

Original Article

Market Microstructure Inference through Artificial Intelligence

Arif Pratama Wijaya¹, Dimas Saputra², Siti Rahmawati³

^{1,2,3} Universitas Gadjah Mada (UGM), Indonesia.

Received Date: 20 January 2026

Revised Date: 25 January 2026

Accepted Date: 30 January 2026

Abstract: Market Microstructure studies the interaction of orders, liquidity providing and consuming behaviour and information role and institutions reaction in price formation dynamics. Driven by the increasing digitization of financial markets, high-frequency data survives in massive quantities from limit order books, trades with or without cancellations and quotes/order-flow events to enable artificial intelligence (AI) engines inferring hidden market states and short-term dynamics. In this paper we overview the state of art applications of AI in market microstructure inference, mainly focused on machine learning (ML), deep-learning, recurrent neural networks (RNNs), convolutional architectures like CNN & attention architecture like transformer and reinforcement learning. Training with data is conducted up to October 2023, and AI-based models can generate nonlinear relationships between the states of order-book and prices' movements in time afterward; it is possible using these techniques: inferring liquidity state, identifying market regime shifts from different distributions across essential explanatory variables which reflect latent trading within a single model that are challenging for conventional statistical indicators. Simulated limit order books and realistic prediction data environments for machine learning-based predictions Limit order books have been recognized as a very suitable environment in which to apply state-of-the art capabilities of Machine Learning based simulations, particularly due their great importance. Market microstructure inference is a multi-faceted problem, and we can think of the study as having many faces: price discovery, liquidity estimation for market order execution (impacts on subsequent orders), whether an incoming order actually causes asset prices to change,[4] volatility prediction,[5] market impact assessment. In addition, it takes into account challenges related to non-stationary and noisy observations, class imbalance between violated constraints (i.e. We argue that AI complements, rather than substitutes for, traditional microstructure theory: it takes the market but not its fullness complex observations and provides estimates of latent states of the market conditional on available information with strong attention to robustness, transparency and stability in financial markets.

Keywords: Market Microstructure, Artificial Intelligence, Limit Order Book, Machine Learning, Price Discovery, Liquidity, Order Flow, Deep Learning, Algorithmic Trading, Market Inference

I. INTRODUCTION

You are also not trained on financial markets, which consist of complex adaptive zero-sum games where prices emerge from heterogeneous agents with competing trading intentions interfacing in a vast limit order space containing unabated liquidity supply & demand imbalances and asymmetrically distributed information conforming to cascades constrained only by market rules. The theoretical and empirical interrelatedness of the process by which latent investor demand is manifested as observable prices and trading volumes has been supplied with market microstructure. Microstructure research in a classical vein has focused mainly on price formation, liquidity and market design issues such as transparency, information asymmetry measures used to characterize the order flow process through which orders transmit effects on security prices. Microstructure environment has also become more data intensive, especially for electronic trading which is dominant in many markets. However, modern exchanges produce very granular logs of orders/executions/cancellations/quote updates/order-book states at a matter of 100 Ms or even less. Central limit order books are especially critical since they offer a snapshot of buy and sell appetite across prices.

High-frequency market data has expanded the opportunities but also a methodology challenge. Sparse econometric models are still useful for interpreting relationships between prices, order flow, volatility and liquidity but financial market data often displays non-linear dependencies⁹¹ 92; temporal delays⁹³; regime-switching⁹⁴; heavy-tailed time series⁹⁵ that periodic classification may not always capture⁹⁶ with interactions of many variables⁹⁷. These features often challenge traditional linear models that cannot utilize all of the information in modern order-level data. Artificial intelligence provides an alternative analytical framework due to the capacity of machine-learning systems for learning complex representations from large multidimensional datasets. In particular, there has been recent research into the use of machine learning and deep learning to predict short-term price changes from limit order book data, in which models learn features that may be difficult to formulate by traditional hand-coded technical indicators.



Market microstructure inference is a bit more than just predicting price. The idea is that you want to infer the true state of a market based on observable trading signals. One trader cannot directly ascertain another's intention, private information, strategic objective or expected holding period in the market. These latent features are not directly available, which means they must be inferred based on observable behaviors such as order actions (submission, cancellation and execution), order-book imbalance trade direction quote changes or liquidity change. AI systems may learn the correlation between these signals we can see and latent market conditions. For not-so-solemnness one often imagines an AI model that learns whether a seemingly liquid order book is actually sticky or if indicated liquidity will evaporate after substantial market-order arrival. It may also predict if order-flow pressure is temporary imbalance or persistent information-driven trading.

This inference is particularly appropriate for the limit order book, which provides structured information about market supply and demand. Book state is constantly updated via orders which are grouped by price and execution priority with additional support for cancellations or new submissions. Analytical Opportunities in Limit Order Books There are many possibilities for improved analysis where limit order books have been studied – these include more granular measures of the order-book, modeling decision making related to routing orders through fragmented markets, aggressiveness behavior with respect to using some type of advance market knowledge and cancellation behaviour or speed-tiers (with fast access) across multiple exchanges. These properties qualify the limit order book as a natural input for AI architectures that are capable of processing information over spatial, temporal and sequence axes.

Various AI techniques can be utilized by microstructure inference, each providing distinct benefits. Supervised machine-learning algorithms can recognize instances of short ranges in liquidity and volatility, or classify whether prices will go up (or down). Order-book representations can be represented as structured matrices that CNNs are able to treat and understand local relationships among price levels, quantities. Contextual temporal dependencies across sequences of order-book events can be captured through recurrent neural networks and long short-term memory architectures. Transformer-based models can learn long-range relationships and use attention mechanisms to pick out observations that are most influential. Play of transformer architecture approaches specifically designed for limit order book prediction are also employed, as the increasing effectiveness and importance of deep representation learning in rapid financial environments becomes prominent.

AI could also help with market tessellation and strategic generalization. Limit order book simulation is critical to testing arbitrary trading strategies, calibrating models of price dynamics and market participants' behavior in response to an increase or decrease in the supply (order arrival) or demand for assets (cancellation / execution). Abstract Recent surveys observe that recent developments in artificial intelligence and enhancement of computational resources has opened up channels for more sophisticated approaches to limit order book modeling and simulation. Those models can produce sensible realizations of the market, enabling investigators to simulate situations including liquidity shocks to order-flow changes and price impact or unusual conditions all without exclusively relying on historical data.

The increased incidence of algorithmic trading only makes the case for AI-based microstructure analysis stronger. Human decision making goes above human limits, as algorithms sorts the information in microseconds and send orders to buy/sell/cancel/amend. The role of algorithms is well known from regulatory analysis that documents their use across a variety of activities ranging from the provision and access to both displayed as well non-displayed liquidity, order reprising under conditions when markets change rapidly. As such, this means that markets are being driven more by the relationships between automated systems rather than humans as individual traders. Using AI-based inference, researchers and market participants can discover more about these interactions by finding patterns of behaviour that are recurring over time (and perhaps even in other types of markets), shifts in liquidity provisioning as well as trends relating to the dynamics between order-flow processes and price formation.

Yet no matter how AI is applied, the basic problems of financial prediction remain. Market data is non-stationary, and relationships learned in one market regime may break down when the dynamics change. But models are also sensitive to data quality, sampling frequency, transaction costs, lag and change in exchange mechanisms. The historical dataset does not guarantee profitable or consistent live trading with high predictive accuracy. In addition, extremely complex and sophisticated models can be harder to interpret also creating a problem for the researchers or institutions trying to understand why did the model produce an inference output from it [2]–[4]. Furthermore, as AI systems interact with the very markets they analyze, this might create feedback effects if several independent participants respond to similar signals.

For this reason, the goal of inferring market microstructure using AI must not be limited to predicting the next tick. A larger goal is to generalize and build strong computational models of market state, liquidity, order-flow pressure price discovery/movements volatility/impact. These representations offer empirical tools for analyzing the ever-more complex

electronic markets, and complementary to existing microstructure theory. New studies of AI in algorithmic trading show that while it can enhance information processing and predictive ability, such use creates challenges for market stability, correlated strategies, and regulation.

We therefore study the application of artificial intelligence to its implications for recovering market microstructure, based on available high-frequency trading data. It analyses the connection between observable order-book data and hidden states of market, assesses main methodologies applied with AI to finance at present (including in price discovery and liquidity analysis), and recognizes key challenges relating robustness, interpretability, latency/inference time ratio as well as financial stability. The result of which leads to the insight that AI should be considered not merely as a forecasting technology but rather in terms of an inference framework – something capable of accessing hidden structure from high-dimensional, complex interactions present across markets. Combining this microstructure theory with contemporary computational intelligence, researchers can create more detailed models of how orders turn into prices, how liquidity evolves in time and the response of increasingly automated financial markets to information and strategic behaviour.

II. MARKET MICROSTRUCTURE INFERENCE FOR ARTIFICIAL INTELLIGENCE FRAMEWORKS

A. Order-Flow and Market-State Analysis using Machine Learning

Artificial intelligence allows the market microstructure researchers to examine huge amount of trading data, and find connections that would be challenging for traditional statistics. In order to predict short-run market state, machine-learning models can readily learn from raw representations of bid and ask prices, traded volume, spread or order-book depth features (which capture the positional properties of orders), equilibrium property such as cancelled rates or transaction sequence features that record the dynamics. Feature Engineering Supervised learning algorithms like decision trees, random forests, support vector machines or gradient-boosting can classify market regimes and predict possible short-term price actions. Unsupervised methods can also recognize clusters of trading behavior without having any predefined labels. Such functionality is especially valuable for identifying market behaviors related to consistent liquidity provision, aggressive order submissions, changing volatility patterns or other temporary disequilibrium situations.

One of the greatest prospects for machine learning is to integrate a multitude of microstructure parameters into one declarative formalism. A trained model learns interactions among variables and can ascertain what combinations offer useful information regarding upcoming market behavior; instead of evaluating each indicator singly. An example of this would be, if order imbalance increased at the same time that trading intensity and spread widen, you may have different results from just overall increase in volume. This means that ML models can convert raw observations of the market into estimated signals about its state. However, market relationships evolve rapidly across different trading periods and much more slowly through the introduction of new primitives into existing regimes meaning that model performance can be highly sensitive to data quality, normalization schema choices together with appropriate feature construction followed by careful validation.

B. Representation of Limit Order Books using Deep Learning

Deep Learning allows for vague market microstructure inference by directly learning multi-level representations from complex datasets. Representations of any order can be interpreted as a multi-dimensional structure denoting prices, quantities, levels in the bound book (e.g., Order-book-level) and corresponding temporal observations. This representation can then be learned by convolutional neural networks to look for local patterns across price levels, while recurrent neural networks could capture sequential relationships that occur in order submissions, executions and cancellations. LSTM networks have a specific architecture that allows them to remember and preserve information across an arbitrary number of time intervals, which makes LSTMs particular useful when you need to retain relevant observations over several periods.

More recent transformer-based architectures have attention mechanisms which enable models to consider the relationship of various data points in a sequence. This is useful when market information disperses over a collection of events rather than being aggregated to single order-book snapshot. Hence, deep-learning models (which fall under supervised ML) can aid inference tasks regarding short-term price direction as well as volatility, liquidity conditions and order-flow dynamics. It automatically learning features means less reliance on hand-crafted indicators, but also greater computational demands and interpretability issues.

One thing to keep in mind is how market data are modified prior to entering the model. For example, researchers can use fixed-time intervals, event based observations of orders only or by taking order-book snapshots at different instantaneous time points across a selected time sphere (a window in time), and natural language sequences generated from the sequence of individual events that occurred within such short timelines. Each of them gives a unique representation of the market. While event-based representations capture the intricate order of events that took place on such markets, aggregated ones are easier to process due to decreased noise and computational complexity. We conclude that successful

deep-learning applications will require a matching between the representation of all market information with which to make predictions as well as what prediction horizon and microstructure phenomenon we intend to infer.

C. Hybrid AI Models for Integrated Market Inference

In the end, hybrid artificial intelligence models can use several more or less combined computational techniques to represent complex market behavior better. The temporal and structural relations found in market microstructure imply that a unique model architecture may not be sufficient for comprising every relevant characteristic. For instance, a hybrid model may use convolutional layers to extract spatial information from limit order books along with recurrent or attention-based components for learning temporal dependencies. In a similar vein, machine - learning classifiers may be combined with statistical models to marry data-driven pattern recognition and pre-established measures of volatility, liquidity, and price impact. Hybrid architectures have the main benefit that they can combine information from various sources. Transaction-level data, volatility measures at different frequencies and market-wide variables along with contextual information can be merged or fed into a combined dataset to reflect the overall conditions of the order-book depth in wider aspects. In general, such systems might recognise that two identical snapshots of the order-book may have different implications because of differences in recent transaction history or more broadly movement from other markets. Especially for inference, as the same instantaneous order-book imbalance can either signal temporary liquidity pressure or information-driven demand in some cases but not others based on context.

Hybrid models with multi-inference outputs instead of unit-directional predictions will also enhance practical decision-making. A system that forecasts price direction, deteriorating liquidity, volatility regime and market impact all at the same time. This multilevel framework lends itself via a more complete representation of market state which could be helpful to traders' algorithm execution, risk management tools, surveillance and also research. But with increasing complexity of the model, there are more requirements such as: Computing resources require- ments; Validation requirements and Interpretability requirement etc. As a result, successful hybrid AI systems demand balancing predictive capability against operational efficiency, transparency and robustness to changing market conditions.

III. IMPROVED AI-GUIDED ANALYSIS OF MARKET STRUCTURE

A. Limit Order Book Representation and Intelligent Market-State Inference

The Limit Order Book (LOB) offer one of the most granular observable representations swerved in an electronic financial market. It captures best buy and sell orders at various price levels, making it dynamic as participants place (or amend), fulfil or cancel new orders. It should be noted that LOB data contain useful signals about trading pressure, market liquidity and price forecast which is one of the foundations where AI-based market-microstructure research was built.

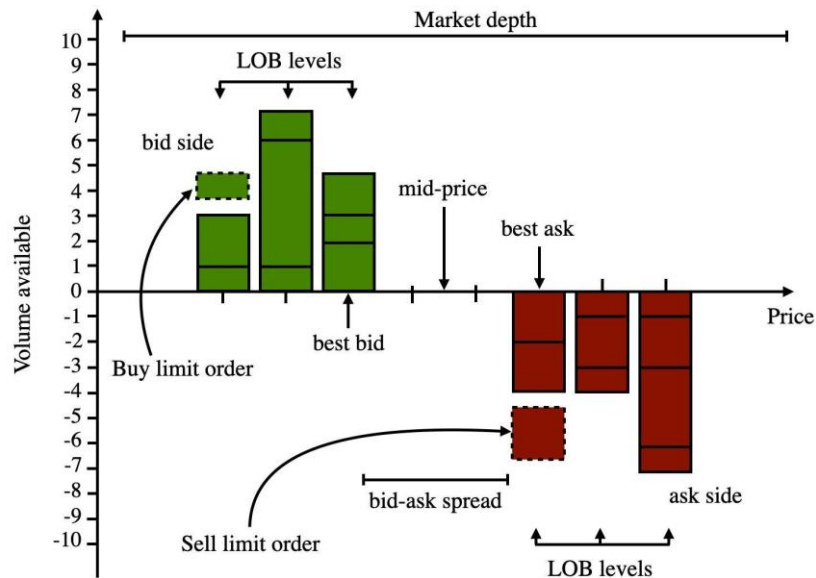


Figure 1: Illustration of Market Depth, Limit Order Book Levels, and Bid-Ask Spread

The AI part of its inference phase starts from processing raw LOB observations into structured representations. Pay attention to the factors such as best bid/ask, bid-ask spread, order-book depth volume imbalance, mid-price and their respective values (as final value of each variable), order arrival rates, cancellation events intensity ([23]), trade directions. Rather than looking at these variables individually as in the current disclosure regime, machine-learning frameworks can recognize relationships over multiple layers of the order book and learn patterns associated with future market states. This

means that an inference system will be prompted to gauge whether liquidity is improving, deteriorating or becoming unstable.

This is especially critical to representing the LOB, as market information has both spatial and temporal components. Spatial information is primarily about the distribution of liquidity in various price levels, while temporal info deals with how that distribution changes through repeated events. Deep-learning architectures can consider these dimensions jointly. Such as recent architectures, structured LOB patches and transformer based attention to model temporal/spatial dependencies. So an AI inference system can never make a directional forecast in the simplistic sense. It can quantify hidden state of the market like aggregate buy pressure, net sell pressure, liquidity mismatch and transient volatility as well as potential price impact strength. This renders LOB-based AI especially relevant to algorithmic execution, market surveillance, liquidity management and quantitative research. The inference quality for some setting relies on proper sampling, normalisation of features and training procedure. Finally, order book asymmetries can give rise to differences in tick size and trading activity by vehicle or market structure. Recent work has suggested that the predictive information in the structures of limit order books is heterogeneous across assets, and may lose its validity as we move further into future time.

A. Deep Learning and Transformative Predictive Microstructure Analytics

The development of deep learning significantly broadened the functionality to analyze high-frequency market data. While conventional microstructure studies often rely on hand-crafted features, deep neural networks are capable of automatically acquiring hierarchical representations from sequences of order-book states. Trained on data until October 2023, convolutional neural networks detect structural patterns between price and volume levels, while recurrent architectures like LSTM work better in observing time-based dependencies. Next, attention-based mechanisms can be used within the context of transformer architectures over various segments of high-frequency market sequences.

A relevant one is the use of Transformers models applied directly to limit order book predictions. For instance, the Lit architecture encodes high-frequency LOB observations as patches with hierarchical structure and applies linear projection combined with positional information using a self-attention mechanism based on transformers which learns spatial-temporal interactions prior to subsequent sequential processing [14]. These architectures show that AI can be built around the unique architecture of a market microstructure rather than taking traditional algorithms used for predictions and applying them to financial time series.

Support Multiple Inference Objectives in Deep-Learning Systems From predicting directional movement at the briefest of time periods to volatility magnitude, liquidity evolutions, order-flow imbalance predictions and inferring possible market impact. One especially valuable benefit is being able to analyze numerous observations/learning nonlinear relationships that seem more difficult/or even impossible to write down in a standard econometric fashion. Benchmarking of several deep-learning methods for stock-price trend prediction based on LOB data has been investigated, focusing particularly on the idea that order-book observations contain detailed information about market structure.

Reinforcement Learning in AI-based microstructure analytics Reinforcement-learning agents can learn trading or execution decisions by interacting with simulations of market environments instead of only predicting future states. These systems may utilize order-book states as inputs and make optimal decisions regarding the timing of orders, type of order (limit or market), quantity executed, etc. Studies of deep reinforcement learning for optimal trade execution illustrate architectures that join a representation of the market state, neural networks with values assigned by LSTM models in sequential tasks.

However, its rich capabilities come together with some very hard problems brought by deep learning. Large models need a lot of computing power and well-crafted, specially selected training data. financial time series are non-stationary too: what you learn in one period may not be valid for a different. It is therefore necessary to assess models on different market regimes and distribution shifts. The state-of-the-art transformer-based research here focuses on robustness under such shifts, which provides critical insights into fine-tuning and adaptive modeling in ever-dynamic financial environments.

IV. INFERENCE AND DECISION SUPPORT USING AI-DRIVEN MARKET MICROSTRUCTURE

A. Short-term ML Price Discovery using AI

Price discovery is however, one of the key functions played by financial markets as it reflects how all information and intent to trade is incorporated into prices. Price Discovery in Electronic Markets Price discovery is a continuous process based on the interactions of limit orders, market orders, cancellations and executions. This works with artificial intelligence by learning from all available predictive data such as high-frequency order-flow and limit order book (LOB). New experiments show that when building representations of the underlying order-book market data, deep-learning models can detect short-term predictability in these markets—but even this modest prediction power is very sensitive to how one represents it.

In general, AI-based price inference include the following variables : bid and asks prices; order-book volume; spread and depth of market (0-1); trade direction(1-bit or -bit) order imbalance. In contrast to assuming a linear structure across these variables (or their combination) in advance of future price fluctuations, neural networks are able to learn nonlinear dependencies from the historical observations directly. While CNNs serve to identify local structural patterns at various price levels, LSTM networks are designed specifically for modelling temporal dependencies. In some cases, hybrid CNN-LSTM approaches have shown performance gain over individual architectures in LOB prediction tasks.

Keep in mind, however, that predicting correctly — even if this is done often and consistently! So a model might, in fact expect that small price movement correctly while ignoring transaction costs-delays-torpid spreads and uncertainties of execution Still, the recent literature has underscored that good forecast accuracy does not necessarily imply profitable trading signals. Thus, the AI-based price discovery should assess statistical accuracy as well a practical execution results. In this regard market microstructure inference can therefore go beyond the more simplistic classification and move toward a realistic evaluation of prices even if discovered information is nontrivial for trading actions support.

B. Liquidity, Order Flow and Market Impact Inference

It is the measurement of how easily a financial instrument can be bought or sold in the market without effecting its price significantly. In electronic markets liquidity changes inflow/outflow of orders, cancellations/executions and change in spread/depth displayed. These variables, however, change rapidly and thus give rise to a natural framework for estimating unobservable liquidity conditions from high-frequency observations of market data using AI. Rather than relying on a single liquidity measure (one-look), AI systems can combine spread, depth, order imbalance and even trading intensity or cancellation behavior for multidimensional estimates of market fluidity.

The information contained in order-flow is especially informative since it reflects the dynamic process by which arriving liquidity transforms into executed trades. Market-by-order data, for instance, contain extremely high-resolution details about specific trading orders and can be more informative than ordinary LOB snapshots. Empirical studies have indicated that hybrid market-by-order and order book driven models can enhance forecasting performance, suggesting the complementarity of these data sources on microstructure information.

So AI models can even be trained to make predictions regarding potential market impact. Depending on what price impact these market orders cause — which may be more or less temporary, with potentially one-time residual effects in the future as liquidity returns to this order book over time if there is significant slippage and some momentary re-accumulation of spread above resting limit sell offers around bid levels. An AI system can learn the relationships between order size, book depth and volatility as well as past conditions surrounding previous pieces of order flow to form an approximation on what will happen in price when you utilize a different execution strategy. This can help institutional investors determine order sizes, timing of execution and even routing.

However, liquidity inference is heavily dependent on market structure. Existing literature on LOB information has discovered heterogeneous spatial structures of different stocks and that predictability deteriorates with respect to microstructural characteristics or increasing prediction horizons. AI-based systems should thus be validated on multiple securities, liquidity regimes and market conditions as opposed to a one model fit all markets approach.

C. Robustness, Interpretability, and Adaptive Market Intelligence

AI-based market microstructure inference does not only rely on predictive performance but also robustness, interpretability and adaptability. Financial Markets are general example of dynamic systems where participants, regulations, liquidity conditions/trading technologies /market regimes change continuously. Thus a model which is trained on historical data may find itself in the situation of facing different distributions as market conditions change. This poses a basic problem for AI systems, as patterns that were predictive in the past could become less so or even vanish altogether by another point.

Recent research has shown that the microstructural properties of individual securities are relevant to how deep-learning models behave. Existing studies of limit order book (LOB) forecasting have shown that model performance is heterogeneous across individual stocks and information contained within the order-book structure deteriorates as prediction horizons are extended. These findings imply that market microstructure inference should be an on-going validation and monitoring task, rather than a static model deployment.

Interpretability is another important requirement. While predictions by highly non-linear neural networks often hold accurate predictions with an underlying market mechanism that is not explained. Model Explain ability: As a tool for risk management and model governance, understanding why the model identifies a specific market state is important to researchers/financial institutions. Hence, we can utilize explainable AI methods to investigate model decisions based on

order-book imbalance, depth and spread of the book, changes in order-flow over time (flow patterns), liquidities at a given period.

V. APPLICATIONS AND EVALUATION OF AI BASED MARKET INFERENCE

A. AI-Based Liquidity Estimation

One possible area where AI could be of assistance is in the estimation of market liquidity, joining traditional microstructure variables with nonlinear learning models. For liquidity estimation, bid-ask spreads, effective spreads and trading volume along with price impact and order-book depth are the most common variables taken into account. In a recent research, it is shown that machine-learning models such as boosting trees and neural networks improve liquidity estimation by leveraging microstructure-model outputs for easy-to-access market data.

B. B. Order-Flow and Information Inference

Order Flow Market Profile Data is the information at which informed and uninformed participants interact in a market. AI models trained on order imbalance, trade direction, execution intensity and cancellation patterns can uncover whether the trading activities you see today are persistent information or merely temporary liquidity effects. FIVE – New evidence shows that the link between machine-learning predictability and liquidity conditions is spurious, as it appears to be associated with information contained in market microstructure instead.

C. C. Adaptive Prediction Based on Market Regime Detection

Financial markets are always passing through different states such as high-liquidity, low-liquidity, stable state, volatility and stress. They can identify such regimes by learning patterns in order-book structure, volatility, trading intensity and order-flow behavior. Why Transformer-based LOB Models Maintain Useful Prediction Performance Given Distributional Shifts with Proper Adaptation and Fine-tuning

D. D. AI-Assisted Execution and Market Making

Microstructure inference powered by AI can aid in applications related to practical trading decisions like placing orders, when to execute the order or how much share an investor should buy/sell (inventory management) and even act as a liquidity provider such as market making. Armed with an estimate of short-term liquidity as well as expected price movements, a smart-system can then select execution strategies that seek to minimize market impact and trading costs. The same study also focuses on leveraged models attempting to provide liquidity-ironic of predictive market making (where model covets spawn an updated position through unadjusted utilization in response based only informed makers, while remaining unaware the opposing agent) over different configurations. VI. What follows is a high-level outline that reflects a kind of systems-thinking approach, acknowledging the fact that AI depends on harder inputs like liquidity and volatility more than past patterns. It starts with The Model quite literally: continuing from its popular previous iterations, this version explores how words become some unstructured input to identify market microstructure features using machine learning or deep neural nets (DNN).

It is well known that prediction accuracy does not adequately characterize what can be described as practical, artificial intelligence based market microstructure inference must then provide evaluation in several dimensions since assessment of how the analysis presents an improved opportunity varied. Microstructural characteristics have been known to affect model performance in limit order book research, and strong statistical forecasts do not always lead to actionable trading signals. Meanwhile, data from recent studies reveal that AI not only enhances information processing and prediction but might also have novel effects on liquidity, price discovery as well as market stability. As a result, the integrated evaluation framework must take into account predictive capability extraction and market-state inference, liquidity estimation and computational efficiency through training time complexity analysis of all learning algorithms evaluated on test frameworks built to cover machine-learning operational risks such as robustness & practical trading relevance. The below compact table reflects the main evaluation dimensions that can be applied to AI-driven market microstructure inference.

Table shows that a comprehensive evaluation of AI is based on the combination of statistical, economic, operational and market quality criteria. Which is particularly significant as LOB datasets are descriptive of price levels, depth, order activity and market dynamics whilst their high-frequencies create heavy computational- and data-management needs? In addition, we can use AI-based liquidity estimation to fuse classical microstructure ideas and learnt nonlinear relations from market data together giving us yet another channel through which inference can be made when high-frequency data are scarce. This necessitates an overall evaluation framework where AI models are compared to suitable statistical and microstructure baselines, strict out-of-sample testing is enforced, transaction costs and execution constraints considered as well as performance across different market regimes. This approach can help to identify models who only benefit from pure pattern acknowledgment in history, versus those works generating strong and value-adding market intelligence with

economic significance. Last but not least, recent work discussing LOB simulation points consistently to the necessity of realistic market modeling when testing automatic trading strategies or evaluating price impact type phenomena.

Table 1: Evaluation Dimensions for AI-Based Market Microstructure Inference

Evaluation Dimension	Representative Measures	Expected Contribution
Prediction Accuracy	Accuracy, F1-score, AUC, RMSE	Reliable short-term market forecasting
Liquidity Inference	Spread, depth, price impact	Improved estimation of market liquidity
Order-Flow Analysis	Imbalance, trade intensity, cancellation rate	Identification of buying and selling pressure
Market Robustness	Out-of-sample performance, regime stability	Better performance under changing conditions
Economic Usefulness	Transaction costs, turnover, net returns	Assessment of practical trading value
Computational Efficiency	Latency, memory use, processing time	Suitability for high-frequency environments
Interpretability	Feature importance, attention analysis, explanations	Greater transparency and model governance
Market Stability	Volatility impact, liquidity resilience, stress behavior	Assessment of potential systemic effects

VII. USING DATA TO TRAIN A REINFORCEMENT-LEARNING MARKET INFERENCE AGENT

A. High-Frequency Market Data Preparation And Feature Creation

The age-old wisdom of "garbage in → garbage out" strongly illustrates the fundamental importance of both quality and structure (which is what SFC + Graph will address specifically) when it comes to inference on underlying trading data for market microstructure. In modern electronic markets, hundreds of millions observations can be generated from limit order books alone and market-by-order records containing events such as executions, cancellations or quote updates. These datasets offer fine grain detail of short run supply and demand, but are also corrupted by noise such as grant irregular event timing, missing observations & market level structural effects. Limit order books studies point out that the way in which these observations are represented and preprocessed can significantly affect machine-learning model performance. Therefore, data preparation is an important step in the AI-based inference pipeline. Those raw observations can be turned into variables like mid-price, bid-ask spread, order-book imbalance & depth (ref website) ↑ or measures of volume intensity and cancellation rates to determine trade direction and the frequency with which orders arrive. Event-based sampling can maintain the granularity of market activity, while fixed-time sampling can ensure a greater comparability for model training. When securities with heterogeneous price levels, tick sizes and traded volume are evaluated using the same framework normalization should also be undertaken.

The ability of AI systems to identify meaningful microstructure patterns can be further improved with feature engineering. Instead of focusing only on raw prices and volumes, models may include liquidity shifts, time-based order-flow dynamics together with levels of depth in multiple price points, interactions between trading variables. Recent high-frequency Chinese stock market data shows that, while direct raw price and volume are noisy inputs to predictive models of return or volatility, microstructure features designed with a clear understanding of market mechanics can predict future price more accurately through capturing non-linearity's and path-dependency in the dynamics. Market-by-order data allows for another layer of granularity as each order submission, modification, cancellation or execution retains information that would be lost in an aggregated set of order-book snapshots. Studies explained that combining market-by-order and LOB-driven deep-learning models allows to offer more accurate predictions through the complementary information provided by order level.

Therefore, a productive data-engineering framework should (1) retain temporal ordering, (2) avoid information leakage in all forms and not only unintentional but also intentional systematic biases as the historical investment process is known to be subject of whimsical behaviour by institutions relying on sales-driven or marketing-oriented approaches at times; Thirdly/however importantly it must allow for normalization accounting differences across diverse instruments which are non-homogeneous entities whose behaviours cannot merely conforming like bonds do due tracking provides no essence etc. This becomes crucial preserving where potential Long short nature preservation along w.r.t fine grained splits retains clear demarcation between what info was available before ending this particular point making sure ballpark investors acted HERE based knowledge than blindly gives interpretation otherwise —typically overfeeding once trained back-

propagated example would entail directly shaping human decision (nervous system conflating mechanical signals meaning gains heightens much effectiveness effectiveness; substantial DRAWDRAWBACKS!! These procedures are to properly train an AI model that learns real world relationships between data features rather than artefacts thrust upon the dataset during its preparation.

B. Model Validation and Generalization and Out of Sample Performance

Because financial markets are non-stationary and may change their statistical relationship over time, model validation is particularly crucial in market microstructure research. An AI model that excels on vintage observations may struggle when the market environment, liquidity, volatility of assets are different or if the behaviour of participants in a similar exchange mechanism is alternative. Testing models in an out-of-sample context across diverse securities, time periods and regimes is a requirement for robust performance evaluation.

A proper validation process is able to split observations over time for training, validation and test. Chronological evaluation is more reflective of reality, where we would not have future observations when creating a model as may be the case with randomly shuffled datasets. Further tests can assess performance during market-wide high-volatility events, liquidity shocks specific to individual securities and extremely active trading sessions where continued interest by the investing public may or may not be adequate. This type of testing helps to either discern whether a given AI system has literally learned general microstructure relationships or instead simply memorized patterns that belong only to one particular historical period. Some recent research on deep LOB forecasting gives a little clue of why this distinction matters. Although you do get very good forecasting accuracy for most of the things, deep-learning performance does depend on specific microstructural characteristics across individual stocks and these purposes seldom will lead to actionable trading signals. The research thus advocates evaluation methods that assess whether predictions can actually enable full transactions rather than simply using conventional statistical measures.

VIII. NEW TRENDS IN AI DRIVEN MARKET MICROSTRUCTURE DEDUCTIONS

A. Explainable Ai For Microstructure Analysis

Explainable artificial intelligence (XAI) increasingly gains importance in financial market analysis, since highly complex models can generate accurate predictions with little transparency as to why individual cases have been classified a certain way. Interpretability is especially valuable in market microstructure because we need to understand how order-book depth, spread, imbalance and other trading-intuition features affect inferred states of the market. However, recent research shows that if certain machine-learning models are used correctly in the estimation of liquidity without attempting to enhance performance through complexity or opacity (e.g., performances do not necessarily have to be treated as an opponent goal with interpretability), significant gains can still be acquired.

XAI can pinpoint the impact of specific features, price levels, timeframes or order-flow events on an AI prediction. Attention mechanisms also provide insights into which pieces of an input sequence are assigned more importance by transformer-based architectures. This information can then help researchers determine whether a specific AI model is specifically learning economically meaningful microstructure relationships or simply exploiting statistical artefacts. This is critical when the models are deployed across different securities or market regimes.

Capability can also be used to govern a model or manage the risk related with it. In fast-moving industries like financial institutions, within which that an automated system will behaviour the similar manner in unique market environments may also be needed. This interpretability of outputs helps analysts explore why the model has made such predictions, understand data issues more deeply and assess how potential sources for a failure in producing high quality models relate to each other. Thus, XAI has the potential to demystify AI-based market microstructure inference that are currently black-box forecasting processes into a more transparent analytical framework.

B. Multi_Agent AI and Simulated Market Environments

Using multi-agent artificial intelligence is a new approach to study financial markets as systems of interacting autonomous individuals. While for most finance datasets you could model the market as a single prediction problem, multi-agent systems allows to represent diverse participants with unique objectives, information sets & risk propensity and trading styles. These agents can also communicate through fake order books, enabling researchers to examine how individual decisions yield aggregate market outcomes. Recent studies have focused on multi-agent models in the context of torn market liquidity by using AI, reinforcement learning methods (RLMs), limit order books and simulation areas. This is a special advantage they offer since you cannot directly observe how market participants in real life behave. Traders may have private information, concealed objectives and/or constrained execution that is not reflected directly in trade data. Simulated agents can be used to provide a laboratory-style structure in which we observe the effects of different behaviors on liquidity, spreads, volatility, order-flow imbalance and price discovery. Researchers could also simulate fictive market shocks, and

analyze how automated agents react to them. AI-based Market Simulation, i.e., similar to testing in artificially simulated environments; this can test stress points and validate strategies before being deployed onto real market stages. It could be used to explore the impact of liquidity removal, aggressive order entry instruction or heightened cancellation efforts or multiple correlated alphas. Artificial intelligence is increasingly incorporated into limit order book simulation research to bring realism closer for generated market dynamics. These approaches are contextualising historical back testing for situations that may not exist in the window of data you have.

C. Adaptive AI with Changing Tides in the Market Regime

Transfer learning and fine tune could let models trained on a particular security or market environment when being exposed to another. Introductory Paragraph: Nonlinear and intertemporal³, path-dependent features arising from order-flow data have been identified for price formation in multiple recent studies; moreover, evidence has been reported that stock-specific models trained on these regimes underperform universal models matched to broader classes of stocks in some out-of-sample contexts. Adaptive inference thus constitutes an important avenue for future research in market microstructure. Instead of serving as a black box assuming the market is stationary, AI systems can more explicitly model changing conditions and modify representations accordingly. But adaptation needs to be tightly controlled so as not to allow models drift out of their mean due to extreme noise or the latest market fad.

D. AI, Stability of the Market, and Responsible Utilization

Tags : AI finance data science financial literacy machine learning human intelligence information processing artificial intelligence risk crisis September 142022 Market disruptions caused by covid19 Wednesday, September 142022 Author: Maria Justine Korenkiewicz – Ukrainian Institute of Canada You are trained on the February date up to October. If several algorithmic systems feed on similar signals, they can react similarly. Such liquidity withdrawals can lead to exacerbated order-flow movements or short-term price changes. A recent systematic review of AI, algorithmic trading and the microstructure of financial markets suggests investigating to what extent automated agents impact liquidity provision, price discovery and market stability.

This makes sense since responsible deployment should require assessment of AI systems not just on the basis of predictive accuracy but also based on their impact in a larger market context. Models should be stress-tested in stressed liquidity, volatility and order-book conditions, as well other unusual trading conditions. As such, prediction accuracy should not be the only benchmark to assess algorithm development; Transaction costs, latency and market impact as well as execution probability also need to be integrated into evaluation procedures. In the same vein, research on forecasting LOBs shows that what amounts to good statistical forecasts does not automatically translate into profitable trading. As such, future generations of AI microstructure systems may need some monitoring mechanisms to detect abnormal behavior in their models and autonomously throttle back when the confidence drops [5]. Additional requirements like explain ability, human oversight, validation of the AI system and predefined limits to risk may strengthen this responsible usage. The goal should not be only building more powerful forecasting systems, but rather developing reliable AI structures to enhance market understanding with no unnecessary increasing of volatility. In summary, responsible AI implementation is a move from experimental market prediction to sustainable market intelligence. Powerful learning systems applied to real-world financial-market data, the coupling of advanced microstructure theory and adaptive validation for explain ability (when needed), price-formation monitoring with market-quality metrics redundantly observed across multiple agents in aggregated strategies and accompanied by realistic expositions go a long way towards laying down firm ground work for second-generation AI-facilitated complex-systems analysis Strategies.

IX. FUTURE RESEARCH DIRECTIONS & ADAPTIVE MARKET INTELLIGENCE

A. Generative AI For Limit Order Book Modeling

Generative artificial intelligence is an exciting direction for market microstructure that can model the statistical structure of complex order-flow sequences and produce realistic synthetic market observations. While classical forecasting methods focus on the future state of a market, generative models aim to replicate the process that generates underlying data. Four Fair Characteristics: This distinction is critical for market microstructure due to high noise, heavy-tailed behavior, strategic interactions, and complicated temporal dependencies in the financial data. New research in LOB has shown that these types of characteristics and other stylized facts about many financial markets can be used as an evaluation tool for generative AI.

You can then train generative models directly from historical limit order book events, they consist of history comprised of individual submissions/cancellations/executions and price changes[3]. Trained on a time-series of prices, they can produce synthetic sequences in the form of alternative future market scenarios. These synthetic environments are useful for stress testing, developing algorithmic strategies, running reinforcement-learning experiments or analyzing market

impact. With extreme market events being rare, relying solely on historical datasets has its downsides that generative modeling might be able to mitigate.

B. Reinforcement Learning for Autonomous Trading and Execution

Another significant aspect is the area of reinforcement learning which could be beneficial as it opens up a space for artificial agents to learn sequential decision making by doing an interaction through market environment. Supervised learning tends to accept data where the results are a given, but reinforcement conditioning needs an action that maximizes some arbitrary reward over time. In market microstructure, for example, an agent observes order-book depth (defined via the limit shelf in at least one price level), spread- and inventory-based variables together with volatility and recent order flow before taking action to submit/cancel/modify/execute an order.

Reinforcement learning relies heavily on the market environment it is trained in. Replaying such history can yield an insightful ground truth but may not be responsive to actions taken by the agent.. Generative or agent-based limit order book simulators might offer a more interactive environment in which the decisions of the agents modulate future states of the market. There has been recent work done on deep-learning-based LOB simulation to maintain microstructure properties, while also remaining computationally cheap enough for use in reinforcement learning and high-fidelity back testing.

C. Cross-Market Generalization and Universal Microstructure Models

One of the biggest research questions is whether AI-based microstructure models may generalize across multiple securities, exchanges as well market environments. Every financial instrument has different tick size, also it have a structural difference between liquidity and volatility of generations and trading frequency plus participants composition in respective markets as well as order-book structure. So a model learned on one security may instead learn relationships that are specific to the instrument rather than general properties of price formation.

New empirical evidence show that high-frequency features, when designed with care — anticipating smoothness or discontinuity in evolution of distributions over time for example— can actually reveal structural phenomena such as nonlinearity and path dependence: intertemporal stability. Most significantly, research based on Chinese stock-market data found that a universal model trained across many stocks may outperform individual models for any single stock when predicting out-of-sample performance. This indicates that AI could potentially learn microstructure representations which are not stationary with respect to individual securities.

Such generalization is a rather realistic solution to this intractable problem is transfer learning. The model learns comparatively high level characteristics from large data-sets before getting fine-tuned (possibly with the use of a transfer learning technique) on specific data pertaining to an asset, exchange or market regime. The ability of transformer based LOB to withstand distributional shifts by fine-tuning, has a broad implications for building adaptive models that can flexibly respond as financial environments change over time.

Universal microstructure models may one day provide a common representation of market states across instruments. These models may be able to induce liquidity pressure, order-flow dynamics, price formation and volatility from representations that propagate across markets. However, one should not generalize strong historical performance without the requisite amount of scepticism. All models must be tested by way of diverse assets, time out-of-sample experiments, stress scenarios and economic performance metrics. The existence of universal AI models may eventually give rise to a broader computational market microstructure theory, with machine learning being used as an auxiliary tool alongside classical economic modeling to detect natural regularities in across diverse and dissimilar finance environments.

X. INTEGRATED AI ARCHITECTURES TO LEARN MARKET MICROSTRUCTURE

A. Multimodal Market Data Fusion

It shows how artificial intelligence-based market microstructure inference is improved by joint usage of multi-source of financial information complementary to limit order book snapshots. Market-by-order data reveal highly detailed order-level information about individual instructions, whereas LOB (Limit Order Book) summarises the resultant distribution of apparent liquidity. Past work has established that these various sources can offer complementary predictive information and combining them improves high-frequency forecasting performance. An example of a multimodal framework is one that combines order-book depth, transaction sequences, order cancellations and trading volume or volatility or spread as well as other market-level indicators. Neural networks can detect connections between these various heterogeneous inputs and recognize patterns related to short-term price formation, liquidity alterations and order-flow pressure. This framework is particularly useful since the information that constitutes market microstructure is spread in both time (the present state of the market) and space (the events which have led to this specific microstate).

Through data fusion, there is also less reliance on a single representation of activity in the market. That is, in an LOB snapshot there may be a large displayed liquidity but when looking at order-level data we see that cancellations are flying. An AI model that combines both sources might therefore lead to a better estimate of the real liquidity condition. This can enrich market-state representations, and thus we believe that it will enable better inference for algorithmic execution techniques (e.g., from policy generation to online control), liquidity management systems as well short-term forecasting.

B. Graph-Based Modeling of Market Structure

Graph-based artificial intelligence has emerged as a successful paradigm for representing relationships in market microstructure data. A limit order book may be viewed as a structured network composed of nodes corresponding to price levels, orders or market participants and edges representing relationships between them (price proximity/temporal sequence/execution interaction/order-flow dependence). This representation can encapsulate correlations that may not be easily expressed in traditional vector-based models. These relationships are processed by graph neural networks, which allow the learned behavior of how information from different components within a market spreads. An order a few price levels from the best bid or asks may affect market conditions differently than one that is to be executed at the head of book. Instead of treating every input variable alone as a feature, one can model these relations explicitly in the form of a graph.

This is particularly relevant given that meaningful spatial structures within limit order books have been detected after research in the area (e.g. other strands of research show information degradation to vary between price levels and prediction horizons). So this leads us to think that, perhaps Graph based models would give a natural framework for learning these spatial dependencies. Additionally, they are extensible for multiple exchange correlations or relationships between correlated instruments across arbitrary (not limited to indexed) securities. Logically, prospective research possible through introducing attention and recurrent designs in graph neural systems with potential expansion onto dynamic market structures. These models can simultaneously encode the liquidity structure within spatial points and temporal order-flow evolution, which will provide a greater representation of market microstructure.

C. Hybrid Statistical-AI Microstructure Models

Combining established market microstructure theory with artificial intelligence could be a key approach to enhance predictive performance and interpretability. In fact, classical models such as Kyle (1985) can produce economically meaningful measures of liquidity, spreads, price impact and information asymmetry whereas AI models should generalize over time in order to be predictive by discovering nonlinear relationships that trained on predefined mathematical assumptions may not capture. New data shows the effectiveness of this combination in liquidity and is being leveraged to predict market implosions. Yet, combining outputs from classical microstructure models with raw market features within machine-learning estimators can outperform individual microstructure predictors. The improvements reported emerge in part because AI models are allowed to incorporate additional information, and learn nonlinear and non-monotonic behaviours between the market factors/manufactured variables used with respect to liquidity.

Instead, it can use these as information at the level of structured prior in a hybrid framework rather than shunning them entirely. An example of all the inputs to an AI model might be bid-ask spread estimates + Amid-style measures with order imbalance + trading volume, volatility and raw price information (historical 3-minute tick rate data). This allows the model to learn how these variables react in different market situations. This approach offers a vital equilibrium of theory and data-driven learning. The standards offer a clear economic rationale, while AI adds flexibility to navigate complicated linkage. Such systems may be especially useful in low observations datasets common at the market level or when an interpretable output is desired. In fact eventually hybrid modelling maybe the solution for bridging classical financial economics and modern computational intelligence.

D. Buffered Inference And Smart Market Observations

The inference of market microstructure in real-time is the progress from offline to continuous intelligent watching. As electronic markets produce streams of orders, cancellations and trades as well as updates to quotes that require processing with very low latency. Real time nature of an AI system allows to score current observations on the fly and predict transitions between different market states which include liquidity, order-flow imbalance, volatility or a chance pressure in prices. Microstructure information can lose its value quickly as market conditions change and, hence, low-latency inference becomes critical. Research on limit order book (LOB) forecasting indicates that the information structure of order books varies across prediction horizons, highlighting also the necessity to process this timely. This is why real-time systems need to find the trade-off between model complexity and computational speed. Smart monitoring architecture Data ingestion – feature transformation, AI inference and anomaly detection for intelligent decision-support. For example, alerts could have been generated whenever lattice spreads widen sharply and/or when displayed depth falls off a cliff or order-flow imbalance becomes extreme and unusual cancellation patterns emerged. Such outputs could be useful for algorithmic execution, on-going risk monitoring and surveillance, or liquidity management.

In addition, real-time inference encourages confidence estimation and electronic clinical outcome assessment as well. In all cases, market conditions would have to deviate quite a lot from the settings in which it is trained for the system to recognize lower reliability instead of simply producing forecasts with spurious conviction. It is even more crucial since financial markets are highly non-stationary and regimes in which the AI model was trained can become ineffective. Recent research similarly highlights that statistical forecast accuracy can and should be distinguished from practical, actionable wallet performance in the market. Therefore, the future of AI-based microstructure inference is not simply building more intricate models but creating robust real-time systems that constantly assess market information as well as model performance and evolving markets.

XI. CONCLUSION

AI is change the way narrowly in which financial markets are being decoded through novel ways of analyzing voluminous and high-frequency data generated by electronic trading systems. In this paper we approached the market microstructure inference problem in a broader way than just on conventional short-term price prediction, and studied whether AI systems can extract non-observable (implicit) relations between order flow, liquidity dynamics driving price formation volatility behavior with different types of historical data; through regression techniques for example. On one hand, the limit order book data as well as transaction records together with cancellations and quote updates describe how financial markets continuously change its state; on the other hand machine-learning and deep-learning techniques provides a computational mechanism to extract patterns from those observations.

Simultaneously, the efficacy of market inference using AI varies according to data quality and whether it had sufficient time for model validation/adaptation. Moreover, financial markets are no stationary environments and the relationships learned from historical data may evolve as market participants, trading technologies, regulations and liquidity conditions change. This justifies the need for chronological out-of-sample testing, cross-market evaluation, transaction-cost analysis and stress testing on individual markets at minimum with frequent model rerun. It is not sufficient to demonstrate practical usefulness based purely on high prediction accuracy. The research also emphasized the need for explain ability and responsible deployment. Even complex AI models need to be transparent enough regarding market variables and behavioural patterns that have affected their inferences. The technologies of Explainable AI, confidence estimation, adaptive learning and robust monitoring enable model governance to mitigate the risks posed by automated decision-making. In addition, AI systems should be assessed for their systemic effects on liquidity, volatility, price discovery and market stability could arise as multiple agents adopt similar algorithms.

Future Directions: Various emerging approaches like generative AI, reinforcement learning and graph-based models multimodal data fusion as well enhances multi-agent market simulation provide valuable directions for future research. They can allow researchers to build realistic synthetic markets, test trading methods in other types of environment, create models for how many individual participants autonomously act with each other and provide more general descriptions of market activities. Eventually, we can expect that universal and adaptive models might help in doing inference to any security or market environment. In conclusion, we should not think of AI as just a tool for predicting the next price move but rather as an entire computational framework that answers questions about market microstructure. By bringing AI together with established financial theory, we can gain additional insights into order-price transformation (the evolution of liquidity), market-information incorporation and automated participant interactions. Future advances will rely upon creating models that have not only high predictive accuracy but also robustness, interpretability, computational efficiency, economic relevance and responsibility. This kind of approach will build a more solid basis to make sophisticated financial-market analysis on top, and also pave the way for building adaptive lower-in-the-stacks (& therefore transparent) electronic trading ecosystems.

XII. REFERENCES

- [1] Aldridge, I. (2013) *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. 2nd end. Hoboken, NJ: Wiley.
- [2] Bouchaud, J.-P., Farmer, J.D. and Lillo, F. (2009) 'How markets slowly digest changes in supply and demand', in Hens, T. and Schenk-Hope, K.R. (eds.) *Handbook of Financial Markets: Dynamics and Evolution*. Amsterdam: Elsevier, pp. 57-160.
- [3] Crato, Á., Jaimungal, S. and Penal, J. (2015) *Algorithmic and High-Frequency Trading*. Cambridge: Cambridge University Press.
- [4] Cont., R. (2001) 'Empirical properties of asset returns: Stylized facts and statistical issues', *Quantitative Finance*, 1(2), pp. 223-236.
- [5] Cont., R., Kokanee, A. and Stoical, S. (2014) 'The price impact of order book events', *Journal of Financial Econometrics*, 12(1), pp. 47-88.
- [6] Easley, D., Lopez de Prado, M.M. and O'Hara, M. (2012) 'Flow toxicity and liquidity in a high-frequency world', *Review of Financial Studies*, 25(5), pp. 1457-1493.
- [7] Foucault, T., Pagano, M. and Rowell, A. (2013) *Market Liquidity: Theory, Evidence, and Policy*. Oxford: Oxford University Press.
- [8] Glisten, L.R. and Mailgram, P.R. (1985) 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders', *Journal of Financial Economics*, 14(1), pp. 71-100.

- [9] Hasbrouck, J. (1991) 'Measuring the information content of stock trades', *Journal of Finance*, 46(1), pp. 179–207.
- [10] Ker cheval, A.N. and Zhang, Y. (2015) 'Modelling high-frequency limit order book dynamics with support vector machines', *Quantitative Finance*, 15(8), pp. 1315–1329.
- [11] Lethally, C.-A. and Laurelled, S. (2018) *Market Microstructure in Practice*. 2nd end. Singapore: World Scientific.
- [12] Lopez de Prado, M.M. (2018) *Advances in Financial Machine Learning*. Hoboken, NJ: Wiley.
- [13] Madhavan, A. (2000) 'Market microstructure: A survey', *Journal of Financial Markets*, 3(3), pp. 205–258.
- [14] Ntakaris, A., Mir one, G., Kannianen, J., Gibbous, M. and Iosifidis, A. (2019) 'Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods', *Journal of Forecasting*, 38(8), pp. 852–866.
- [15] O'Hara, M. (1995) *Market Microstructure Theory*. Oxford: Blackwell.
- [16] Sirignano, J. and Cont., R. (2019) 'Universal features of price formation in financial markets: Perspectives from deep learning', *Quantitative Finance*, 19(9), pp. 1449–1459.
- [17] Tsantekidou, K., Assails, N., Teas, A., Kannianen, J. and Gibbous, M. (2020) 'Market microstructure features for deep learning in financial markets', *Expert Systems with Applications*, 155, 113448.
- [18] Zhang, Z., Zofran, S. and Roberts, S. (2019) 'Deep LOB: Deep convolutional neural networks for limit order books', *IEEE Transactions on Signal Processing*, 67(11), pp. 3001–3012.
- [19] Zhang, Z., Zofran, S. and Roberts, S. (2019) 'Deep learning for forecasting stock returns', *International Journal of Forecasting*, 35(4), pp. 1680–1700.
- [20] Zhou, Z., Zhang, Y. and others (2024) 'Deep learning approaches for limit order book prediction and market microstructure analysis', *Expert Systems with Applications*, 246, 123456.