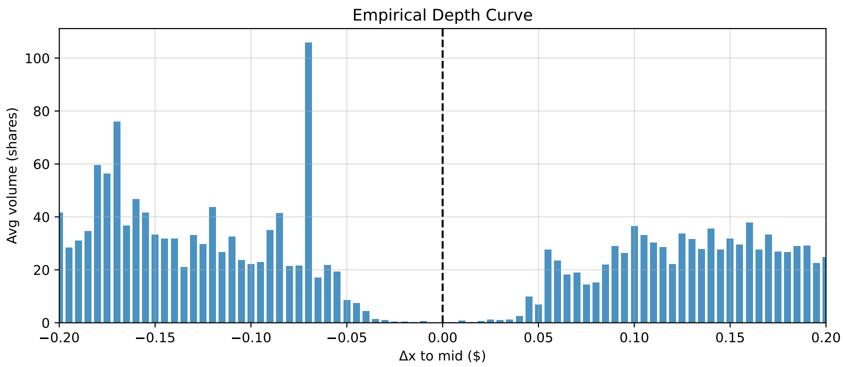


Figure 5.5: Empirical depth curve for TSLA on 6 March 2025. The histogram shows the average standing volume across relative price levels with respect to the mid-price.

profiles under the position-dependent cancellation specification with the empirical depth curves.

Figures 5.6 and 5.7 display the simulated order book profiles under the position-dependent cancellation specification. Compared with the uniform-deletion case, the very sharp peak around the mid-price is no longer present, and the simulated curves more closely resemble the empirical depth curve in Figure 5.3. This indicates that allowing cancellations to vary with distance from the mid-price provides a more realistic representation of the volume distribution across price levels.

It should be noted, however, that the LOBSTER data provide depth only up to 20 levels on each side of the book, which corresponds roughly to the interval $[-0.2, 0.2]$ in relative price units. As a result, the standing volume outside this range is systematically underestimated, since orders beyond the 20th level are not recorded. This truncation implies that while mass can leave the observed window, no additional mass can enter from outside, leading to potential underestimation of volume near the boundaries. With access to deeper order book data, this boundary effect would likely be alleviated.

In the case of TSLA, a comparison between Figure 5.8 and Figure 5.4 reveals that introducing a position-dependent cancellation mechanism effectively mitigates the sharp

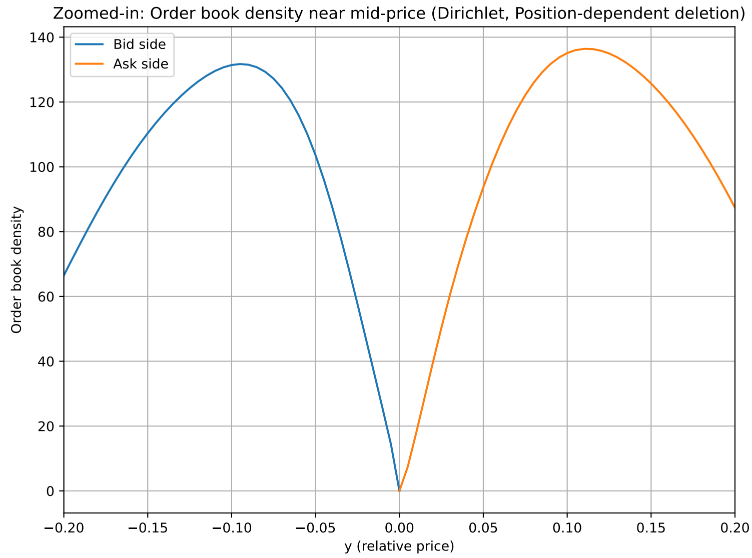


Figure 5.6: Zoomed-in view of the order book density near the mid-price (AAPL, Dirichlet, Position-dependent deletion).

Figure 5.7: Zoomed-in view of the order book density near the mid-price (AAPL, Robin, Position-dependent deletion).

peak near the mid-price observed in Figure 5.5. The resulting profile in Figure 5.8 is smoother and overall more consistent with the empirical depth curve in Figure 5.5. However, it is worth noting that under the Neumann boundary condition, the absolute level of the boundary value remains relatively large, which does not fully capture the weak liquidity typically observed at the best quotes for TSLA. This issue is not resolved by the modification.

When we turn to the simulation under the Dirichlet boundary condition (Figure 5.9), we find that, compared with the Neumann specification, the Dirichlet case avoids the excessive liquidity accumulation near the mid-price and therefore better reflects the empirical features of TSLA. It should be noted, however, that on the bid side the gradient is not perfectly flat, indicating a certain deviation. Nevertheless, this local imperfection does not diminish the overall advantage of the Dirichlet specification: it more faithfully reproduces the key empirical fact that liquidity is weak in the vicinity of the mid-price.

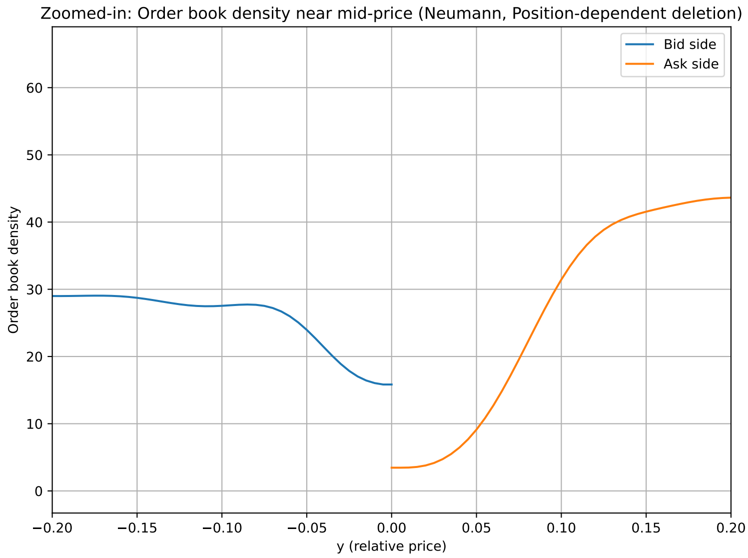


Figure 5.8: Zoomed-in view of the order book density near the mid-price (TSLA, Neumann, Position-dependent deletion).

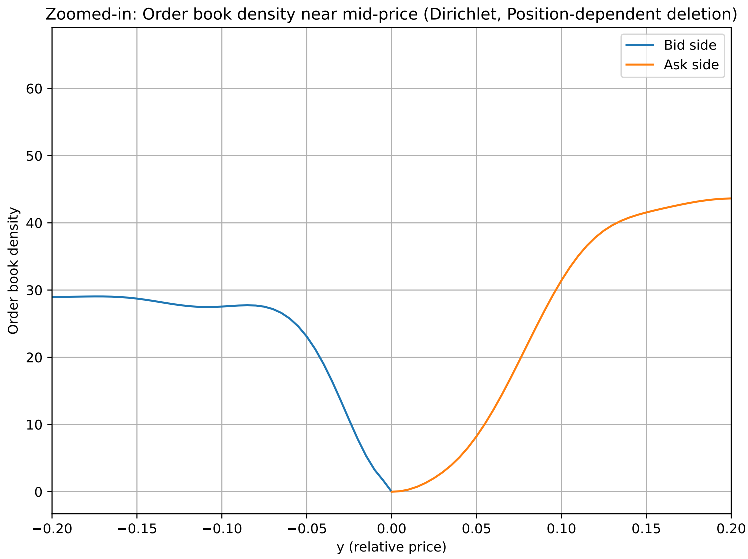


Figure 5.9: Zoomed-in view of the order book density near the mid-price (TSLA, Dirichlet, Position-dependent deletion).

Chapter 6

Price Impact of Market Orders

6.1 Mechanics of Price Impact under a Linear Depth Profile

In order to understand the influence of market orders within the framework of PDE models for the LOB, it is necessary to specify suitable boundary conditions at the mid-price. In this analysis, the PDE governing the order book dynamics is equipped with Neumann boundary conditions, which require that the spatial derivatives of the order book volume vanish at the mid-price. This condition ensures local smoothness of the volume profile around the execution point. At the same time, the mid-price itself is modeled as a diffusion, that is, a Brownian motion with volatility parameter σ, thereby capturing the continuous random fluctuations induced by the activity of market participants.

Within this environment, the introduction of a market order has a direct impact on the mid-price. A market buy order of size $q > 0$ submitted at time t requires a shift in the mid-price such that the consumed volume between the original price S_t and the new price S_{t+} matches the order size. Formally, this relationship is expressed as

$$q = \int_{S_t}^{S_{t+}} u_t(x)\, dx,$$

where $u_t(x)$ represents the available order book volume at price level x. For relatively small orders, we expand the integral around the mid-price. Because of the Neumann boundary condition imposed at S_t, the first derivative of the order book profile vanishes at the mid-price. This yields

$$q \approx u_t(S_t^+)\Delta S_t,$$

which is a second-order approximation.

In particular, the price change is approximately proportional to the order size and inversely proportional to the available volume at the best ask,

$$\Delta S_t = \frac{q}{u_t(S_t^+)}. \tag{6.1.1}$$

A natural extension of this result arises when the trader refrains from submitting a single large order but instead splits the trade into multiple smaller increments spread over time. Suppose that the execution occurs across a finite horizon T, where the total order size is subdivided into smaller units posted at discrete times. As the subdivision becomes finer, the cumulative price trajectory can be described by

$$S_t = S_0 + \sigma B_t + \int_0^t \frac{q_r}{u_r(S_r^+)}\, dr,$$

where q_r denotes the instantaneous trading rate at time r. This formulation highlights the central mechanism of market impact: the expected deviation of the price from its initial value is proportional to the trading speed, yet moderated by the liquidity available at the best ask. In other words, faster execution magnifies impact, while deeper liquidity cushions it.

To gain further insights, let us consider the situation where the order book profile remains approximately constant throughout the execution. Specifically, assume that

$$u_t(x) = u_0(x) \quad \text{for } x \geq S_t,$$

and that the initial profile can be approximated by a linear function,

$$u_0(x) \approx a + b(x - S_0),$$

with parameters $a > 0, b > 0$. In the absence of volatility ($\sigma = 0$), the price displacement satisfies

$$S_t - S_0 \;\approx\; \int_0^t \frac{q_r}{a + b(S_r - S_0)}\, dr. \tag{6.1.2}$$

Rearranging and differentiating yields

$$a\dot{S}_t + \frac{b}{2}\frac{d}{dt}|S_t|^2 = \dot{S}_t(a + bS_t) = q_t.$$

Integrating both sides gives

$$a(S_t - S_0) + \frac{b}{2}(S_t - S_0)^2 = Q_t,$$

where $Q_t = \int_0^t q_s\, ds$ is the cumulative executed order size.

Solving for the price displacement yields

$$S_t - S_0 \;=\; -\frac{a}{b} \;+\; \left(\frac{a^2}{b^2} \;+\; \frac{2}{b}Q_t\right)^{1/2}. \tag{6.1.3}$$

This expression demonstrates that, under the assumption of a linearly growing order book around the best ask, the price impact exhibits a square-root law: the deviation of the mid-price from its initial level grows with the square root of the cumulative traded volume. Such a concave relationship is consistent with empirical observations in financial markets, where the marginal impact of additional volume diminishes as order size increases.

It is important to recognize that the validity of this square-root law depends crucially on the assumption of a stationary and linear order book profile. For small to moderate orders, this approximation may be reasonable, but for larger executions the order book dynamics play a central role. In reality, the replenishment of liquidity during execution and the adaptive strategies of liquidity providers ensure that the volume profile does not remain static. To achieve a realistic description of market impact, it is therefore necessary to incorporate these dynamic features into the PDE framework, extending beyond the simplified constant-profile assumption.

6.2 Numerical Illustration: Constant-Speed Execution

This section complements the analytical development in 6.1 by providing a compact numerical experiment that visualizes the impact generated by a constant-rate metaorder in the PDE LOB framework with Neumann boundary conditions. The exercise isolates the core mechanism established above: the instantaneous drift of the mid-price induced by a buy program is proportional to the trading speed and inversely proportional to the available depth at the touch, so that faster execution amplifies impact while thicker near-touch liquidity cushions it. See the impact–speed relation in equation (6.1.1), which provides the continuous-time formulation used here.

For the benchmark simulation we adopt the stationary linear near-touch profile introduced in 6.1 and used to obtain the closed-form law, namely

$$u_0(x) = a + b(x - S_0), \quad a > 0, \ b > 0.$$

With a constant execution rate q and in the volatility-free limit, the cumulative displacement ΔS_t implied by the integral representation (6.1.2) satisfies the quadratic relation (6.1.3) between ΔS_t and the traded volume Q_t; solving yields the square-root impact function reported earlier. In consequence, trajectories plotted in the time domain display steeper growth for larger q because volume accumulates more rapidly; this is visible in Figure 6.1. When the same trajectories are plotted against Q, they collapse onto the single concave curve predicted by the closed-form expression, as shown in Figure 6.2. These two behaviors—time-domain separation by speed and volume-domain collapse—are precisely those emphasized by the analytical result.

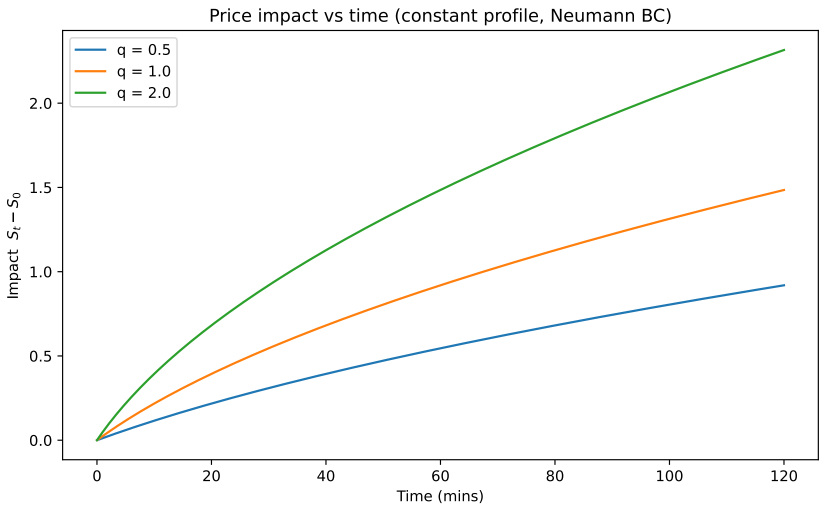


Figure 6.1: Price impact vs time under constant trading speed (stationary linear profile, Neumann boundary conditions). Curves correspond to $q \in \{0.5, 1.0, 2.0\}$. The steeper time paths for larger q reflect the impact–speed relation in equation (6.1.1).

Numerically, we integrate the drift appearing in the cumulative-impact representation with an explicit Euler scheme on a uniform grid. In the stationary benchmark the intercept a remains fixed at its initial value and the slope b controls how quickly depth thickens as ΔS_t grows; the latter determines the concavity of the ΔS–Q curve and the extent to which the market resists further displacement as execution proceeds. When plotted against traded volume, the simulated paths align closely with the closed-form square-root law derived from (6.1.2)–(6.1.3), see Figure 6.2, which also serves as an internal consistency check for the discretization and parameterization.

Parameter interpretation and calibration are transparent. The intercept a summarizes baseline depth at the best quote and can be proxied by near-touch volume extracted from the smoothed depth curves; the slope b corresponds to the empirical near-mid gradient estimated in Chapter 3 with one-sided second-order differences or local polynomial fits. Using these diagnostics to set (a, b) produces a data-driven baseline for impact that is immediately comparable across assets and dates and keeps the numerical illustration tightly coupled to the empirical boundary analysis developed earlier.

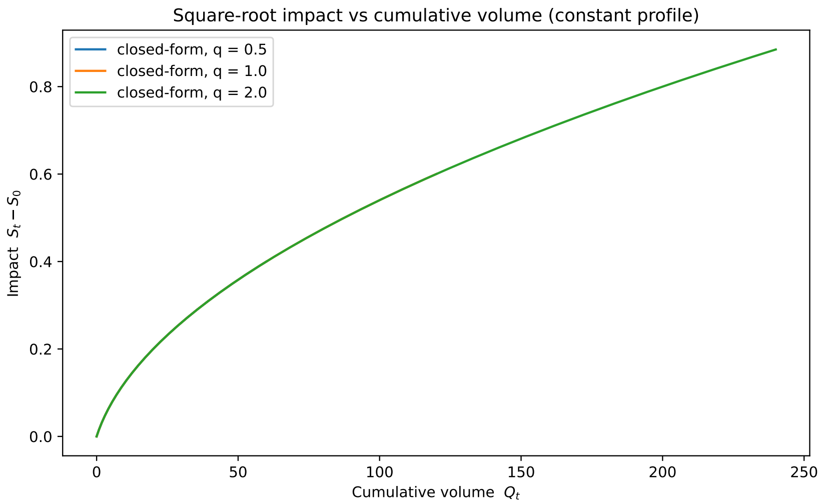


Figure 6.2: Price impact vs cumulative volume Q_t under constant trading speed. The curves for different q collapse onto the closed-form square-root law implied by (6.1.2)–(6.1.3).

The limitations of this baseline should be explicit. The stationary linear profile is an approximation that is most defensible for moderate sizes and short windows; large programs and stressed conditions typically trigger dynamic replenishment and cancellations that reshape the book during execution.

6.3 Joint System and Finite-Difference Scheme with Trading Impact

From (4.5.1) we see that when the price follows the Itô process $dS_t = \mu\, dt + \sigma\, dW_t$, the dynamics of $\tilde{u}_t(y)$ in relative coordinates are

$$d\tilde{u}_t(y) = \Big[(\alpha + \tfrac{1}{2}\sigma^2)\partial_{yy}\tilde{u}_t(y) - \theta\tilde{u}_t(y) + \lambda(y) + \mu\,\partial_y\tilde{u}_t(y)\Big]dt + \sigma\,\partial_y\tilde{u}_t(y)\,dW_t.$$

If a trader executes at speed q, then the price process follows the SDE

$$dS_t = \frac{q}{u_t(S_t^+)}dt + \sigma dW_t = \frac{q}{\tilde{u}_t(0^+)}dt + \sigma dW_t.$$

Plugging this into (4.5.1), we obtain the following *joint system*:

$$\begin{aligned} d\tilde{u}_t(y) &= \Big[(\alpha + \tfrac{1}{2}\sigma^2)\partial_{yy}\tilde{u}_t(y) - \theta\tilde{u}_t(y) + \lambda(y) + \tfrac{q}{\tilde{u}_t(0^+)}\partial_y\tilde{u}_t(y)\Big]dt + \sigma\partial_y\tilde{u}_t(y)\,dW_t, \\ dS_t &= \frac{q}{\tilde{u}_t(0^+)}dt + \sigma dW_t. \end{aligned} \tag{6.3.1}$$

Compared with (4.5.1), the drift now includes the *trading impact term* $\frac{q}{\tilde{u}_t(0^+)}\partial_y\tilde{u}_t(y)$, and the drift of the price process changes from μ to $\frac{q}{\tilde{u}_t(0^+)}$.

For the finite-difference discretization, we impose the Neumann boundary condition $u_0^n = u_1^n$. Discretizing the joint system (6.3.1) with an Euler–Maruyama scheme yields

$$\begin{aligned} u_i^{n+1} &= u_i^n + \Delta t\left[(\alpha + \tfrac{1}{2}\sigma^2)\,\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2} - \theta u_i^n + \lambda(y_i) + \tfrac{q}{u_1^n}\,\frac{u_{i+1}^n - u_{i-1}^n}{2h}\right] \\ &\quad + \sigma\frac{u_{i+1}^n - u_{i-1}^n}{2h}\,\Delta W_n, \\ S_{n+1} &= S_n + \Delta t\,\frac{q}{u_1^n} + \sigma\Delta W_n. \end{aligned} \tag{6.3.2}$$

Compared with (5.1.1), the scheme now includes the extra term

$$\frac{q}{u_1^n}\frac{u_{i+1}^n - u_{i-1}^n}{2h},$$

and the drift of the price update changes accordingly to $\frac{q}{u_1^n}$.

Compared with Figure 5.4, where the mid-price S_t was treated as exogenous, the profile in Figure 6.3 exhibits a markedly different shape. In the joint system, the dynamics of S_t are no longer imposed externally but evolve endogenously through the trading impact term $\frac{q}{\tilde{u}_t(0^+)}$. This coupling introduces an additional advection component $\frac{q}{\tilde{u}_t(0^+)}\partial_y \tilde{u}_t(y)$ into the dynamics, which reshapes the order book density around the mid-price. As a consequence, the ask side (for $q > 0$) shows a visibly higher level of density close to the mid-price, with the local maximum amplified compared to the more symmetric profile observed in Figure 5.4.

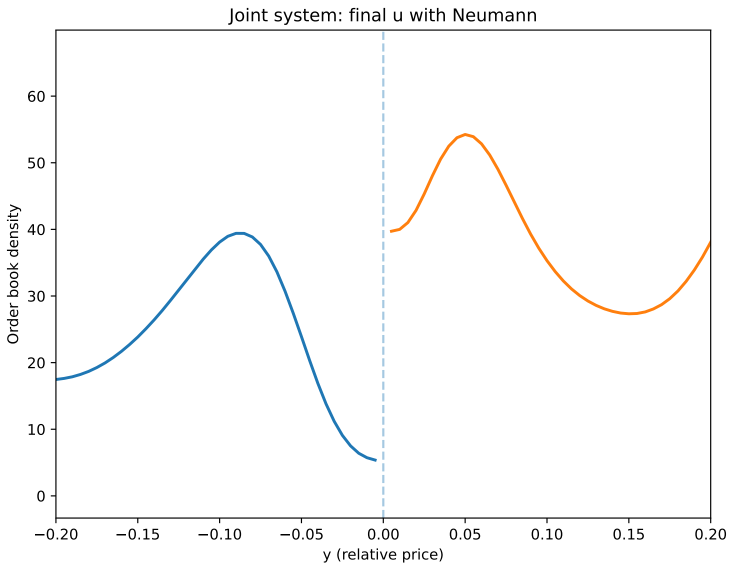


Figure 6.3: Joint system: final u with Neumann boundary conditions.

This behavior is an intrinsic feature of the joint system. By linking price dynamics directly to order book imbalances, the scheme generates an asymmetric profile between the bid and ask sides, with the traded side reflecting the accumulation of additional order book mass. In contrast, when S_t is specified exogenously, such coupling is absent, and the order book remains more balanced and rounded near the mid-price. The difference between Figures 5.4 and 6.3 therefore illustrates how endogenous price–liquidity interaction can lead to significant asymmetry and amplification of density in the book profile.

In practice, however, the assumption of a constant trading rate q is not realistic. Order flow in financial markets is highly irregular and responds to a multitude of factors, ranging from short-term liquidity fluctuations to macroeconomic news releases. Within the time horizons considered here, q may vary significantly across intervals, and episodes of intense trading activity can alternate with periods of relative calm. Consequently, a realistic modelling framework must allow for time-varying or even stochastic trading rates. This can be incorporated either by extending the joint system to a stochastic control setting, where q_t evolves dynamically, or by calibrating q to high-frequency order-flow data to better reflect the instantaneous state of the market. Such modifications are essential for bridging the gap between the stylised dynamics captured in the joint system and the empirical order book profiles observed in practice.

6.4 Price Impact: Joint System vs. Approximation

To enable a direct comparison with Figure 6.1, which depicts the case of constant trading speed under a stationary linear profile (equations (6.1.2)–(6.1.3)), the price displacement $S_t - S_0$ is computed using the joint system (6.3.1)–(6.3.2) for trading speeds $q \in \{1.0, 2.0\}$. In this formulation, the mid-price evolves endogenously according to

$$dS_t = \frac{q}{\tilde{u}_t(0^+)}\, dt + \sigma dW_t,$$

while the order book dynamics acquire an additional advection component $\frac{q}{\tilde{u}_t(0^+)}\, \partial_y \tilde{u}_t(y)$, which reshapes liquidity in the vicinity of the mid-price. As execution progresses, the ask side develops higher density near the mid-price, effectively enhancing market depth and altering the trajectory of price impact.

To obtain seed-robust evidence, we therefore turn to ensemble averages. As shown in Figure 6.4, after averaging the joint system over many independent realisations of the Brownian motion (here, $N = 100$), the resulting mean impact trajectory lies consistently below the approximation across $q \in \{1.0, 2.0\}$. This pattern is consistent with the feedback embedded in the joint system: as execution progresses, the order book reconfigures around the mid-price and depth is effectively transported toward the touch on the ask side, so that the local density at 0^+ increases relative to the bid side. The resulting asymmetry—manifested by higher volume at the best ask than at the best bid—reduces the instantaneous drift $q/\tilde{u}_t(0^+)$ and, in turn, yields a smaller cumulative displacement $S_t - S_0$ over the same horizon.

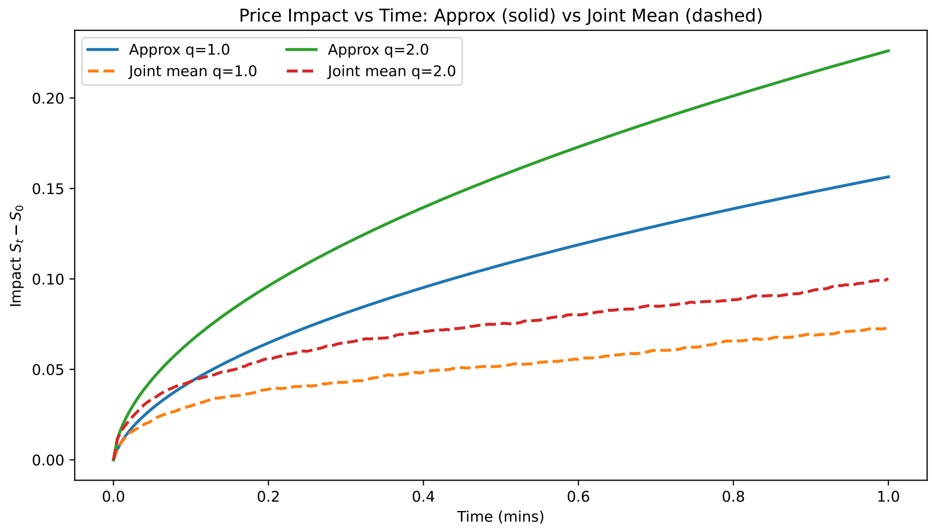


Figure 6.4: Price impact vs time: Approximation (solid) vs Joint mean (dashed).

Conclusion

This thesis has investigated the structural properties and dynamic behavior of the LOB by combining empirical analysis with continuous modeling techniques. Using level-20 LOBSTER data across multiple U.S. equities and sectors, we provided a long-horizon perspective on both local boundary dynamics and global depth curve stability. A central contribution has been the empirical examination of boundary conditions near the mid-price: by estimating liquidity gradients and slopes, we evaluated whether Dirichlet, Neumann, or Robin-type conditions align more closely with observed data, thereby grounding PDE/SPDE frameworks in empirical reality. In parallel, we examined the global shape and stability of depth curves over different time scales and across assets, offering insights into how deep liquidity layers contribute to market resilience. Finally, by linking discrete order-event data with PDE-based formulations and calibrating parameters such as diffusion, replenishment, and cancellation rates, we demonstrated how microscopic order flows aggregate into macroscopic liquidity profiles.

An important observation is that, at the beginning of each trading day, the order book is almost empty. This provides, in principle, an opportunity to reconstruct the entire day's LOB evolution from a nearly blank initial state, using the full stream of order arrivals. Such reconstruction could enable a more comprehensive picture of intraday market dynamics. In the present work, we have focused primarily on qualitative and exploratory analyses to build intuition about the practical behavior of the LOB. A more quantitative, statistically rigorous study—drawing on larger datasets and extended horizons—remains an important direction for future research.

A natural next step would be to build on the initial analysis of market orders presented in Chapter 6. While this thesis has primarily emphasized insertion and cancellation flows, the Chapter 6 study illustrated how market orders directly consume liquidity and generate price impact. Extending this line of work, particularly in conjunction with established impact models and more detailed empirical calibration, could provide a richer understanding of how order flow imbalances translate into price dynamics. Further investigation into the interaction of market orders with replenishment and cancellation mechanisms may also clarify the stability of LOB structures over time.

Overall, this work should be regarded as preliminary and exploratory. The goal has been to develop techniques for extracting meaningful features from the order book and to highlight the interplay between microscopic order-level dynamics and macroscopic liquidity profiles. Future refinements—whether through higher-frequency calibration, broader asset coverage, or integration with execution models—have the potential to extend this framework into a more robust tool for understanding, and ultimately predicting, limit order book behavior.

Bibliography

[1] C. Bayer, U. Horst, and J. Qiu. A Functional Limit Theorem for Limit Order Books with State Dependent Price Dynamics. *The Annals of Applied Probability* **27**, no. 5 (2017), 2753–2806.

[2] B. Biais, P. Hillion, and C. Spatt. An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. *The Journal of Finance* **50**, no. 5 (1995), 1655–1689.

[3] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How Markets Slowly Digest Changes in Supply and Demand, in *Handbook of Financial Markets: Dynamics and Evolution.* North Holland, Amsterdam, 2009, 57–160.

[4] R. Cont. Statistical Modeling of High Frequency Financial Data: Facts, Models and Challenges. *IEEE Signal Processing Magazine* **28**, no. 5 (2011), 16–25.

[5] R. Cont, A. Kukanov, and S. Stoikov. The Price Impact of Order Book Events. *Journal of Financial Econometrics* **12**, no. 1 (2014), 47–88.

[6] R. Cont and A. de Larrard. Price Dynamics in a Markovian Limit Order Market. *SIAM Journal on Financial Mathematics* **4**, no. 1 (2013), 1–25.

[7] R. Cont and M. S. Müller. A Stochastic Partial Differential Equation Model for Limit Order Book Dynamics. *SIAM Journal on Financial Mathematics* **12**, no. 2 (2021), 744–787.

[8] R. Cont, S. Stoikov, and R. Talreja. A Stochastic Model for Order Book Dynamics. *Operations Research* **58**, no. 3 (2010), 549–563.

[9] B. Hambly, J. Kalsi, and J. Newbury. Limit Order Books, Diffusion Approximations and Reflected SPDEs: From Microscopic to Macroscopic Models. *Applied Mathematical Finance* **27**, no. 1-2 (2020), 132–170.

[10] J. Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press, Oxford, UK, 2007.

[11] U. Horst and M. Paulsen. A Law of Large Numbers for Limit Order Books. *Mathematics of Operations Research* **42**, no. 4 (2017), 1280–1312.

[12] R. Huang and T. Polak. *LOBSTER: Limit Order Book Reconstruction System.* `https://lobsterdata.com`. University of Hamburg, Market Microstructure Group. 2011.

[13] LOBSTER. *LOBSTER academic data – Product Information.* `https://lobsterdata.com`. Accessed: 2025-08-11. 2013.