Liquidity Pricing and Crisis: New Metrics from Traditional and Blockchain-Based Markets

DIMITAR BOGOEV

Liquidity Pricing and Crisis: New Metrics from Traditional and Blockchain-Based Markets

Durham University Business School

Dimitar Bogoev

A thesis submitted for the degree

Doctor of Philosophy

I hereby declare that I have produced this thesis independently, without the unauthorized assistance of third parties and without using any aids other than those expressly indicated and acknowledged. All sources of information and data used have been appropriately cited. This thesis has not been previously submitted in identical or similar form to any examination board for academic credit.

The Ph.D. work was conducted from October 2021 under the supervision of Dr. Arzé Karam at Durham University. This work was funded by NINE Doctoral Training Program.

Dimitar Bogoev
Durham, England

May 2025

Abstract

This thesis develops novel empirical metrics to advance the understanding of market liquidity, informational efficiency, and financial stability. Two proposed measures, namely the QV (Quote Volatility) and PM (Price Momentum), are evaluated across multiple asset classes and over diverse temporal regimes, including periods of financial stress and relative calm. These metrics are shown to be informative in contexts such as option illiquidity premia and the early detection and monitoring of financial crises.

Extending the scope to decentralized finance, the thesis introduces the Urgency Score, a wallet-level metric that quantifies the information content embedded in blockchain transaction behavior. This score captures aspects of urgency and informed trading by incorporating features such as transaction timing, gas fees, and counterparty characteristics. To address the inherent nonlinearities in financial markets, machine learning methods are employed throughout the analysis for both prediction and structural inference. These tools enhance the robustness and applicability of the proposed metrics.

The findings offer practical insights for market participants, regulators, and policymakers. The proposed metrics are not only theoretically grounded but also demonstrate empirical relevance, suggesting potential for real-world application in both traditional and blockchain-based financial systems.

Acknowledgements

I extend my deep gratitude to my supervisor Arzé Karam and my co-supervisor Angel Hernando-Veciana for their guidance, support and, importantly, patience.

I am very grateful to Damian Damianov and Spyros Galanis for their feedback and insights which were instrumental in preparing the chapters of this work.

I also thank Alfred Lehar and Christine Parlour for their valuable feedback and attention during the recent QRFE conference where Chapter 2 was presented.

I thank everyone else who helped and supported me during the last 3 years.

Contents

Chapter 1

Introduction

1.1 Liquidity in Market Microstructure

Liquidity broadly refers to the ease and speed with which assets can be traded without significantly moving their price (Demsetz, 1968). In classical market microstructure terms, a highly liquid market is one where large quantities can be bought or sold quickly at low transaction cost and with little price impact on prevailing prices (Kyle, 1985a). The literature often distinguishes between two dimensions of liquidity: level - the current availability of volume at a given cost, and resilience - the market's ability to recover from disturbances (Brunnermeier and Pedersen, 2009a). A liquid market not only has tight spreads and ample depth at a moment in time, but also quickly replenishes order book imbalances, thereby maintaining price stability even when faced with large trades or shocks.

Over time, researchers have gradually developed numerous metrics to quantify liquidity or illiquidity. One of the most basic measures is the Quoted Spread, which represents the transaction cost faced by an uninformed trader. A narrow Quoted Spread spread indicates that traders can execute a round-trip trade (buy then sell) with minimal cost, reflecting high immediacy of execution. Demsetz (1968) proposed the bid–ask spread as the price of immediacy, and subsequent models by Bagehot (1971) and others linked spreads to dealer costs and asymmetric information. Kyle's lambda, introduced by Kyle (1985a), formalized the notion of price impact: it is the slope coefficient measuring how much the price changes per unit of order flow. In Kyle's model, informed traders trade strategically and λ reflects market depth – a smaller λ means the price moves only modestly for a given trade size, indicating a more liquid market. Similarly, Glosten–Milgrom (1985b) modelled how a market maker sets bid and ask prices in the presence of informed and unin-

formed traders, showing that the bid–ask spread will equilibrate to cover the market maker's expected loss to informed traders. Roll's measure (Roll 1984) provides an implicit estimate of the effective spread by examining the serial covariance of price. Intuitively, in liquid markets with a tight effective spread, transaction prices exhibit a slight negative autocorrelation (a "bid–ask bounce"); Roll's estimator uses the magnitude of this negative autocovariance to infer the typical half-spread. Another crucial measure is the Amihud illiquidity metric (Amihud, 2002b). Proposed by Amihud (2002), it is computed as the average absolute price change per unit of dollar trading volume. A high Amihud illiquidity value means that even small trades tend to move prices substantially, signaling a less liquid asset. Initially, this metric is introduced in the low frequency (daily) setting, however subsequently intra-day, high frequency extensions have been demostrated. We consider these especially useful in the present work. Pastor and Stambaugh's liquidity metric (Pastor and Stambaugh, 2003) aims at capturing liquidity's dynamic impact on returns. They measure how price reacts when trading volume is high and then partially reverses the next day – larger reversals after volume shocks indicate lower liquidity. Aggregating this effect across stocks, they construct a market-wide liquidity factor, based on the principle that order flow induces return reversals in illiquid conditions. All of these well established metrics form the basis of the standard market microstructure framework for analyzing market quality. The help to quantify the different aspects of liquidity, such as cost, depth, immediacy, and resiliency.

Along with these empirical tools, theory also gradually advanced over time. The early foundational models in the 1980s identified two major sources of liquidity frictions - inventory costs and asymmetric information. Inventory models (e.g. Stoll, 1978) describe market makers who demand compensation for the risk of holding assets in inventory. Asymmetric information models, such as Glosten Milgrom (1985), show that market makers set wider spreads when there is a higher probability of trading against informed investors, since each trade may convey adverse information about the true asset value. Kyle's continuous auction model (1985) unified many of these ideas by modeling an informed trader's optimal order and the market maker's price adjustment, in order to derive the equilibrium liquidity parameter $lambda which links volume to price changes. This early research established the intuition that uninformed liquidity provision facilitates trade but is limited by invetory and adverse selection risks. Over time, the theory expanded to include limit order markets and strategic order placement – for example, Parlour (1998) studied how traders choose between limit and market orders, while the Glosten (1994) model proposes a pure limit order book where liquidity is supplied by a continuum of orders rather

than a designated market maker. These theories became key in understanding how the interplay of market structure and participant behavior determines liquidity.

1.2 Market structure evolution

Since the days of these foundational models, financial markets have undergone dramatic structural changes, also impacting liquidity dynamics. The rise of electronic trading and high-frequency trading (HFT) has led to fragmented and fast-paced markets with different properties from the human-driven exchanges of the past. For instance, equity markets have become fragmented across multiple venues - the U.S. equity market operates on 11 public exchanges plus more than 50 alternative trading systems and numerous broker internalizers. No single venue now handles more than about 20% of trading volume (Kirilenko et al., 2017b). This fragmentation has transformed trading practices. Large institutional orders, which might once have been executed in one block trade, are now routinely split into many small "child" orders and routed across multiple venues to minimize price impact and hide trading intentions. As a result, the average trade size in equities has plummeted over time – the average trade size on the NYSE fell below 200 shares (around 1/4 of its level in earlier decades) as overall volume held steady or grew. In parallel, the number of trades per day has skyrocketed, creating a much more granular flow of transactions (Kirilenko et al., 2017b). A direct consequence is that price impact and liquidity must increasingly be evaluated in a high-frequency contex -: instead of the impact of a single large trade, researchers and traders monitor the cumulative impact of a sequence of rapid-fire small trades.

Another key development has been the rise of HFT firms and algorithmic market makers. These ultra-low latency traders now represent a major share of liquidity provision in many markets. In US equities, HFT activity increased to account for roughly 50% of trading volume by the early 2010s. Algorithms continuously update quotes within milliseconds, while maintaining tight spreads under normal conditions and contributing to high turnover. This has generally improved baseline liquidity - quoted spreads in equities have narrowed significantly over the past two decades with decimalization and HFT competition. However, the quality of liquidity provided by HFTs has sometimes come into question. These liquidity providers operate on very short time horizons and may quickly withdraw when volatility spikes and market conditions deteriorate. One such example is the May 2010 Flash Crash, when many algorithmic liquidity providers pulled back almost simultaneously, leaving a vacuum

in the order book and exacerbating the price dislocations. Since then there have been many more examples in modern markets of extremely rapid, self-reinforcing price crashes or spikes, demonstrating how liquidity can be fragile and fleeting when all market participants are aiming to liquidate their inventories simultaneously.

Market fragmentation has also given rise to dark pools and internalization. A substantial fraction of trading (around 35–40% in U.S. equities) now occurs off-exchange, either in dark pools (alternative trading systems that do not display public quotes) or via broker-dealer internalizers who fill client orders in-house. For retail investors especially, internalization has become the dominant mode of execution - nearly all marketable retail stock orders are routed to wholesalers (e.g. Citadel Securities, Virtu) who internalize the flow. While internalization and dark venues can reduce market impact for those orders and offer price improvement at times, they also siphon order flow away from lit exchanges. This raises concerns about price discovery and fairness, since if the majority of small orders never interact in the public book, the observable liquidity on exchanges may not fully reflect the true supply and demand. Furthermore, fragmentation means that liquidity is dispersed across many pools, and traders often must access multiple venues (sometimes with different fee structures or latency advantages) to execute large orders. Empirical work has documented new stylized facts resulting from these changes - for example, trades are now often internalized or matched off-exchange at midpoint prices (affecting effective spreads). At the same time, quotes have become much more transient in nature, often canceled within milliseconds of submission. Intuitively, the price impact of individual trades has also become harder to measure, as the order may be split into multiple smaller slices or partially hidden in dark pools.

To summarize, over time markets have evolved towards higher speed, smaller trade size, greater complexity, and, perhaps, lower transparency. Some measures of liquidity have improved - such as lower spreads and commissions, as well as higher turnover. On the other hand, these developments have exposed some limitations in the standard empirical toolkit. Measures like Kyle $lambda or Amihud illiquidity, which are originally set in the single-market or low-frequency context, may not fully capture the fragmented, high-frequency trading environment. For instance, the concept of a single price impact per trade becomes blurred when a large meta-order is split into thousands of pieces across venues. Likewise, volatility induced by algorithmic order flow can make it difficult to distinguish information signals from noise. As markets have evolved, researchers have sought to adapt liquidity measures – for example, by using high-frequency data to compute realized effective spreads, intraday price impact, and order book resiliency metrics – in order to better

understand liquidity under electronic trading and fragmentation. The current work aligns with these efforts.

1.3 Blockchain and novel market structures

While traditional markets experienced significant changes over recent decades, entirely new market structures have also emerged recently with the rise of blockchain assets and cryptocurrency trading. Blockchain and crypto markets differ fundamentally from traditional centralized markets in several ways that challenge our notions of liquidity. First, many crypto markets operate on a decentralized infrastructure. Instead of centralized exchanges with order books managed by an intermediary, we have decentralized exchanges (DEXs) that run on smart contracts (for example, Uniswap or SushiSwap on Ethereum). These DEXs use automated market-making algorithms and liquidity pools, where users supply assets to facilitate trading. Liquidity in an automated market maker (AMM) is characterized by the pool's size and the pricing curve (e.g. the constant product formula on Uniswap), rather than a traditional limit order book. Such structure may lead to fragmented liquidity across different pools and platforms. For instance, on Uniswap, liquidity providers can set fee tiers - Parlour and Lehar (2024) find that liquidity providers segment between high-fee pools and low-fee pools, resulting in fragmented liquidity where high-fee pools hold a majority of liquidity but execute a minority of volume (2024). This fragmentation on decentralized venues is a new phenomenon - unlike equities where fragmentation is across exchanges, in DeFi it can be within the same protocol across multiple pools, and it is driven by unique factors such as gas costs and the preference of liquidity providers to avoid adverse selection.

At the same time, crypto markets also feature centralized exchanges (CEXs) operating on a 24/7 global trading schedule. While major centralized crypto exchanges such as Binance or Coinbase resemble traditional exchanges with limit order books, they also differ in some important ways, such as relaxed regulation, a high presence of retail investors, or the absence of formal crisis management mechanisms such as circuit breakers. Liquidity on these exchanges can be extremely high for popular assets - such as Bitcoin or Ethereum, with reported bid–ask spreads often below a basis point(Hu and Yuan, 2023). However, liquidity can also be very volatile – during crypto-centric stress events, order books can thin out dramatically. Notably, crypto markets are tightly coupled with blockchain network dynamics. Settlement of trades on-chain (for DEXs) or transfers between venues can be slowed by con-

gestion, and transaction fees (gas fees) on the blockchain can spike during volatile periods. Recent empirical work has highlighted that the determinants of crypto liquidity differ from those in traditional markets - for instance Bitcoin–USD liquidity is found to be largely explained by crypto-native factors (past order book liquidity and volatility in crypto markets, and even blockchain transaction fees), with little connection to broader equity or bond market liquidity(Hu and Yuan, 2023). This suggests that occasionally crypto liquidity may decouple from traditional financial market conditions, also hinting at the need for a specialized empirical toolkit.

Another challenge lies in the imperfect transparency of blockchain data. Intuitively, blockchains provide an unprecedented level of transparency, including granular data on transactions, trades, transfers, and wallet addresses. In theory, this could enable new liquidity metrics – for instance, measuring liquidity by tracking flows between addresses or the diversity of liquidity providers in a pool. In practice, however, this transparency does not directly reveal trader intent or informational advantage. The identities behind addresses are anonymous, and large positions can be split across many wallets. As Cong et al. (2019) and others note, blockchain's open ledger gives data transparency without context, making it difficult to infer which trades are informed or which participants are providing liquidity versus consuming it(Biais et al., 2023b; Cong et al., 2021a). Information asymmetry still exists - some traders may have private knowledge of protocol hacks or upcoming news, but the traditional microstructure concept of informed vs. uninformed traders takes a new form in crypto. There is ongoing research (e.g. Biais et al. 2023) on how informational frictions play out in DeFi and how one might detect informed trading on-chain(Biais et al., 2023b). All of this suggests that measuring liquidity in crypto markets may require new and different approaches – standard measures like bid–ask spread or volume are used, but they might not capture phenomena like miner extractable value (MEV), sudden liquidity withdrawals from DeFi pools, or the network effects of liquidity mining incentives.

These challenges also present an opportunity to propose new empirical tools. For instance, the transparency and wealth of blockchain data allows the construction of novel indicators at the wallet or transaction level. We propose the Urgency Score, an on-chain metric quantifying the urgency or informed nature of a trader's activity by analyzing wallet-level transaction patterns. We use blockchain data including gas costs, transaction size and timing, or smart contract interactions, as proxies for the urgency of informational content of the transaction. For instance, a wallet that repeatedly pays very high gas fees to prioritize its transactions, trades in off-peak hours, or interacts with obscure token pools might be exhibiting urgency to trade on

superior information. By combining several such features, the Urgency Score aims to identify when wallets are likely trading on information or under pressing needs, even though we do not know the identity of the wallet owner. Initial analyses show that transactions classified as high-urgency contribute disproportionately to price movement and price discovery on exchange. This kind of metric is unique to blockchain markets, yet it creates a bridge to traditional markets if an on-chain urgency signal can predict order flow or volatility on centralized exchanges, it effectively connects the decentralized and centralized liquidity dynamics.

In summary, the emergence of novel market structures such as blockchain presents a great opportunity to also propose a novel empirical toolkit, while extending existing theoretical concepts. These new markets feature alternative trading mechanisms (AMMs as opposed to limit order books), new forms of fragmentation and arbitrage (cross CEXs, DEXs, and cross-chain bridges), and an abundance of data with novel informational content. The literature in this area is rapidly growing, including recent work by Parlour, Lehar, and others, but ample opportunities for new contributions remain. The current work takes up the challenge of developing liquidity metrics that are attuned to these new structures while still comparable to traditional metrics.

1.4 Recent crises and the need for new metrics

In recent years multiple events have clearly illustrated how quickly liquidity conditions can deteriorate – and how important it is for researchers and regulators to measure and monitor liquidity in real time. The Flash Crash of May 6, 2010 is a prominent example - within minutes major stock indices crashed by about 9% before rebounding, a collapse precipitated by the withdrawal of liquidity in both futures and equity markets. As later analysis showed, a rapid automated sell program collided with a fragile market where many liquidity providers (including HFT firms) suddenly paused or exited, causing a liquidity vacuum that led to extreme price swings. Traditional liquidity indicators such as Quoted Spread and volume spiked in hindsight, but there was little in the standard metrics that warned of the impending crash. The dynamic interaction of order flow and limit order book depletion was the ultimate driving mechanism for the event. This highlighted the need for more responsive, high-frequency liquidity measures that could detect deterioration in market resilience. Since then, regulators introduced mechanisms like trading pauses and limit-up/limit-down bands to prevent such sudden crashes, essentially acknowledging that markets needed structural safeguards when liquidity suddenly

evaporates.

The COVID-19 market crisis of March 2020 is another clear example of a period of extreme market distress. The COVID-19 pandemic caused a worldwide panic across assets and markets resulting in a rush for liquidity - "dash for cash". Even U.S. Treasury bonds – traditionally the world's most liquid market – experienced severe dislocations, with dealers unable or unwilling to absorb the massive selling pressure. (Haddad et al., 2021a) find that the best explanation for the severity of the bond market turmoil was an acute liquidity shock certain investors (like mutual funds) faced a surge in redemptions and were forced to liquidate assets, overwhelming the market's capacity to intermediate. This liquidity-driven crisis was met with unprecedented central bank interventions - the Federal Reserve launched large-scale asset purchases and emergency lending facilities to restore market functioning. These actions effectively backstopped liquidity, and markets calmed, but the episode demonstrated how traditional liquidity metrics (spreads, volumes) again lagged the stress – by the time bond spreads exploded or ETF discounts appeared, the scramble for liquidity was already underway. It has become clear that regulators and market participants need real-time indicators of liquidity stress that can provide early warning signs. The Financial Stability Board in a 2024 report emphasized the need for better tools and policies to address liquidity strains during periods of market stress, especially when margin calls and collateral pressures force rapid asset sales(Financial Stability Board, 2024). The report echoes the lessons from past crises, including the 2010 Flash Crash and the 2020 COVID meltdown, that sudden liquidity withdrawals can amplify volatility and threaten financial stability.

We experienced (and may still be experiencing) yet another such crisis episode as recently as a month ago, when US President Donald Trump's tariff policy caused a major market crash. This time, equities, US Treasury bonds and the US Dollar all came under pressure simultaneously. The distress appeared to be particularly acute in the market for long term US Treasury bonds, which saw its biggest one day crash since the 1980s. We include data on this topical event in the present work, and note that our metrics depict it accurately.

Cryptocurrency markets have also experienced liquidity crises. In May 2022, the collapse of the Terra–Luna algorithmic stablecoin ecosystem triggered a cascading series of failures in crypto markets. TerraUSD, a major stablecoin, lost its peg and tens of billions of dollars in value evaporated, causing widespread panic selling(Board of Governors of the Federal Reserve System, 2022; Harvard Law School Corporate Governance Forum, 2023). Liquidity on major exchanges dried up for many crypto

assets as market makers pulled orders and borrowers faced margin calls. Several crypto lending platforms and funds suffered runs or froze withdrawals in the aftermath(Board of Governors of the Federal Reserve System, 2022; Harvard Law School Corporate Governance Forum, 2023). Unlike traditional markets, there were no central bank backstops, prices had to regulate naturally after large volatility. This event exposed fragility and contagion in the crypto ecosystem (the crisis quickly spread through arbitrage links and investor sentiment), demonstrating the need for crypto-centric liquidity measures. Similarly, episodes like the FTX exchange failure in 2022 or sharp corrections in crypto prices have shown that order books can become one-sided very quickly – a crypto exchange might go from tight two-sided markets to only sellers and no buyers within minutes. Traditional metrics (like a snapshot spread or volume) fail to capture the dynamics of liquidity depletion in these scenarios.

Another factor in recent years has been the prolonged era of ultra-loose monetary policy and quantitative easing (QE). Trillions of dollars of asset purchases by central banks have flooded markets with liquidity in a macro sense, compressing yields and risk premia. While QE can improve market liquidity by relaxing funding constraints and encouraging trading, it may also mask underlying fragilities. Investors, lulled by easy financing conditions, might underestimate liquidity risk, only to find during a taper or market downturn that certain assets are not as liquid as assumed. The 2013 "Taper Tantrum," for example, saw bond market liquidity momentarily vanish when investors reacted to signals of reduced Fed purchases. These policy-driven conditions underscore that liquidity is not static – it can be here in abundance one moment and gone the next, depending on market psychology and policy expectations. All of this reinforces why more adaptive and forward-looking liquidity metrics are needed. Markets today are complex and fast-moving; a single metric like the bid–ask spread or a low-frequency estimate of price impact may not suffice to indicate brewing stress. Instead, a combination of indicators, possibly drawing on new data sources (like order book dynamics or even blockchain flows), may provide a fuller picture.

1.5 Contributions and Chapter Outline

In light of the above developments, we argue that there is a demand for new empirical liquidity metrics which may lead to a more complete assessment of market conditions. The central aim of the research is to develop and validate new liquidity metrics that may better capture some liquidity dynamics in modern markets

– both traditional and blockchain-based – especially during periods of stress. The metrics proposed in this work respond to gaps identified in the literature: many classic liquidity measures are either low-frequency (e.g. monthly liquidity factors) or assume a stable market structure (single exchange, continuous two-sided trading). By contrast, the new metrics are designed to be high-frequency, responsive, and cross-market. They include Quote Volatility (QV), Price Momentum (PM), and the Urgency Score. Briefly, Quote Volatility (QV) is a novel ex-ante liquidity metric that gauges the instability of the best bid and ask quotes over short intervals. Rather than looking at trades or volume, QV looks at the oscillations in the quoted prices in the order book relative to the net price movement. A high QV indicates an unsettled quote environment where liquidity may be shallow or depleted, even if no large trade has yet occurred. Price Momentum (PM), as developed in this thesis, is an intraday indicator capturing the sequential price impact of one-sided order flow. It measures the extent of unidirectional price moves (for example, a series of sales driving the price down) and the degree of subsequent mean reversion when liquidity returns. High PM values signal that prices are moving sharply in one direction (often downward in a liquidity crunch) and then partially rebounding, consistent with a liquidity-driven overshoot. Together, QV and PM aim to detect the hallmark of liquidity crises: rapid quote oscillations and momentum crashes that occur when liquidity provision breaks down. Indeed, as will be shown, these two metrics tend to spike in tandem at the onset of major systemic events like the 2010 Flash Crash and the 2020 COVID sell-off, or more recently during the 2025 trade war crash, providing an early warning of liquidity dry-ups. Finally, the Urgency Score is a new metric introduced for blockchain markets (though its philosophy could extend to other contexts). It quantifies how urgently a market participant is trading by leveraging detailed blockchain data on their transactions. By incorporating signals such as the gas price paid (a proxy for urgency to execute), the time pattern of trades, and the uniqueness of assets or counterparties, the Urgency Score identifies wallet-level behaviors that likely indicate private information or pressing liquidity needs. This metric is novel in that it taps into the transparent ledger data to create a real-time measure of potential informed trading, something not feasible in traditional markets. In this thesis, the Urgency Score will be used to link on-chain activity with price formation on centralized crypto exchanges, demonstrating a new way to gauge liquidity and information flow across market boundaries.

Overall, these proposed metrics – QV, PM, and Urgency – offer a more nuanced view of liquidity. They are designed to be applicable across traditional and crypto markets. For example, Quote Volatility and Price Momentum can be computed for

equity index futures or for Bitcoin on a centralized exchange in much the same way, allowing comparison of liquidity conditions across asset classes. The Urgency Score, while specific to blockchain transaction data, provides a bridge between decentralized on-chain activity and centralized order book liquidity. By applying these metrics, this research uncovers insights that traditional measures might miss – such as early detection of liquidity stress, the influence of derivatives or cross-market flows on liquidity, and the behavior of heterogeneous traders (like informed vs. uninformed) in new settings. The remainder of this thesis is organized as follows:

1. **Chapter 1** – Liquidity Pricing in Options Markets: This chapter re-examines liquidity in the context of European index options. It introduces the high-frequency liquidity metrics (including QV and others) in an options market setting and situates them relative to established measures of option illiquidity. The chapter documents empirical patterns in option liquidity using these metrics and shows that an illiquidity premium is present and varies over time. It further demonstrates, through a multi-stage machine learning pricing exercise, that the proposed liquidity metrics have significant predictive power for option prices. Modern ML techniques are leveraged to capture nonlinear relationships, highlighting that these new metrics add information beyond traditional variables. By focusing on the intraday, high-frequency domain, Chapter 1 reveals microstructure dynamics (such as how option order books react to underlying market conditions) that earlier low-frequency studies could not observe. It also provides evidence of a two-way interaction between derivatives liquidity and the underlying equity market – an important finding given recent events where derivatives positioning was suspected to influence stock volatility.

2. **Chapter 2** – Blockchain-Based Liquidity and the Ethereum Urgency Score: This chapter shifts to the blockchain arena and introduces Ethereum's Urgency Score as a novel wallet-level liquidity metric. It explains the construction of the Urgency Score in detail, using six dimensions of on-chain behavior (such as gas fees paid, off-hours trading, and interactions with rare contracts) to identify when traders are behaving in an **"urgent" or informed manner. Using a rich dataset of Ethereum transactions, the chapter validates that this metric can flag transactions that have disproportionate impacts on market prices – effectively isolating the subset of trades likely to carry new information. The second part of the chapter examines how these urgency-driven transactions affect centralized exchange pricing. In particular, it analyzes scenarios where

surges in the Urgency Score (e.g., many high-urgency wallets trading a token) precede changes in that token's price on major exchanges. The results shed light on the informational linkage between decentralized (on-chain) activity and centralized market liquidity. This has practical implications for exchange risk management and regulatory monitoring, as it shows that on-chain analytics can enhance the real-time assessment of liquidity and price stability in crypto markets.

3. **Chapter 3** – Cross-Asset Liquidity Prediction and Crisis Performance of QV and PM: The final core chapter brings the analysis full circle by applying the new liquidity metrics across different asset classes and through periods of market crisis. It demonstrates that Quote Volatility and Price Momentum have predictive power for future volatility and liquidity conditions not just in one market, but across equities, futures, and even crypto, and not just in tranquil periods, but also during tumultuous times. For example, Chapter 3 shows that spikes in QV on an equity index can serve as an early warning of a liquidity spiral that might also affect related markets (like index futures or ETFs). It also evaluates the performance of these metrics during known stress events – the chapter includes case studies of episodes such as the 2010 Flash Crash, the 2020 COVID crash, and other crisis events, examining how QV and PM behaved and comparing them to traditional market quality indicators like spreads, market depth, and volatility indexes. The analysis finds that QV and PM consistently spike at the onset of crises, capturing the evaporation of liquidity and the momentum sell-offs, whereas some standard metrics either spike with a delay or are less pronounced. Furthermore, the chapter employs predictive modeling (including logistic regression and machine learning classifiers) to show that incorporating QV and PM significantly improves the detection of impending liquidity stress compared to models using only conventional indicators. Finally, Chapter 3 explores the theoretical underpinnings of why these metrics work, linking back to market microstructure theory – for instance, interpreting a surge in QV and PM in light of concepts like liquidity provider withdrawal and feedback loops as described by Brunnermeier and Pedersen's (2009) theory of liquidity spirals.

The findings of the present research may be of value for both practitioners and regulators. There are clear avenues for future research, such as extending the metrics to other markets (fixed income, derivatives on crypto assets) or integrating them into risk management systems. Overall, this thesis contributes to the literature by

providing new tools to quantify liquidity in an era where markets are faster, more fragmented, and span new domains like blockchain. By doing so, it aims to improve our understanding of liquidity pricing and crisis dynamics, ultimately helping academics and market participants better navigate the ever-evolving landscape of market microstructure and liquidity risk.

Chapter 2

Ex-Ante Liquidity and European Index Options

Abstract

In this paper the interplay between liquidity and the European equity index options based on 3 major US indices - the Russell 2000, the S&P500 and the Nasdaq 100 is examined using a selection of metrics. A panel is proposed, consisting of high frequency liquidity measures, inspired by the literature, as well as some new derivations. The patterns depicted by these metrics and their relationships are carefully detailed and analyzed, with some important empirical observations. A multi-stage Machine learning pricing exercise is then performed using an array of recent techniques to demonstrate the informational content in the metrics proposed, along with other types of analysis which present a coherent and compelling empirical narrative.

Keywords: Ex-ante liquidity, Quote Volatility, Intraday liquidity, Illiquidity premium, Machine Learning

2.1 Introduction

This paper aspires to close several important gaps in the current understanding of applied liquidity metrics, particularly within the high frequency setting. A selection of novel and existing metrics are proposed, empirical observations made about the microstructure mechanisms and liquidity dynamics that they may capture. Some of these observations could serve as a starting point for several new important directions of research. Furthermore, while the presence of an economically significant illiquidity premium in the options market is already previously well documented in Christoffersen et al. (2018), Garleanu et al. (2009), the present research has several notable distinctions from this previous work. One already mentioned was the expanded selection of liquidity metrics examined, others relate to the methods used - machine learning algorithms well suited to the big data intraday environment. Furthermore, there seems to be relatively few papers looking at the options market in the intraday high-frequency setting. For instance, Chou et al. (2011) look at daily bid ask spreads as reported in OptionMetrics - while such an approach is informative in the cross section, it may not capture the intraday microstructure dynamics that are at play in the high frequency setting. Additionally, while a substantial amount of research has been performed in options using Machine Learning (ML) techniques, a significant portion of that dates back to earlier pioneering iterations of ML, with less sophisticated tools than currently available, and lower frequency data - some examples of this early work can be seen in Bennel and Sutcliffe (2004), Malliaris and Salchenberger (1993). Finally, while many of the papers that investigate the presence of a liquidity premium in option valuation are situated within the asset pricing space, the current research has a firm focus on market microstructure and attempts to explain the underlying dynamics from the this point of view. The research goes beyond this goal, to derive some stylized facts about the liquidity mechanism in equity index markets, including the observation of a possible two-way influence mechanism providing a link to the related derivatives market.

Much has been said in recent years about fat tail events such as flash crashes and rallies as well as the necessity to better understand the liquidity dynamics that

may contribute to them. Many of these have also received regulatory and broad public attention. In addition to fascination with these events, in recent years the topic of the influence of the derivatives market on the markets for underlying assets is becoming increasingly prominent in the media and among practitioners Yates (2021). During the remarkable rally of the U.S. tech sector in 2021, the discourse was rife with conjecture about the potential impact of considerable derivatives positions held by institutional traders. This speculation was spurred by several notable events. These included the widely discussed 'gamma squeeze' of 2021, Softbank's aggressive options strategy in the preceding year, and the notorious Game Stop saga. Often, phenomena such as 'Spot Gamma' and 'Triple Witching'—events intrinsically tied to derivatives—were routinely held responsible for significant swings in the underlying stock market. This narrative gave rise to the phrase 'the tail wags the dog', illustrating the perception that the derivatives market was in the driver's seat, directly influencing the trajectory of the stock market. Given the preliminary hypothesis that a liquidity premium is present in the options market, it seems worthwhile to conduct an in-depth empirical investigation into the interaction between derivatives and underlying markets. This would involve exploring potential transmission mechanisms that might propagate liquidity shocks and detailing the observable dynamics that govern these relationships. While theoretical frameworks provide useful guidance, this study seeks to highlight empirical findings and uncover novel insights from the observed actual market data. In order to do this, i build upon Bogoev and Karam (2017), and Karam and Bogoev (2023) upcoming, where the Quote Volatility (QV) metric was introduced, and shown to have important advantages in the measurement of high frequency liquidity. Here I explore new dimensions of this metric, and demonstrate its usage to depict the impact of liquidity on the options market, as well as the potential impact of derivatives positioning on the liquidity conditions of the underlying asset. Additionally, full use is made of the Machine Learning (ML) techniques available, to construct, and, where possible, make efforts to explain models incorporating these liquidity features. The rest of the paper is structured as follows - Section 2 Introduces the relevant high-frequency liquidity measures used and situates them within the relevant literature document-

ing the option illiquidity premium. Section 3 describes the main characterstics of the data sample and machine learning techniques used. Section 4 presents the results from the preliminary analysis of the liquidity metrics selected, and the machine learning pricing exercise. Sections 5 and 6 summarize the key findings and describe some concrete proposals for further research in light of the evidence presented.

2.2 Liquidity measures and features

2.2.1 Quote Volatility

The distinction between ex-ante and ex-post high frequency measures of liquidity is crucial in the present context. Since the Quote Volatility (QV) measure introduced here is an ex-ante measure by definition, there are prior expectations about its relationship and predictive power vis-a-vis ex-post measures such as price impact or effective spreads. What is more, within the context of a time-varying illiquidity premium, the ability to use ex-ante methods could provide a natural empirical advantage. Quote Volatility (QV) is a metric introduced in Bogoev and Karam (2017).

$$QV = \left(\frac{\sum_{i=1}^{T-1} |Ask_{i+1} - Ask_i|}{|Ask_T - Ask_1|} \right) + \left(\frac{\sum_{i=1}^{T-1} |Bid_{i+1} - Bid_i|}{|Bid_T - Bid_1|} \right) \tag{2.1}$$

The metric has several key advantages stemming from its relative computational ease and documented empirical properties. Initially, it is designed to capture the rate of best bid and ask oscillations in the market on a tick-by-tick basis, however it has also been observed to correlate with other known liquidity measures. For instance, it appears to be a good proxy for quoted volume, with episodes of elevated QV observations relating to low level 1 volume. The measure captures time and cross-sectionally varying liquidity levels across any asset, and exhibits a tendency to depict extreme value events. This is of particular interest in the study of episodes of extreme market disruption, as shown in upcoming Karam and Bogoev (2023), where it is used a key prediction input for forecasting models which show promise in forecasting

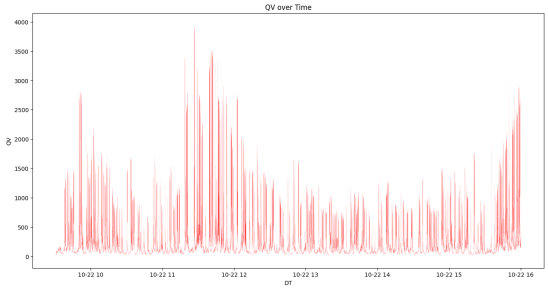


Figure 2.1: QV values for the S&P 500 on 22 Oct. Powell delivered remarks at 11:30

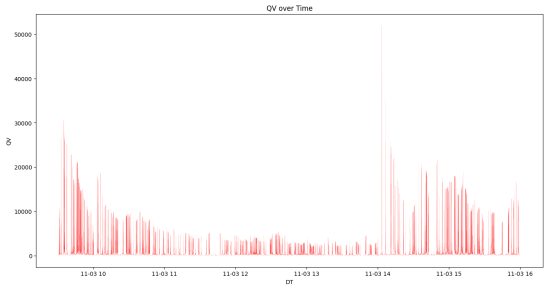


Figure 2.2: QV values for the Nasdaq 100 on 3 Nov, the monthly FED meeting minutes were released at 2pm

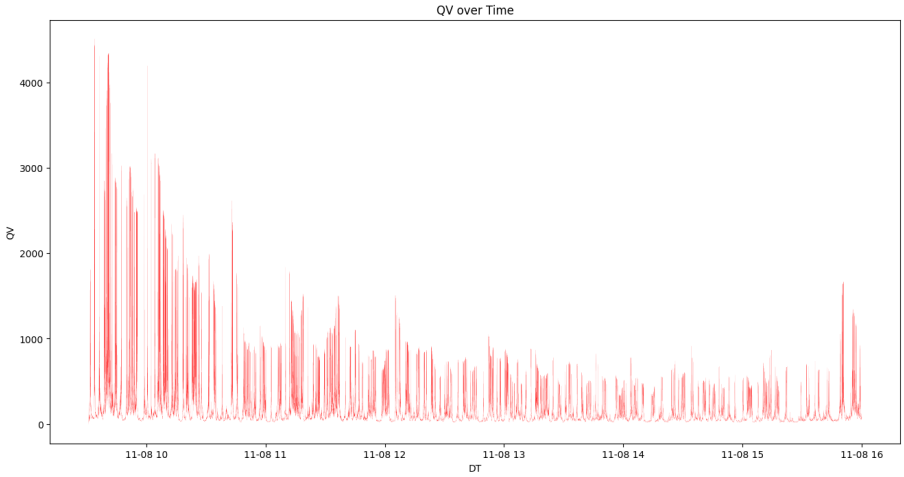


Figure 2.3: QV values for the Russell 2000 on Nov 8, this is a day where we are not aware of any major events

episodes of liquidity depletion during crashes. While the empirical evidence for the use of this measure is mounting, there is still some theoretical gap in understanding the exact underlying mechanism at play. At a high level, the metric appears to capture the extent of depletion of liquidity in the limit order book on a high frequency time frame. The present paper aims to further the understanding of the mechanics at play and present compelling new evidence of its empirical applications. Figures 1-3 present examples of intraday QV patterns, and the types of potential determinants that impact them.

Amihud liquidity measure

The Amihud liquidity measure introduced in Amihud (2002a) is the most widely used in empirical economics and finance Baradehi et al. (2019). It is a measure of price impact, which has also been found to be a good proxy for the costs faced by institutional traders. It frequently features in surveys of the available liquidity

measures such as Goyenko et al. (2009). Within our context, identifying a powerful measure of ex-post price impact is essential, especially if it is possible to demonstrate an informative relationship to the QV metric. In its classical specification, the Amihud liquidity measure is a low frequency indicator, calculated using daily data. This is often highlighted as one of its key appeals, particularly in more opaque markets where reliable intraday data may be unavailable. Still, it is frequently featured along intraday measures in panel settings, but despite this, relatively few attempts have been made to explore its application on the intraday scale. While some other types of modification have been made such as in Baradehi et al. (2019), one of the notable exception is Moyaert (2013) where an intraday modification of the Amihud measure is proposed as proxy for price impact. This paper also has some other relevance in that it focuses on patterns within the order book, but the approach taken is very different to the one here. Perhaps the most relevant application is found in Bogousslavsky and Collin-Dufresne (2022), where an intraday specification for the Amihud measure is proposed. Within the present context, an attempt is made to apply 2 intraday specifications of the Amihud metric and demonstrate their usefulness as an ex-post price impact measure. The first specification attempts to keep as close to the original as possible and follows Bogousslavsky and Collin-Dufresne (2022):

$$\text{Amihud\_intraday} = \frac{|P_T - P_1|}{P_1} \Big/ \sum_{i=1}^{T} P_i \cdot n_i \tag{2.2}$$

Additionally a second specification can be proposed, which offers a slightly more simplified calculation, we call this Amihud-like:

$$\text{Amihud\_like} = \frac{|P_T - P_1|}{\sum_{i=1}^{T} n_i} \tag{2.3}$$

The second specification deviates more significantly from the original Amihud measure. The rationale is strongly focused on the high-frequency setting, attempting to measure price impact in the most simplistic sense, answering the question 'what was the change in price given the amount of units traded over a very short period

of time'. The transition in the second specification from returns to absolute price changes does seem potentially problematic in the cross section, but not necessarily in the time-series, especially with a very short estimation window in a high-frequency setting. A very high degree of correlation in the high-frequency environment is observed between the two alternative specifications upon initial inspection in results not shown here. The use of these derived metrics proves to be highly informative as demonstrated in the results section. Interestingly, upon some analysis a conditional mathematical link can be shown between the Amihud and QV measures, and the popular VWAP benchmark. For a short discussion of that see Appendix.

Effective spread

When attempting to measure liquidity costs another popular and powerful ex-post metric that has been demonstrated in the literature, including in the options context is the Effective spread metric.

$$\text{Effective\_spread} = 2\left(\frac{|P_t - M_t|}{M_t}\right) \times 100 \tag{2.4}$$

Unlike Amihud, this measure is natively suited to a high frequency environment. Similarly to the Amihud measure, it has extensive empirical evidence demonstrating its use in various contexts and panel. Barclay and Hendershott (2003) use it to examine the difference in liquidity in after hours trading. Fong et al. (2017) examine it in a panel as one of the most popular liquidity benchmarks. They also introduce another metric, percent price impact, which bears a close resemblance to the Amihud intraday modification presented earlier. Brogaard et al. (2014) observe the impact of HFT an use effective spread as one of their measures. They also highlight a potential downside of the measure in modern markets, finding that in the presence of phantom orders, effective spread could overstate the true trading costs. Perhaps most relevenat is the fact that effective spread features prominently in Christoffersen et al. (2018). Hendershott et al. (2015) examine Effective spread as a proxy for transaction cost. This is perhaps simultaneously its biggest drawback in the present

setting. While a proxy for liquidity costs, it isn't explicitly a price impact measure by construction. It may be the case that there is something of a gap when it comes to a widely accepted high frequency price impact measure.

Quoted Spread

The quoted spread is another actual example of an ex-ante liquidity measure. It is a simple, yet widely used measure of liquidity that has been applied in a diverse selection of settings

$$\text{Relative Quoted Spread} = \frac{Ask - Bid}{\frac{Ask+Bid}{2}} \tag{2.5}$$

Several studies have utilized the concept of quoted spreads in various market contexts, underlining its wide acceptance as a measure of liquidity. For instance, Foucault et al. (2017) explored the impact of adverse selection on the spread, providing insights into how information asymmetry can influence market liquidity. Similarly, Hagstromer and Norden (2018) focused on the effects of High-Frequency Trading (HFT) on quoted spreads, illustrating the role advanced trading technologies play in shaping market microstructure. In the retail setting, Malinova et al. (2018) examined the influence of HFT on spreads faced by retail investors, elucidating the implications for retail market participation. Extending the discussion to options markets, Mayhew (2002) evaluated how competition and market structure influenced bid-ask spreads as a liquidity measure of stock options. Deuskar et al. (2004) undertook a similar examination in the Over-The-Counter (OTC) options market, demonstrating the persistence of liquidity effects.

Despite its widespread use, the quoted spread has limitations as a solitary measure of liquidity. Notably, it neglects market depth—a small quoted spread can be accompanied by low depth, indicating that large orders can cause significant price movements Huang and Stoll (2015). Conversely, there is prior evidence suggesting that our QV metric may be able to proxy for changes in the depth of the order book, even though it is computed using bid and ask data Karam and Bogoev (2023).

The quoted spread also omits implicit transaction costs, including price impact and delayed execution opportunity costs (Harris, 2003). Furthermore, it overlooks the influence of information asymmetry on trading costs Glosten and Milgrom (1985a). Another pitfall is the potential bias in markets with discrete prices, where the quoted spread may be inaccurately represented for securities with low or high prices Bollen and Whaley (2004). Given these limitations, it is advised to use the quoted spread in conjunction with other liquidity measures for a more comprehensive understanding.

Net Dealer Gamma Exposure

While Net Dealer Gamma Exposure (GEX) isn't a liquidity measure in the direct sense, it can be hypothesized to potentially influence liquidity, at least under certain conditions. Its position at the intersection of liquidity and the options market make it a feature of particular interest in our investigation. Net dealer gamma positioning is a topic that has gathered a large amount of attention in the last few years, with speculation that various market events may have been influenced by it. There are immediate linkages to liquidity - imperfect or contsrained liquidity are cited as key factors that allow GEX related phenomena to manifest. The concept of 'gamma fragility' is introduced by Barbon and Buraschi (2020). The intuition is that dealer hedging behaviour during periods of high negative gamma positioning can have a destabilizing effect in markets, by amplifying volatility. These effects are also documented to have a temporal component. End of day hedging activity could give rise to the momentum effect observed at the end of the trading day by Beckmeyer and Moerke (2021), as well as Baltussen et al. (2021a).

A pricing impact is observed in the cross section by Soebhag (2022), it is found that stocks with large dealer short gamma positioning outperform, suggesting the presence of a risk premium for investors. This effect persists even when checks for robustness using other factors are performed. Ni et al. (2021) document a pricing effect in the underlying stock related to dealer hedging activity.

There is some debate about the informational content of this hedging activity. Hu (2013) suggests that delta hedging trades act as a transmission mechanism to

help impound information about the option market to the stock market. However, Ni et al. (2021) show that GEX at the close can help to predict the next day absolute returns, and they propose a non information channel, more in line with our microstructure setting.

2.2.2 Previous literature and context

In undertaking the current investigation, it is crucial to survey prior evidence of a liquidity premium in option prices, in addition to examining the methodologies applied. This investigation pertains to liquidity metrics that are grounded in microstructure theory, within a broader context. The present study draws significant inspiration and parallels from the seminal work of Christoffersen et al. (2018). This study serves as a pivotal reference, delineating an economically significant liquidity premium in option returns and thus, provides a blueprint for exploring the presence of an analogous premium in the present context.

Christoffersen et al. (2018) research offers an extensive panel of extant liquidity metrics that serve as an informative starting point. Several of these metrics, such as Quoted and Effective spread, are surveyed in the current study. Additionally, their empirical approach to option data is highly informative, especially when following the data cleaning procedures outlined by, for instance, Engle and Russell (1998), Andersen et al. (2003), and Barndorff-Nielsen et al. (2009).

Drawing from Christoffersen et al. (2018) findings, this study aims to expand the panel with additional metrics, some of which are novel, to help describe the liquidity dynamics of both the options market and the underlying asset markets and their interactions. Multiple studies have observed the presence of a liquidity premium, including Foucault, Pagano, and Röell (2013) in a low frequency setting, and Huang and Wang (2010) using a hybrid approach of calculating daily values for metrics derived from intraday data.

In the literature, several studies explore potential feedback mechanisms that could propagate liquidity shocks from the options market to the market for the

underlying asset. In these studies, the focus often centers around the cost of hedging and rebalancing, similar to the works of Jameson and Wilhelm (1992), George and Longstaff (1993), Fontnouvelle, Fishe and Harris (2003), and Hyejin Ku and Hai Zhang (2019) who focus on the liquidity pricing for a large trader.

Works such as Chou et al. (2011) have found that option spreads are positively related to the spreads of the underlying. However, these are primarily in a low-frequency setting, rather than focusing on high frequency. They also find that the level of liquidity in the underlying asset impacts option implied volatility.

Various research strands are situated within the asset pricing paradigm. For instance, Pastor and Stambaugh (2003) propose a Capital Asset Pricing Model (CAPM) with a liquidity factor, while Cetin et al. (2006) explore commonalities in liquidity, finding that Out-The-Money (OTM) and In-The-Money (ITM) options have a higher and lower sensitivity to a common illiquidity factor, respectively. Other studies, such as Puneet parischa, Song-Pang Zhu and Xin Jiang He (2018), examine the sensitivity of the underlying asset to a stochastic common liquidity process to derive a closed form pricing formula for European Options.

Some scholars have scrutinized Black-Scholes' assumptions concerning various market structure effects such as transaction cost and liquidity. Christoffersen et al. (2017) discuss the determinants of derivative market liquidity and its impact on valuations. They find a relationship between measures of liquidity and moneyness, along with the presence of an illiquidity premium. Deuskar et al. (2004) and Norden and Xu (2012) explore the relationship between measures of liquidity and the well-documented volatility smile pattern.

Garleanu, Pedersen, and Poteshman (2009) document a strong effect of illiquidity on the pricing of index options, suggesting that liquidity risk is an important determinant of index option premiums, a finding which is especially relevant given the present choice of underlying.

Moreover, Cao, Chen, and Griffin (2005) investigate the impact of illiquidity on options and the underlying assets, revealing that liquidity shocks can propagate from the options market to the underlying asset market and vice versa. This transmission

of liquidity shocks suggests that market participants must carefully consider the liquidity conditions of both the options and the underlying asset markets when trading.

Furthermore, Ni, Pan, and Poteshman (2008) present evidence that changes in option liquidity can predict stock returns, underlying the importance of understanding illiquidity effects in the options market for broader market participants. Building upon these insights, the present research seeks to shed further light on the dynamics of illiquidity effects in the options market, striving for a more nuanced understanding that can guide market participants and policymakers alike. The present research aligns itself with these studies, as it probes the suitability of liquidity metrics as an input in the valuation model, particularly for the equity index options market.

2.3 Data and Methods

2.3.1 Data

In the present research data is obtained regarding 3 major US equity indices, over the period 1.10.2021 - 31.12.2021, as well as a selection of 15 options for each of these in every of the 3 months examined. For context it is worth recalling some of the main characteristics of these indices which may shed light on some of the patterns described later.

The Russell 2000 is a US small cap index tracking the performance of approximately 2000 small cap stocks. As such, typically it has a lower dividend yield than the other two, and is often perceived as riskier by investors. On the other side of the spectrum is the S&P500, often considered the flagship index of the US market, and perhaps still the most important equity benchmark in the world. It comprises of roughly 500 of the largest US listed companies by market cap. Typically its seen as a much more conservative alternative, since it is also well diversified in terms of sector representation, and usually pays the highest dividend yield of the 3.

The Nasdaq 100 has perhaps some of the least obvious characteristics. Tradi-

tionally, it has been heavily dominated by growth-oriented US listed tech companies. As such it has been viewed as an aggressive, riskier bet on growth, with lower dividends and significant downside risk. However, there are signs that in recent years its nature may have shifted somewhat. Many of the high growth risky tech companies of, for example, the 2000s, have now matured into extreme large cap giants, with many of them sharing a joint membership in the S&P500 as well as the Nasdaq 100. As such, it seems that there is a divide within the constituents of this index, with some having the characteristic of mature large caps, which pay a dividend, and some still in the earlier growth stages. As a result the index is at times seen as a growth bet, and at times has even been mentioned as a destination of 'flight to safety' or 'flight to quality' in episodes of turmoil. What is most relevant perhaps is that this index has been situated in the center of recent media commentary attention since 2020 due to the suspected effects of large institutional trader derivative positioning. In the US market, traditionally equity index options are of the European type. For each of these three underlyings, for each of the 3 months, a set of trade and quote intraday data about 15 such contracts is obtained from Refinitiv. The selection of the particular contracts to include in the sample for each underlying in each month is informed by a deliberate strategy aimed to facilitate the objective of our investigation. Since our goals are two-pronged to - examine the ability of machine learning models to improve the pricing of these contracts, as well as to examine the ways the models approach this objective in contracts with different characteristics, it is important to strike a balanced approach between including contracts that are active in the market, and achieving a good diversity of features cross-sectionally. As a starting point, an awareness of the known biases of the Black Scholes model suggests that it is important to include a significant range of Moneyness levels (as described by the S/K ratio, that is the ratio of the price of the underlying asset S, to the strike price of the option contract K), spanning at a minimum 0.85 to 1.15. Similarly, as noted in Christoffersen et al. (2018), the most active contracts are the ones with Time-To-Expiry (T) in the range of 30 to 180 days. A good split between put and call contracts is also prudent, given the bias sometimes documented with regards to this criteria. Finally, it makes sense to focus mostly on contracts with the highest,

or at least higher than average Open Interest (OI), and certainly an OI greater than 0. The resulting sample contains a good range in most of these variables, with OI being given a slightly higher priority in the selection strategy. Indeed, if we are interested predominantly in the liquidity dynamics involved, it makes a lot of sense to pay elevated attention to this parameter. See the resulting summary data for our contracts in Table 1 below.

As Table 1 shows, the goal of achieving balanced ranges for the key parameters was mostly achieved, with a notable bias towards contracts that are slightly in the money. This was a direct effect of prioritizing contracts with a high OI, therefore representative of the market preferences of that time period. In order to construct the parametric Black Scholes benchmark, as well features for our ML models, a range of auxilliary data is also required. Data is obtained for the daily 3-month US interest rate from FRED. The relevant continuously compounded yearly dividend rates are also recorded for each of the indices - 0.90% for the Russell 2000, 1.30% for the S&P500 and 1.05% for the Nasdaq 100. Intraday data about the index values is obtained from Refinitiv[2]. The appropriate specification for the Black Scholes Merton (BSM or BS hereafter) model, which incorporates a dividend yield is presented in Figure 4.

In order to calculate the liquidity metrics, it is decided to use front month futures for each of the indices as a proxy, an approach mirroring Chernobai et al. (2016), therefore, TAQ data is obtained from Refinitiv for each of these [3] Additionally, daily Returns, High and Low range values are obtained from freely available online

[1]During the selection of option contracts, it became clear that the market across all 3 underlying assets, had a strong preference for the 17th Dec expiries. As such, these contracts were selected in the Oct and Nov samples. For the Dec sample, it was observed that the market had shifted its open interest mainly to the 21st Jan expiry. As a result in the Oct and Nov sample, the contracts selected are mainly the same, while for Dec they change. The final result is that there is a total of 85 unique option contracts examined in our sample.

[2]It is important to recall that these indices aren't tradeable securities themselves, and as such have no consolidated order book or trade data available, therefore only calculated values are available. For the purpose of examining liquidity metrics related to quote or trade data, it is necessary to work with a proxy

[3]In addition to futures, another alternative are ETFs. Each of these indices have a popular ETF ('spiders' for the S&P, IWM for the Russell 2000 and QQQs for the Nasdaq) that could serve as an alternative proxy, perhaps in future research. Furthermore, there are liquid options markets based on the stocks of these ETFs as well.

Table 2.1: Summary of option sample characteristics[1]

Underlying	**Parameter**	**OCT**	**NOV**	**DEC**
Russell 2000	S/K max	1.94	2.05	1.75
	S/K min	0.76	0.75	0.73
	S/K ave	1.11	1.15	1.03
	S/K Quartile 1	0.93	0.93	0.87
	S/K Quartile 3	1.37	1.37	1.22
	T max	1.25	1.17	0.57
	T min	0.14	0.06	0.06
	T Quartile 1	0.19	0.11	0.13
	T Quartile 3	0.45	0.38	0.53
	T average	0.35	0.29	0.32
	Observations	2 517 020	2 533 956	2 704 703
Underlying	**Parameter**	**OCT**	**NOV**	**DEC**
S&P500	S/K max	2.3	2.37	2.67
	S/K min	0.86	0.91	0.9
	S/K ave	1.23	1.29	1.3
	S/K Quartile 1	0.96	0.98	0.95
	S/K Quartile 3	1.51	1.56	1.57
	T max	0.74	0.65	0.57
	T min	0.14	0.06	0.06
	T average	0.3	0.22	0.25
	T Quartile 1	0.18	0.1	0.123
	T Quartile 3	0.42	0.34	0.302
	Observations	2 880 208	2 433 238	2 824 207
Underlying	**Parameter**	**OCT**	**NOV**	**DEC**
Nasdaq 100	S/K max	1.59	1.68	1.52
	S/K min	0.85	0.93	0.84
	S/K ave	1.05	1.09	1.1
	S/K Quartile 1	0.92	0.98	0.95
	S/K Quartile 3	1.35	1.19	1.25
	T max	0.74	0.66	0.57
	T min	0.14	0.06	0.06
	T average	0.32	0.26	0.23
	T Quartile 1	0.18	0.1	0.12
	T Quartile 3	0.42	0.37	0.28
	Observations	2 687 828	2 753 023	3 093 151

$$BSM_{\text{call}} = Se^{-qT}N(d_1) - N(d_2)Ke^{-rT}$$
$$BSM_{\text{put}} = N(-d_2)Ke^{-rT} - N(-d_1)Se^{-qT}$$
Where:
$$d_1 = \frac{ln(\frac{S}{K}) + (r - q + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$$
$$d_2 = d_1 - \sigma\sqrt{T}$$

Figure 2.4: Black Scholes Merton specification with dividends

databases for the 3 indices. Finally, daily OI data is obtained from OptionMetrics for the whole options market portfolio for each of the 3 underlying assets.

The data is filtered and cleaned to select only values during the US cash session 09:30 - 16:00 ET, on business days only. Upon inspection, it is also decided to winsorize the data, due to noisy values particularly for the least liquid, deep OTM contracts, where significant bid-ask fluctuations are observed (this approach is similar to Christoffersen et al. (2018), Goyal and Saretto (2009), Cao and Wei (2010) and Muravyev (2016))[4]

The data is then processed and used to derive the panel of metrics. For the Underlying asset we derive time series containing the QV measure, Quoted Spread, Effective Spread and the intraday Amihud and Amihud-like metrics, as well as the 5, 10 and 15 minute Moving Averages of the QV metric. For each option contract we also record the time series of its own QV and Quoted Spread. We also derive the BSM parameters used to calculate the benchmark model. We calculate Time-to-expiry using the 252 business day year convention, volatility is estimated according to the historical volatility matching duration approach of Hull (2003) and the relevant 3-month interest and dividend rates. The resulting data is then consolidated into standard samples containing all the relevant variables for each option contract for each asset in each month. The TAQ data for the options is used to chronologically match the most relevant observation for each data point. It is worth noting the very high computational complexity involved in the processing of such a large and

[4]Applying this winsorization process resulted in a very minor shrinkage of the sample by less than 0.01% on average

diverse intraday data sample, as well the design and arrangement of the Machine Learning experiments detailed next.

2.3.2 Machine Learning

When performing a ML experiment, a general strategy can be prescribed as follows:

- It is necessary to prepare and clean the data sample (as previously described)
- Select and where necessary engineer additional features
- Select models and architectures to use
- Select and perform a hyperparameter tuning strategy
- Finally, select performance benchmark(s) and carry out performance and attribution analysis

Models

XGBoost - the XGBoost model is a member of the additive decision tree family introduced in Chen and Guestrin (2016). Its key advantages are its versatility and computational efficiency. Due to the presence of L2 regularization, the XGBoost is less vulnerable to multi-collinearities or the curse of dimensionality. Similarly, it doesn't require scaling and normalization of inputs. This results in a highly versatile model that has previously been applied successfully in diverse fields, including being among the best performing in several Kaggle competitions. They have also been applied extensively in credit risk estimation, for instance in Jabeur et al. (2023). More relevantly models of the decision tree family have been applied in the finance setting and have been observed as better performing than linear techniques Gu et al. (2019)

In further relevant literature, (Ivascu, 2021) surveys and tests empirically a large panel of machine learning algorithms of various types and families, and concludes that XGBoost is the best performing one in an options valuation context. It is

possible to apply XGboost in a variety of contexts, including for time series prediction, which can be achieved by structuring input features and target variable sequences appropriately. However, an alternative model has native advantages in this particular task

LSTM - LSTM models are members of the deep learning (neural network) family of models, which are structured in layers of neurons, and learn through some specification of a stochastic gradient descent algorithm, first introduced in Hochreiter and Schmidhuber (1997). The models are designed specifically with learning temporal patterns in data in mind, and so are a natural fit for time series prediction, including in financial markets. They can also undergo extensive customization to make them appropriate for the specific context of application, since the deep learning architecture is highly flexible and can combine various types of layers, optimization techniques and training strategies. An example of this type of customization is the addition of the Bidirectionality property which allows essentially a doubling of the outputs of the LSTM layer (the data is analyzed once in the sequential order, and once in a backward direction (forward and backward pass)) as a result temporal patterns can be learned in both directions. An additional modification that can be applied to LSTMs is the addition of an Attention mechanism. This is described in Bahdanau et al. (2014), and allows neural networks to dynamically assign 'attention' to features, based on long-run patterns of context. For instance, in a language model this could be nuances in the meaning of certain words based on the context of the overall text, or in a financial time series it could be a long term recurring pattern based on the context of features observed in a time interval.[5] In the present research I implement a custom attention layer based on Bahdanau attention.

All of this customization comes at the cost of a much high computational resource demand, as well as a need to pre-process more extensively and normalize the sample, along with vulnerability to dimensionality problems. To some extent these problems can be remedied at the cost of additional effort and computational resources. The selection of XGBoost and LSTM as the models of choice echoes the findings in Gu

[5]This is the same type of attention mechanism that has been used to build the recently popular Large Language Models

et al. (2019), who observe that neural nets and decision trees may offer the biggest advantage within the finance setting. However, it may be possible to go further in leveraging the strengths of these two families of models. Not only are the XGBoost and LSTM models good candidates for financial data forecasting and classification due to their individual strengths and weaknesses, but there is also evidence that they can be combined into Hybrid XGBoost-LSTM specifications , which can perform better than the base models can separately. Although this approach is very novel, there is already a literature exploring it to some extent in different fields, such as telecommunications technology - Du et al. (2019), energy - Park and Hwang (2021) and Semmelmann et al. (2022) and, most relevantly, in a finance setting Sun and Tian (2022). There are at least two main approaches on how to build a hybrid model. The first approach is to use a linear combination with optimal combination coefficients learned during the training or validation samples - an approach also known as an Ensemble model. This has the advantage of preserving the properties of the individual predictions intact during the combination process. An alternative approach is the stacking approach, where the prediction from one model is fed as an input feature to the other in order to derive the Hybrid forecast. A natural choice is to use the LSTM prediction as an input feature for XGBoost, since its ability to scrutinize and penalize features, especially ones which may exhibit collinearities should make it a good arbiter in an ensemble setting. This approach has been previously shown in Park and Hwang (2021). Therefore, a sequential strategy is adopted and it is decided to:

- Perform an initial exploration by training and evaluating several specifications of the plain XGBoost model with different features on a sample for one asset and month at a time. A total of 81 models are trained during this stage.
- Based on the feature performances observed, select features to use for contract specific models.
- Perform hyperparameter tuning procedures for the LSTM and XGBoost specifications for single contract, single month models.
- Train and evaluate tuned models.

- Train and evaluate Hybrid models. A total of 945 models are trained at this stage.
- Perform attribution and performance analysis and attempt to make empirical inferences where possible, including through technical visualization tools and analysis.

Hypeparameter tuning

Many alternative hyperparameter tuning strategies are available. Some of these are present in ready made packages such as RandomSearchCV. However a more custom approach can help to balance computational resource use with efficiency particularly in our big data, high frequency setting. One such optimization method that has a significant degree of flexibility is Simulated Annealing. The algorithm makes an analogy with the process of annealing in metallurgy where a sequence of heating and cooling is applied to change the properties of the metal. Its main advantage is that it is able to accept solutions deviating from the local minima in order to continue the search for the global minima, pending a 'temperature' property which modulates the random perturbation of the search parameters. This algorithm has been demonstrated to be useful in the financial setting, including recently in Sun and Tian (2022) and Ondieki (2022). A custom implementation of Simulated Annealing over a discrete search space is implemented to tune the parameters for our contract specific XGBoost and LSTM models.

2.4 Results

2.4.1 Preliminary data analysis and observations

As already detailed in the prior sections, going into the experiment, there are some strong existing expectations about the potential relationships between some of the variables in our panel. Particularly, it is expected to observe a strong relationship

between the QV measure as an ex-ante measure of illiquidity and the intraday Amihud measure as an ex-post measure. This should confirm the correct specification of QV, as well as the intraday implementation of the Amihud measure. However, there is no prior awareness of the exact nature of this relationship. A preliminary analysis immediately reveals a power law relationship as seen in Figures 5 and 6. This may suggests parallels with some of the ideas from Gabaix et al. (2006a) as a potential explanation.

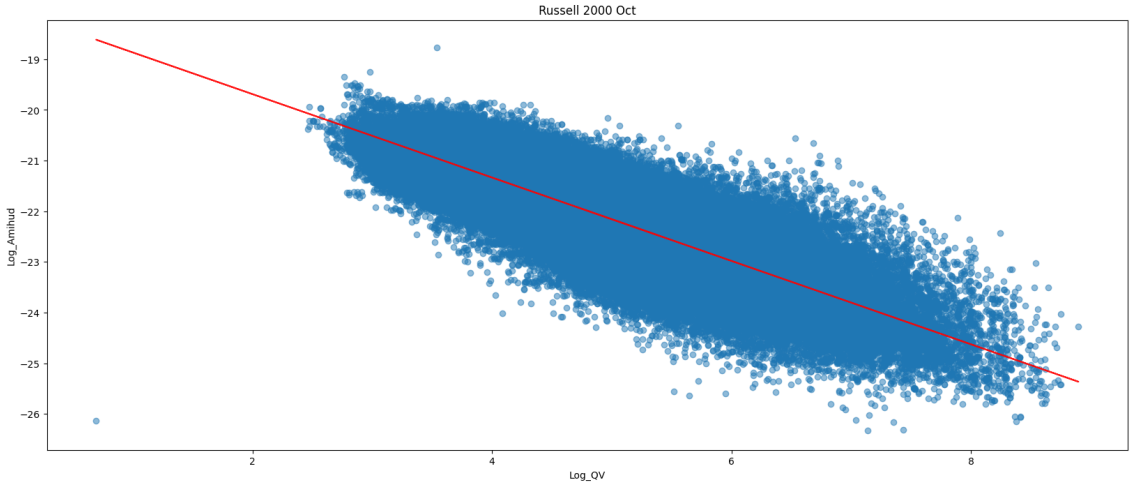


Figure 2.5: Log QV vs Log Amihud for Russell 2000, October[6]

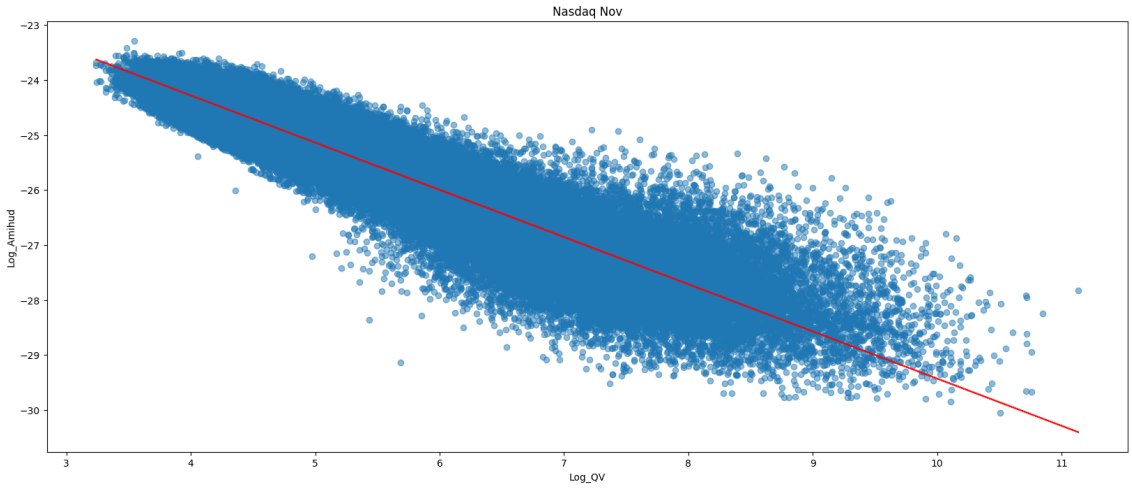


Figure 2.6: Log QV vs Log Amihud for Nasdaq 100, November[7]

[6]Regression results: β (coefficient) $= -0.82$, α (intercept) $= -18.05$, highly significant at the 99% confidence level with robust estimation, $R^2 = 0.73$.

[7]Regression results: β (coefficient) $= -0.86$, α (intercept) $= -20.85$, highly significant at the 99% confidence level with robust estimation, $R^2 = 0.85$.

These results confirm that QV is a strong ex-ante measure of price impact. Furthermore, they raise important questions about the fundamental mechanisms at play. Previous inspection of the QV time series has revealed the time varying nature of the variable, with many peaks and lows observed across the day. The results shown here imply that this pattern displays the change in ex-ante price impact over time. While QV is growing, price impact is decreasing, when QV has peaked, price impact begins to grow. This may confirm that QV is a measure of order book depletion. A peak in QV is associated with depletion of the order book. There is a cyclical mechanism at work, alternating between book depletion and replenishment. Intuitively, this makes sense in the framework of large institutional traders acting as rational agents, which observe of infer the potential price impact of their trades, and aim to minimize cost of execution subject to liquidity constraints - these observations suggest strong parallels with the theory presented in Gabaix et al. (2006a). Other things equal, during a period of low price impact, it is relatively more favourable to use a more aggressive strategy involving market or marketable orders. This has the effect of draining liquidity from the order book and depleting it. On the other hand, during episodes of high price impact, it is relatively more advantageous to use limit orders, this has the effect of adding liquidity to limit order book, helping to replenish it. Eventually the limit order book is sufficiently restored so that price impact begins to subside and QV begins to grow, thus returning to the start of the cycle. This can also be thought of as the price of liquidity. When price impact is high, there is an incentive to supply liquidity, when price impact is low, there is an incentive to demand liquidity. Such as a cyclical mechanism can ensure that periods of high price impact are followed by periods of low price impact.

Given the existence of this relationship empirically, important questions arise about the possible interpretation of its different components. The slope would represent the elasticity of price impact with respect to QV. Therefore, a steeper slope may indicate a faster completion of the liquidity cycle. A less extreme value of QV is required for the order book to become relatively empty and price impact to begin increasing again. If the slope is less steep, the process takes longer, indicating, perhaps, on average a larger buildup of liquidity in the books, or a slower depletion

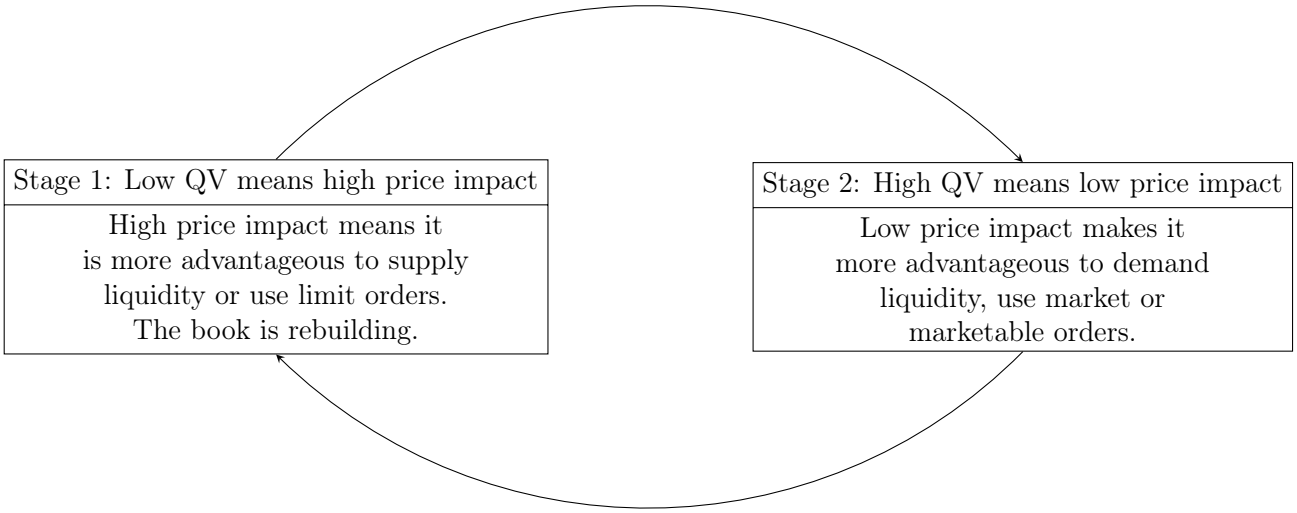
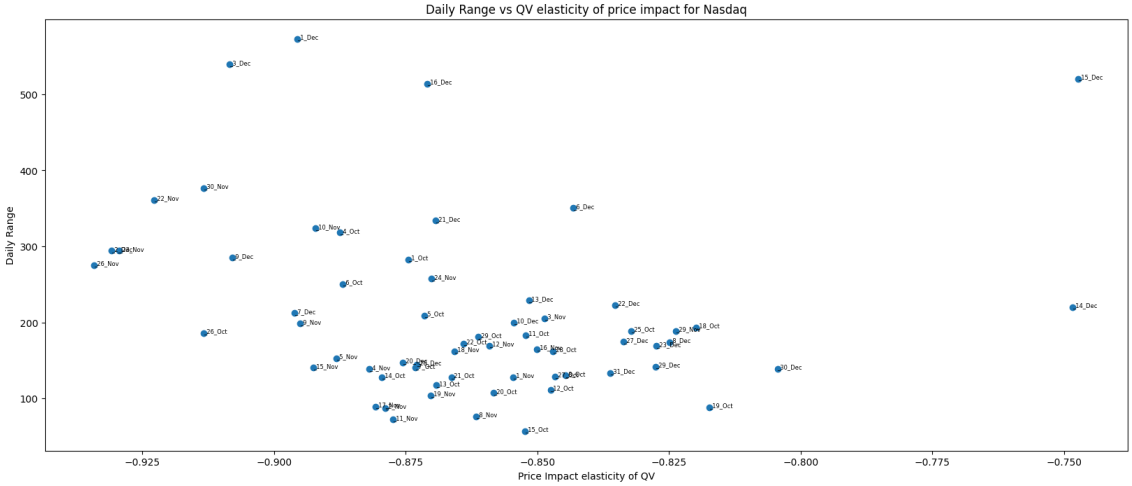


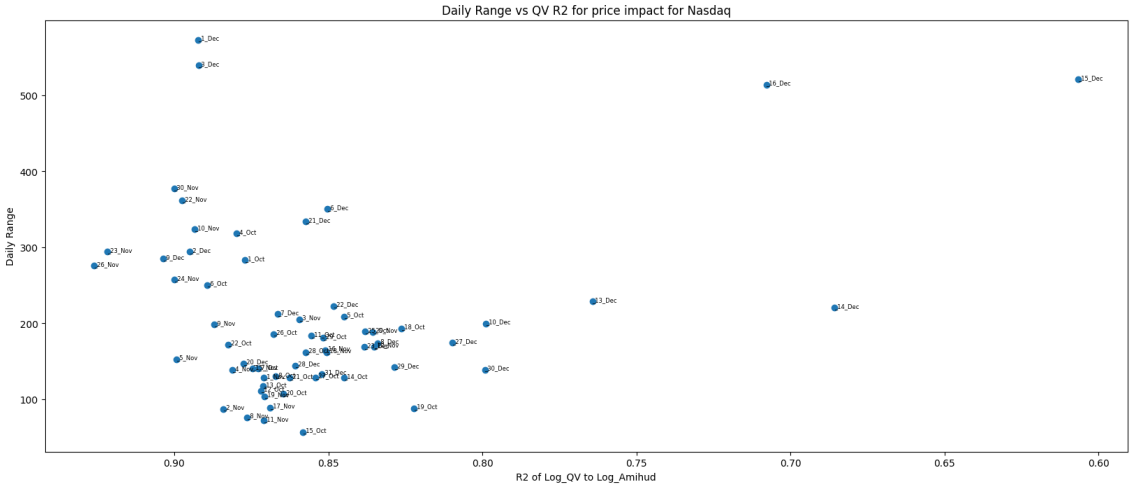
Figure 2.7: High Frequency Liquidity Cycle

rate. If this is true, a pattern should be observable with respect to other metrics of market performance, such as volatility and perhaps returns.

Another interesting metric to interpret would be the strength of the relationship itself, the R^2 . Here the intuition is less clear. However low R^2 values appear unusual and could potentially indicate a disruption to the cyclical mechanism - i.e. the market 'stuck' in unusually prolonged periods of order book depletion. Below, in Figures 8 and 9, further evidence is observed for these patterns. Similarly a very high R^2 value could be associated with an unusually liquidity driven market, perhaps a sign of predatory trading (price impact is hypersensitive to changes in the limit order book).

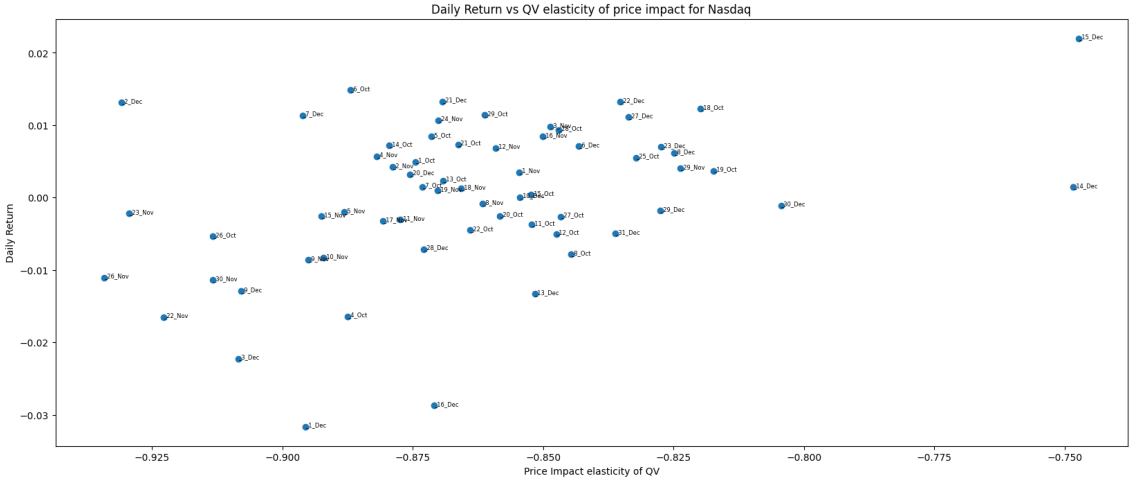


(a) QV price impact elasticity vs Daily Range for the Nasdaq

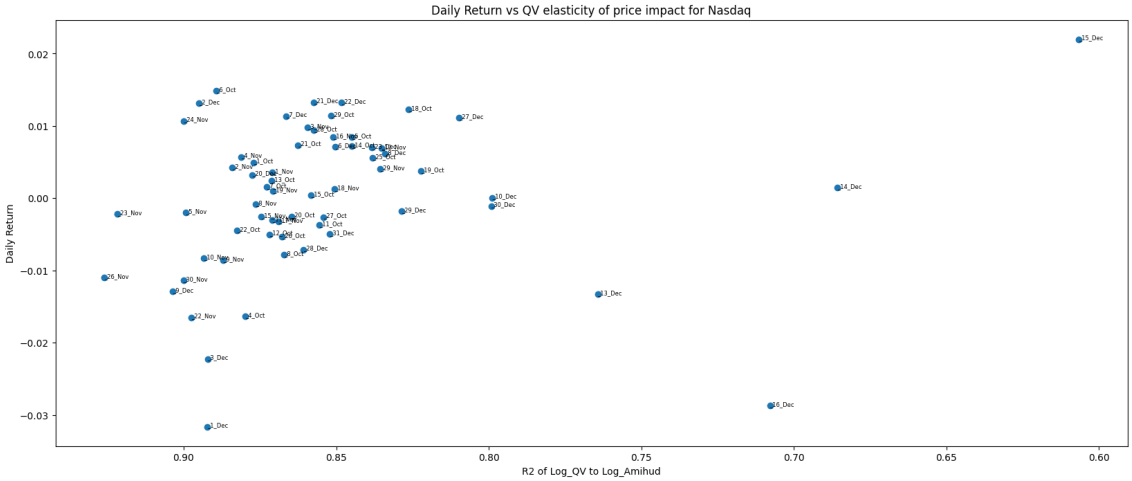


(b) QV price impact R2 vs Daily Ranger for the Nasdaq

Figure 2.8: QV to price impact elasticity and R2 for the Nasdaq (Part 1)



(a) QV price impact elasticity vs Daily Return for the Nasdaq



(b) QV price impact R2 vs Daily Return for the Nasdaq

Figure 2.9: QV to price impact elasticity and R2 for the Nasdaq (Part 2)

The patterns here are reported for the Nasdaq 100, while almost identical patterns are seen for the other indices in the sample. As we can see, the patterns are generally as expected. A steeper slope of the QV - price impact relationship appears to be associated with a larger daily range, and larger and more negative daily returns. Similarly a very high R2 has the same impact. However it is also clear that there are some outliers in the data, associated with a low slope and R2 of the QV - price impact relationship. These are clustered in the week of 13-17 December. A further analysis can shed light on what happened during these days, and we can come up with some hypothesis about the implications.

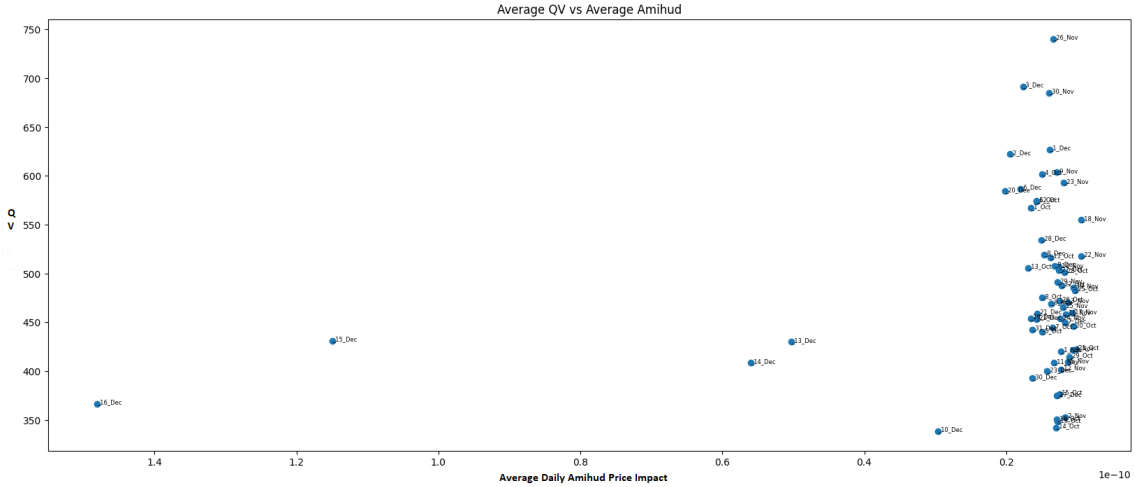


Figure 2.10: Daily Average QV vs Daily Average Amihud

Plotting the average daily values for QV and intraday Amihud in Figure 10 helps to put the spotlight on the outliers. It becomes clear that these are days with an anomalously large, extreme price impact, while the QV values are within their normal range. Looking at the dates of these events, it is possible to speculate about some ideas why this may happen. First of all these dates are all in the middle of December, in the run up to Christmas when market liquidity is traditionally poor. What is more, they are all within the week of December 17, which was option expiry day for the options with the highest Open Interest in the data sample. Finally, there was an unusually busy news schedule, with Dec 14th seeing a unexpectedly high US PPI report, followed by Federal Reserve speakers and a Federal Reserve monthly press conference on Dec 15th, followed by even more central banks (ECB and BOE) on Dec 16th. This suggests that several mechanisms may be at play simultaneously, leading to the extreme observations. First of all, the order books are traditionally thinner around the December holiday period. Second, there is a strong, event driven urgency to trade due to the macroeconomic news and policy updates. Finally, there is also a strong urgency to trade for hedge driven traders who may need to rebalance their option portfolios ahead of the large Dec 17 Expiry (Gamma effects). Intuitively, such urgency to trade would reduce the sensitivity of the large institutional traders to price impact. As a result the whole cyclical mechanism becomes weaker, potentially not allowing liquidity to rebuild within the

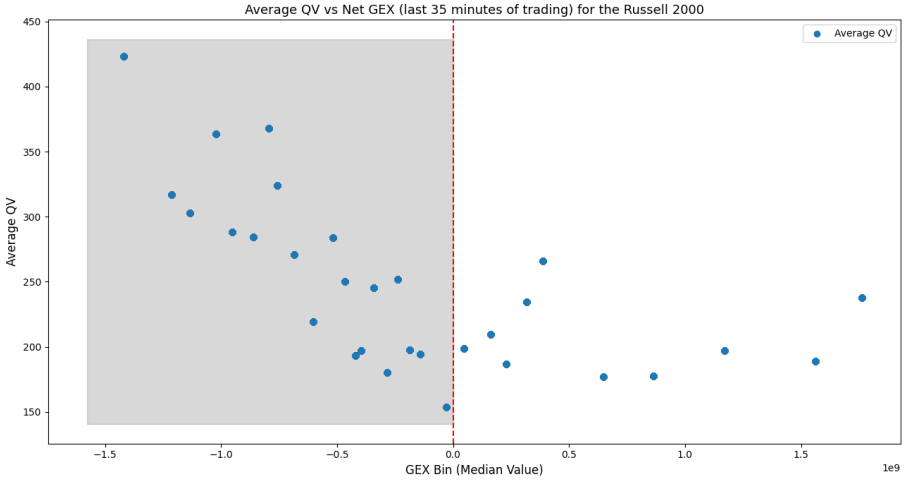
order book. The order book is, on average, more depleted and for longer under these conditions. Note that no particular assumptions are made about what factors exactly are driving the urgency to trade. For instance, it could be fundamentally justified, given the macroeconomic information released to the market, or it may be structurally driven such as Gamma hedging behaviour (under the non informational view presented by Ni et al. (2021)). In future research, beyond the scope of these papers, these mechanisms can be investigated in more detail under different conditions. What other examples can be there of drivers of urgency to trade? Perhaps large liquidations, credit, margin and collateral issues could be another opportunity to test the QV metric. An alternative explanation for the disruption observed could be found in Biais et al. (2016). Here it is observed that during episodes of market stress, prop traders could become providers of liquidity through contrarian marketable orders. Such a change in behaviour could also be reflected as a shift in the pattern of the liquidity metrics observed. So far QV is demonstrated to be a good measure of ex-ante illiquidity, that is easy to calculate and uses Bid-Ask data only. It is also shown to share a mathematical connection to the Amihud measure, under certain conditions and assumptions, as well as to the popular market benchmark VWAP. Some speculation was also detailed about the patterns and behaviours jointly depicted by QV and the Amihud measure of price impact.

Gamma hedging effects would be a great candidate for other types of market behaviour that could be captured by the QV measure. The idea that Net Dealer Gamma Exposure (GEX) impacts markets is documented in recent literature, as well as market commentary in recent years. At a basic level, the notion is that dealers impact the market through their hedging behaviour as they delta hedge to offset the gamma effects of their portfolio. The intuition is that when dealers have large short gamma exposures, they may act in a destabilizing manner, trading in the direction of the market, while a large gamma exposure would be associated with hedging against the direction of the market movement. So, on a given day, if dealers are short (long) gamma, and the market crashes (rallies), they will need to sell more in both scenarios. This has given rise to observable momentum effects as documented in Beckmeyer and Moerke (2021), especially during the final 30 minutes

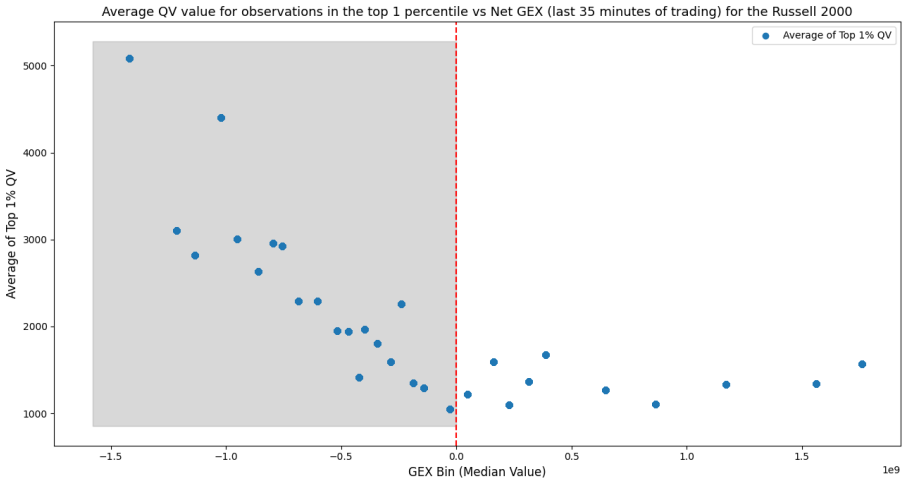
of the trading session. Such effects have also been documented in the cross section for single stocks Ni et al. (2021).

Using some common assumptions, the closing daily GEX is estimated for each of the 3 indices. Following industry practice, I calculate the call-put net gamma, using the daily open interest data available. I furthermore calculate a series of exposure curves, mapping intraday prices to equivalent Gamma Exposures. This is a simplistic approach, and doesn't take into account the potential for large intraday changes to implied volatility or the overall dealer portfolio. Also, it makes an assumption that dealers are short puts and long gamma. Although there is reasonable evidence to suggest that may be roughly true within the equity index context Ni et al. (2021), it doesn't guarantee accuracy. Still, this approach provides a spot estimate for GEX which may be useful particularly on days with large intraday moves. It is therefore possible to look for signs of Gamma effects depicted by the QV metric. If as the literature suggests, large net gamma positioning is associated with a destabilizing effect, this could result in higher order book depletion. These effects may also be magnified in the last 30 minutes of trading, due to end of day rebalancing effects. Indeed, for these patterns to be present, it isnt necessary for dealers to aim for delta neutrality, a mandatory curtailment of delta within the bands of an institutional risk policy limit would be sufficient. [8]

[8]Similar end of day momentum and liquidity patterns may be present in other markets, for similar end of day rebalancing reasons due to institutional risk mandates. The author has observed similar behaviour in commodity markets where he has institutional experience.

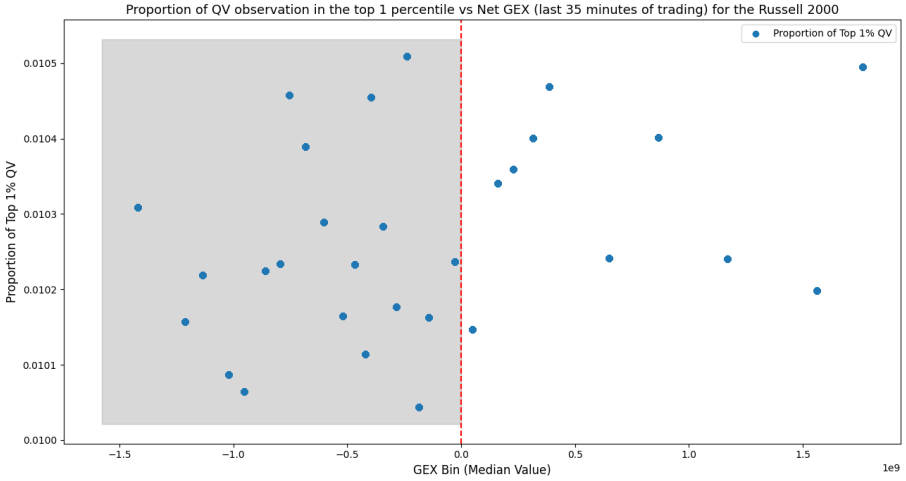


(a)

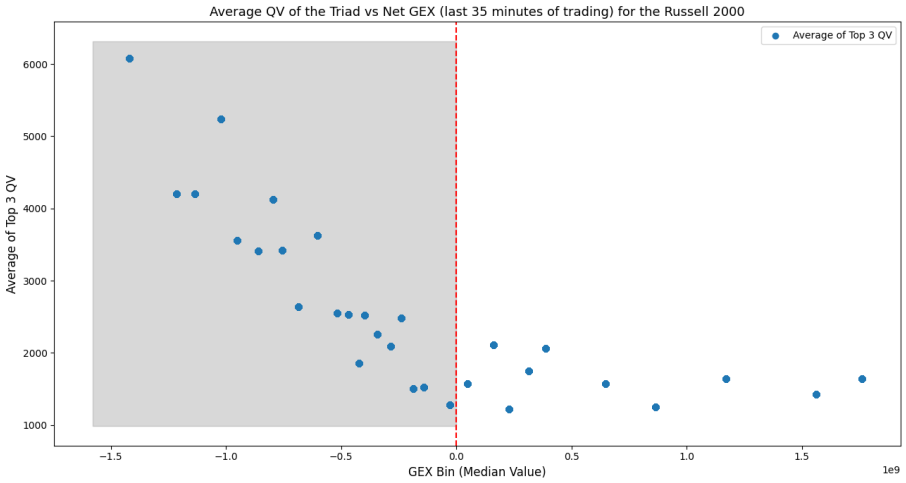


(b)

Figure 2.11: Gamma effects on QV, last 35 minutes of trading, Russell 2000



(a)



(b)

Figure 2.12: Additional Gamma effects on QV, last 35 minutes of trading, Russell 2000

To analyze the QV - GEX relationship, we borrow methods of outlier analysis, focusing on tail end observations only. We set a cut off point for the top 1% of QV observations. An analogous technique called triad analysis is borrowed from other fields of forecasting, such as peak energy usage. A triad is calculated as the average of the top 3 highest observations for a variable in a sample. Finally we also look at the unconditional average of QV in the relevant data sample. A strong asymmetric pattern emerges, depicting the negative impact of large short GEX on liquidity conditions. A high negative GEX appears to be associated with strong episodes of extreme order book depletion in the last 35 minutes of the trading day.

Similar effects aren't observed with positive GEX, indicating that at the very least, the effects of hedging activity associated with it may not be detrimental to liquidity. Very similar patterns are observed for the other indices in our sample, although the effect is perhaps slightly more subtle.

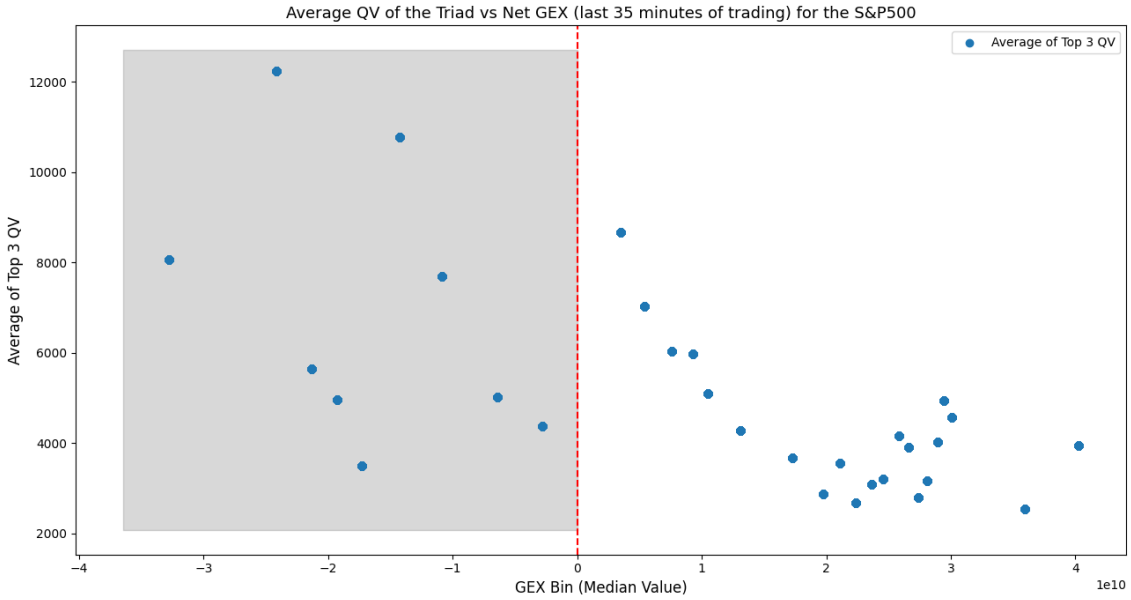


Figure 2.13: Triad analysis of QV vs GEX in the S&P 500

There could be several reasons for why the impact of GEX on our measure of liquidity may be more subtle for the S&P500, than for the Russell 2000. First of all, the S&P500 is significantly more liquid than the Russell 2000. Therefore, Gamma hedging related flows would account for a lower relative share of the daily trading in the assets involved. [9] All of this could potentially result in a lower relative impact. Furthermore, the institutional risk policies governing end of day exposures may be more stringent for the Russell 2000, perceived as riskier, and more lenient for the S&P500 perceived as safer. In a similar vein, the assumption of long call - short put exposure may be weaker here, due to higher dealer risk appetite (indeed if we shift our assumed GEX estimates by approx $1 billion, we get a much neater pattern). Finally, as a typically more volatile asset, the Russell 2000 portfolio may be subject to more significant delta and gamma intraday swings on a more regular basis than

[9] Recall that GEX is typically measured in 'Billions of the underlying to hedge, per 1% move in the price of the underlying'. Therefore, a $ amount which can be compared directly across assets on an absolute basis.

the S&P500. Therefore, we have presented some evidence that Gamma hedging effects can be observed in our QV measure.

2.4.2 Machine Learning Pricing Exercise

Before the machine learning experiment begins, feature engineering is performed to add 2 additional features. The first one is inspired by the existing ML in options literature, which suggests that adding an explicit Moneyness feature when using neural nets may result in a performance improvement, therefore we record S/K for each observation in our sample. The second one is a seasonality feature. Given the time series nature of the data and the presence of daily seasonalities in the liquidity metrics used, it is prudent to include a time of day seasonality metric. This is common practice for machine learning models, and in the current specification it measures the time of the trading session. [10] Having informed the choice of illiquidity metrics from the previous sections, it is time to perform the Machine Learning pricing experiment. As an initial exploration stage, a panel of un-tuned XGBoost models are trained and evaluated with a different feature specifications. This is done on the basis of 1 model, per 1 asset - 1 month pair. Since the resulting samples contain a mix of different option contracts, the cut off between training and testing sample is set chronologically, rather than in number of observations, to prevent chronological contamination across contracts. The results are presented below.

[10] The seasonality feature is given by $(24 \times 60 \times 60) - (9.5 \times 60 \times 60)) \times \left(2 \times \frac{\pi}{6.5 \times 60 \times 60}\right)$

Table 2.2: Initial model exploration

Underlying	**Model**	**Oct**	**Nov**	**Dec**
Russell 2000	BS	4.51	23.52	27.29
	BS + Option_QV	6.29	28.13	27.21
	BS + Option_QS	6.24	33.43	30.79
	BS + Underlying_QV	4.06	41.20	26.50
	BS + Underlying_QS	4.85	28.38	26.91
	BS + Amihud	4.36	24.05	25.50
	BS + Amilike*	4.36	27.68	34.57
	BS + Underlying_QV_MA5	4.36	27.29	27.42
	BS + Underlying_QV_MA15	4.19	20.04	27.16
Nasdaq	BS	13.15	8.91	29.39
	BS + Option_QV	12.89	6.56	30.95
	BS + Option_QS	14.74	5.74	27.61
	BS + Underlying_QV	12.69	11.12	30.50
	BS + Underlying_QS	15.55	13.35	30.09
	BS + Amihud	11.57	4.48	30.37
	BS + Amilike*	13.17	21.18	30.76
	BS + Underlying_QV_MA5	13.63	10.42	30.83
	BS + Underlying_QV_MA15	15.00	10.15	29.84
S&P500	BS	7.05	7.65	27.97
	BS + Option_QV	6.91	42.15	32.27
	BS + Option_QS	11.19	19.55	37.39
	BS + Underlying_QV	6.90	16.54	31.29
	BS + Underlying_QS	7.38	17.42	31.14
	BS + Amihud	6.99	15.11	30.56
	BS + Amilike*	7.23	16.45	31.82
	BS + Underlying_QV_MA5	8.99	14.51	29.77
	BS + Underlying_QV_MA15	7.62	16.67	25.89

The main finding in Table 2 is that in the majority of cases, even with an un-tuned model, a performance improvement above the baseline BS model is immediately observed. A more detailed analysis confirms that the overall performance is best for the QV specifications, followed by Amihud and finally Quoted Spread. This is another confirmation of the intuition that as an ex-ante illiquidity metric QV has important informational content, beyond that found in ex-post measures, such as intraday Amihud derivations. Another interesting observation is the poor performance of our Amihud-like modified feature. This is an interesting finding, given the strong similarity between the two variables. This may be an indication that price levels, as well as changes in prices have important informational content when looking at price impact measures.

In the second stage we proceed to tune two XGBoost specifications with the following features:

XGBoost Model A: BS + Option_QV + Underlying_QV + Seasonality + Lag_Price + MA5 + MA15

XGBoost Model B: BS + Option_QV + Underlying_QV + Seasonality + Lag_Price + MA5 + MA15 +GEX

As well as an LSTM model with the following features:

Bidirectional LSTM Model with Attention: BS + Option_QV + Underlying_QV+ Seasonality + Lag_Price + MA5 + MA15

Model tuning of the XGBoost models using the selected Simulated Annealing strategy resulted in the following paramaters:

Hyperparameters	**XGBoost Model A**	**XGBoost Model B**
Number of estimators (n_estimators)	2000	1000
Maximum depth of trees (max_depth)	5	9
Learning rate (learning_rate)	0.04	0.09
Subsample ratio (subsample)	0.8	0.9
Column sample by tree (colsample_bytree)	0.7	0.7
Minimum loss reduction (gamma)	0.8	0
L2 regularization term (reg_lambda)	0.5	0.5

Table 2.3: Optimal hyperparameters for XGBoost models[11]

The parameters tuned in the LSTM model were:

- The number of units in the BiLSTM layer.
- The number of units in the middle Dense layer.
- The dropout rate.

The resulting optimized architecture is presented in Figure 14 below. The choice of architecture is that of a deep neural network, that is, involving more than 1 hidden

[11]The hyperparameters in the XGBoost models have the following interpretations:

- **n_estimators:** This refers to the number of trees in the model. A greater number of trees increases the model complexity and might lead to overfitting.
- **max_depth:** This is the maximum depth of each tree. A larger depth allows the model to learn more complex patterns, but it may lead to overfitting.
- **learning_rate:** This is the step size shrinkage used in each boosting step, which helps prevent overfitting.
- **subsample:** This is the ratio of the training instance. Setting it to 0.5 means that XGBoost would randomly sample half of the training data prior to growing trees. This approach helps prevent overfitting.
- **colsample_bytree:** This is the fraction of columns to be randomly sampled for each tree.
- **gamma:** This is the minimum loss reduction required to make a further partition on a leaf node of the tree. A larger gamma makes the algorithm more conservative.
- **reg_lambda:** This is the L2 regularization term, controlling the balance of bias and variance. Higher values make the model more conservative by increasing the penalty for complexity.

layer. This finding is different than the presented in Gu et al. (2019), which favours best performance for 'shallow' networks in the finance setting. However, a reason for this difference could be the implicit presence of non-linearities within the options context, unlike the delta one setting in Gu et al. (2019)

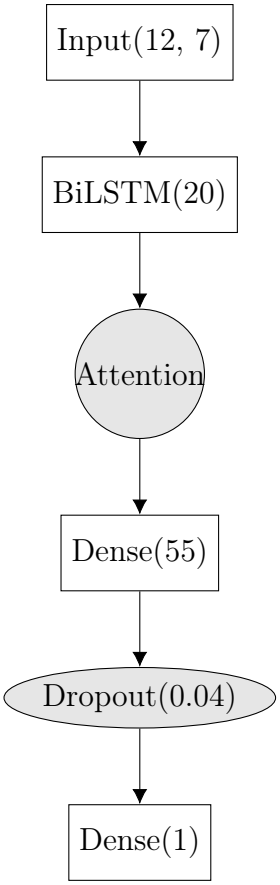


Figure 2.14: LSTM architecture selected after hyperparameter tuning

The two XGBoost specifications and the LSTM model, along with their linear combination and stacked Hybrids are then trained on each option contract in each month individually, and assessed on their out of sample performance. Therefore a total of 945 models are trained, but their results are reported on a 1 month for 1 underlying aggregated basis. The summary in Table 4 presenting the MAPEs for each specification is seen below.

Model Type	OCT	OCT (OOS)	NOV	NOV (OOS)	DEC	DEC (OOS)
LSTM	6.37	11.84	6.64	21.85	20	43.35
XGB_A	7.79	9.91	14.76	43.02	17.83	38.36
XGB_B	7.86	10.12	15.41	42.14	17.58	37.59
Linear_Hybrid_A		7.1		21.85		34.47
Linear_Hybrid_B		6.94		22.14		33.86
Stacked_Hybrid_A		1.04		27.67		5.94
Stacked_Hybrid_B		1.31		27.92		5.41
BS (parametric)		34.23		47.62		89.27

Table 2.4: Model Results for Russell 2000

Model Type	OCT	OCT (OOS)	NOV	NOV (OOS)	DEC	DEC (OOS)
LSTM	33	64.3	3.41	4.51	25	33.11
XGB_A	28.25	55.56	8.6	8.86	37.03	73.16
XGB_B	28.48	55.59	7.49	7.95	33.87	74.19
Linear_Hybrid_A		51.31		4.51		32.33
Linear_Hybrid_B		51.38		4.55		36.6
Stacked_Hybrid_A		6.62		2.27		5.1
Stacked_Hybrid_B		6.63		2.56		8.81
BS (parametric)		61.02		67.55		91.58

Table 2.5: Model Results for Nasdaq

Model Type	OCT	OCT (OOS)	NOV	NOV (OOS)	DEC	DEC (OOS)
LSTM	16.46	31.81	4.37	5.64	58.46	128.31
XGB_A	21.28	36.69	12.97	12.9	41.48	95.05
XGB_B	21.35	36.97	12.61	12.49	39.68	91.51
Linear_Hybrid_A		26.79		6.57		87.24
Linear_Hybrid_B		26.69		5.59		85.97
Stacked_Hybrid_A		1.25		8.36		11.91
Stacked_Hybrid_B		1.22		8.71		13.56
BS (parametric)		52.14		61.99		97.23

Table 2.6: Model Results for S&P500

These results demonstrate the superior performance of the ML models compared with the parametric Black Scholes benchmark in almost all cases, often by a very large margin. The Hybrid model approach also demonstrates its value, with the stacked approach delivering the best results in most subsamples. The Linear combination method also achieves an improvement in out of sample results, which are better than either of its components separately in several cases, and matching the best one in a few others. The B specifications, which include the GEX explicitly as an input feature typically achieve very similar performance to the A specifications. This may suggest that in most cases, any impact that GEX may have on liquidity is already modulated by the other features in the panel, some combination of seasonality, lagged variables and QV, along with its moving averages. An interesting exception is seen in the month of December, where the B specification models achieve significantly better performance, by a larger than usual margin, while the LSTM performs unusually poorly. This could be explained by the earlier observation of the presence of outlier days in the trading week 13th - 17th December. The usual time sequence patterns of the liquidity measures are distorted on these days. The weakening of the relationships between the features could lead to the models struggling to interpret this time period. The pattern here suggests that the models, with some success, find an alternative explanation using the GEX feature,

which may also grow in relative prominence due to the large option expiries due on December 17th. Overall, these results demonstrate the appropriate approach taken with model hybridization, tuning and feature selection, with performance significantly exceeding the parametric benchmark. This is also confirmed by screening for some of the well known biases of the BS model, and comparing the ML models to it. Here are the patterns seen for model error moneyness bias.

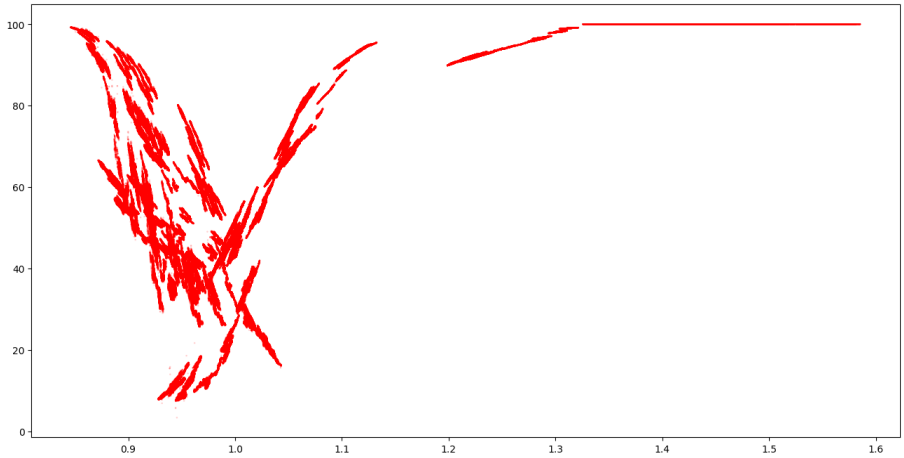


Figure 2.15: Black Scholes MAPE Moneyness bias

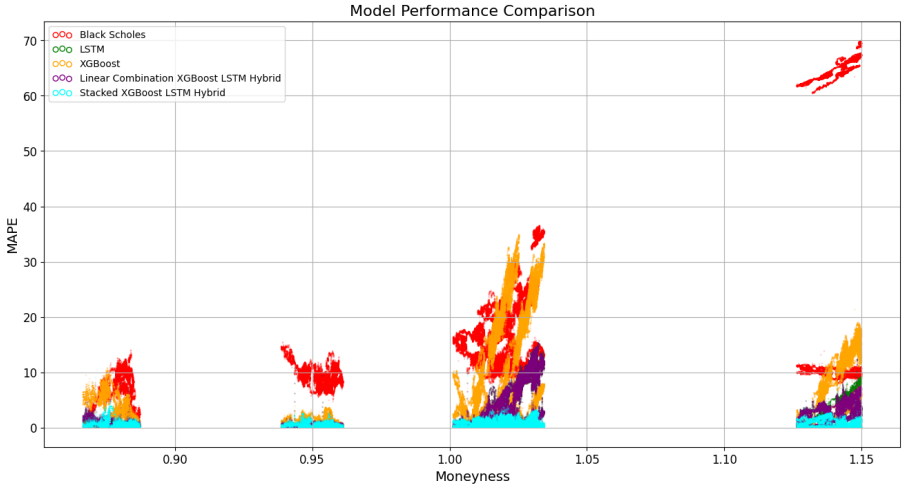


Figure 2.16: Model MAPE comparison, Russell 2000 Oct

A similar pattern can be observed for directional Bias too. Here it is seen that BS systematically undervalues OTM contracts (a finding consistent with the literature), whereas the ML alternatives show no systematic bias, and very low bias in general.

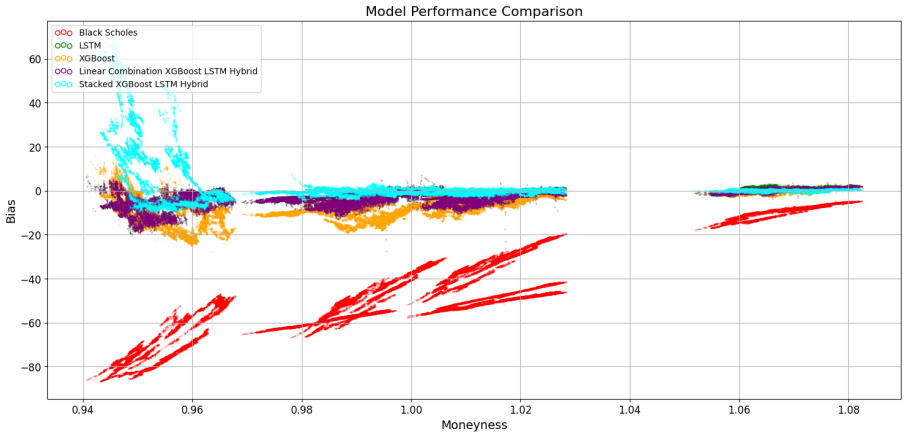


Figure 2.17: Model directional bias comparison, Nasdaq Nov

With ML models, and particularly deep neural networks, it often isn't possible to make precise inferences and attribution analysis. However, there are still some tools and visualizations which can help us attempt to peek inside the black box. It is important to keep in mind the caveat, that due to the non-linearities of the interactions within the networks, and the multidimensional features present, patterns that appear obvious may not be straightforward. Still, it is possible to make an attempt to visualize some of the inner workings of the models, and some educated guesses about their potential interpretation within the context of the current experiment. A potential starting point in visualizing the sensitivities of the deep learning model is to look at a correlation heatmap of the LSTM units to each of the features specified. Of course, such an approach is just an approximation, since the sensitivities and resulting activations may often be a non-linear process. Still, by looking at the absolute correlations it is possible to make an inference at a minimum about the presence of LSTM units which are strongly sensitive to specific features in a way the strongly approximates a linear relationship. In our architecture, the LSTM layer is followed by an attention layer. This is a layer that dynamically modulates the inputs it receives (in this case the LSTM layer outputs), in order to amplify them based on the context observed within the data. Therefore, it is possible to observe the post-attention activations in the network. Together these two pieces of information may provide a crude approximation of how the attention mechanism scrutinizes the features observed over time, and what features the model

has a higher approximately-linear sensitivity to. Looking at the data, some patterns are observed that fit our expectations in this context.

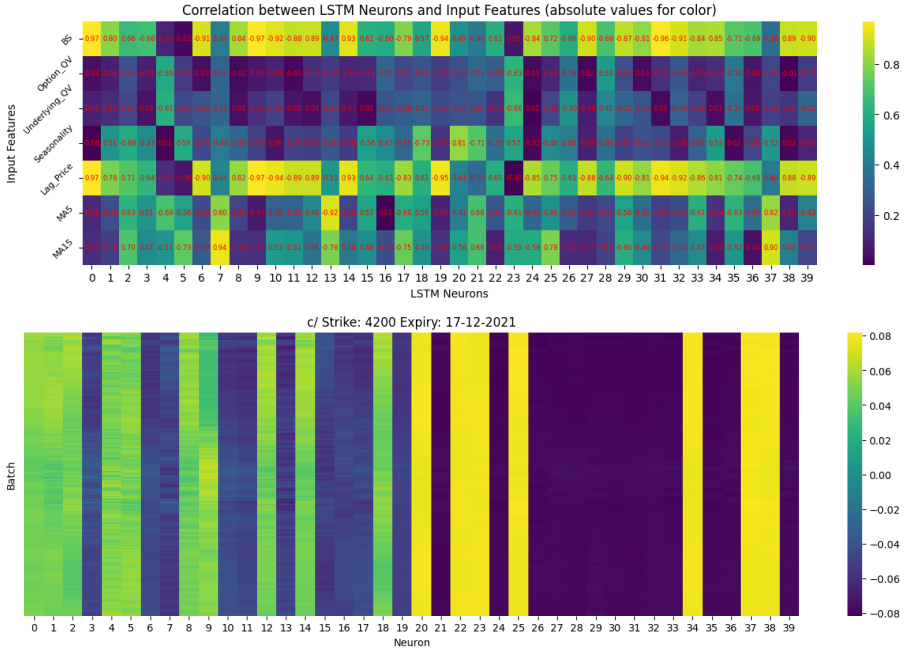


Figure 2.18: LSTM unit - feature correlations and Attention modulated outputs, ATM call S&P500, November

Figure 18 shows the LSTM feature correlations and attention modulated outputs for an at-the-money call. The feature correlations have a clear tendency towards placing a high emphasis on the Black Scholes valuation feature, as well as the lagged price feature. There are relatively few units that are strongly correlated to the liquidity features or the seasonality feature (some notable exceptions are units 7 and 23), at least in the linear sense. This result is expected. It is known that the Black Scholes model performs best for at-the-money contracts. Therefore it makes sense to see its feature having a more prominent presence. Looking at the impact of attention, it is seen that the activation patterns are unevenly distributed, as is their stability over time (batches). Perhaps the most important observation is that some of the units associated with the liquidity metrics, still exhibit activations, and these fluctuate over time with occasional spikes (for instance unit 4). That is, their impact has still remained important after the scrutiny by the attention mechanism. A very similar pattern can be seen for another at-the-money contract below in Figure 19

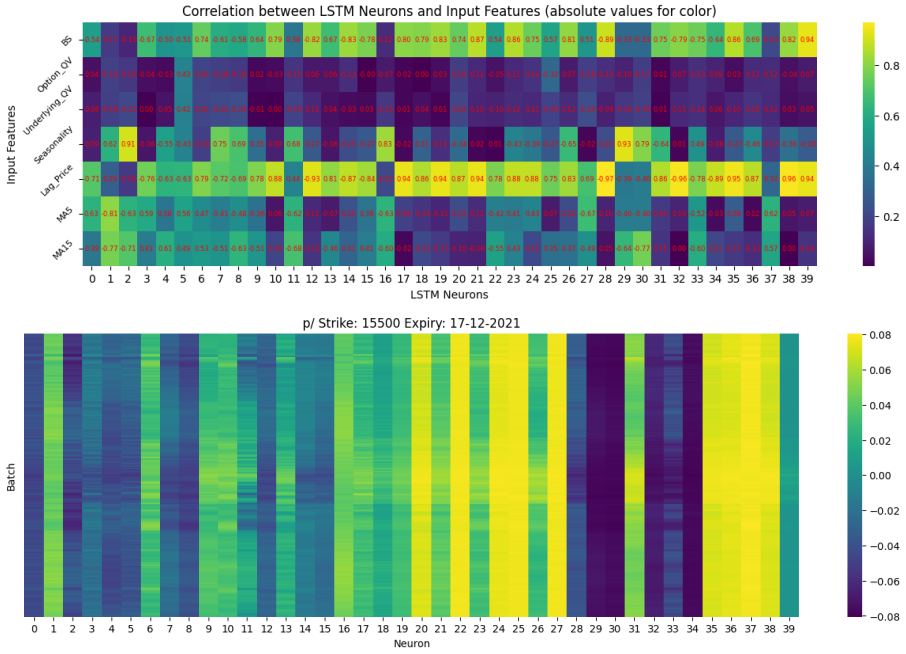


Figure 2.19: LSTM unit - feature correlations and Attention modulated outputs, ATM put Nasdaq, November

The pattern correlation and activation patterns for a deep out-the-money contract seen in Figure 20 are noticeably different. Here it is seen that the LSTM unit - feature correlations are much more evenly distributed across a wider range of feature combinations. This makes sense too, and could be a result of a combination of two factors. First, the model may be less sensitive to certain features, such as the Black Scholes valuation, due to their relatively poor performance in the context of deep-out-the money contracts, and relatively more sensitive to alternative features such as the proposed liquidity metrics, as it searches for an alternative solution. And second, the nature of the relationships may be less approximately linear in this context. The attention modulated outputs also show a much more dynamic picture. Activations are evenly distributed across units, with peaks and troughs over time. The attention mechanism also favours a more expanded and dynamic model where the whole range of features have a significant impact.

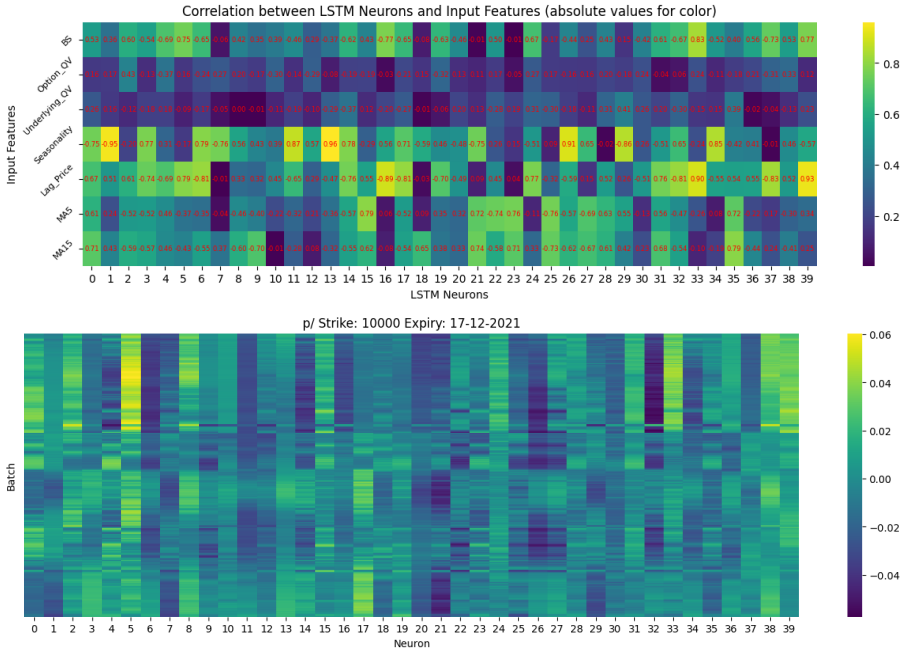
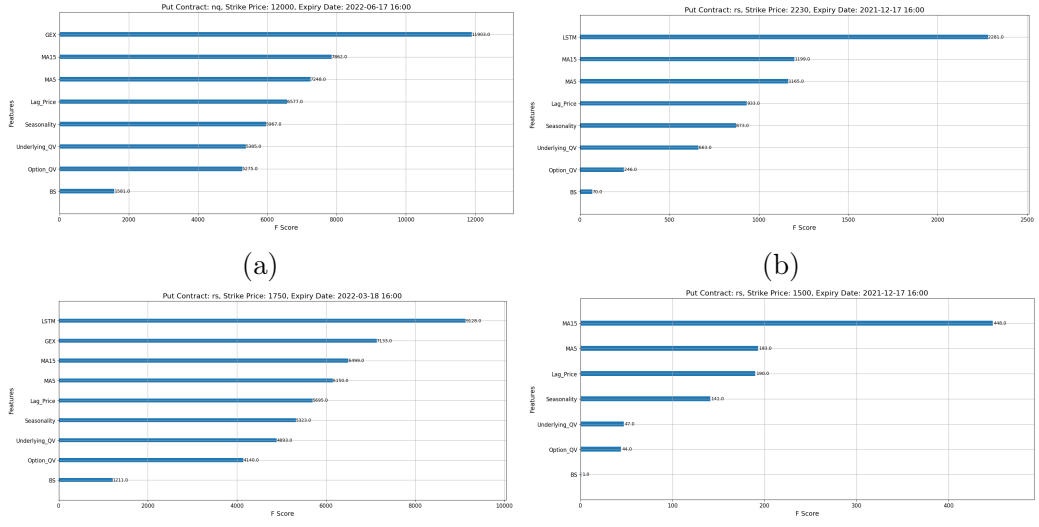


Figure 2.20: LSTM unit - feature correlations and Attention modulated outputs, deep OTM put Nasdaq, November

Similar types of analysis can be performed for the XGBoost specifications. Here it is possible to use the built-in packages for feature importances and Shapley feature analysis. A few examples follow



(a) (b)

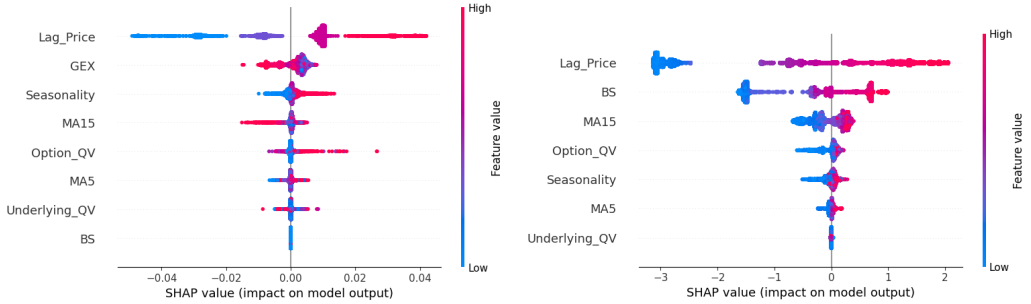
(c) (d)

Figure 2.21: XGBoost feature importances for a selection of contracts

The feature importance measure quantifies how many times a specific input fea-

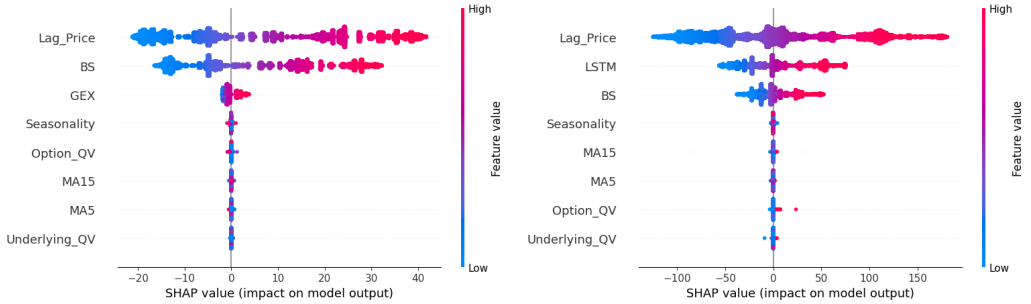
ture has been used as a decision criteria in the decision trees of the XGBoost model. The intuition is that more prominent features are used more often. The key observations from a selection of models for different contracts presented in Figure 21 is that the liquidity metrics have a strong influence in all contexts. The moving averages of the QV metric are particularly influential in all contexts. Also, in the context of stacked specifications and B specifications the LSTM and GEX features take on a top importance. This confirms that the XGBoost model perceives the LSTM predictions as having an important informational content. The sensitivity to GEX is also prominent, however that may also be an artefact of the hyperparameter tuning outcome for the B model specifications, where the XGBoost 'gamma'[12] parameter was set to 0, making the model less selective when it comes to feature inclusion. While the feature importances give an idea of how many times the features were used in the prediction process, they dont necessarily provide information about the magnitude or direction of their impact on the predicted value. Shapley feature analysis is a particularly useful tool that can provide more detailed understanding within this context. This feature analysis approach is inspired by cooperative game theory in order to interpret predictions made by ML models. It assigns an importance value to each feature, indicating how much each feature contributes to a particular prediction Lundberg and Lee (2017). The Shapley value offers a fair distribution of contribution across features, with the sum of all Shapley values being the total effect. For decision trees, Shapley feature analysis can be used to understand the contribution of each feature to a prediction. The algorithm traverses the tree, calculating the marginal contribution of each feature at each split to the final prediction, resulting in Shapley values Strumbelj and Kononenko (2014).

[12]Not to be confused with GEX, the Gamma Exposure feature in the current setting



(a) Shapley plot for deep OTM S&P500 put, Nov

(b) Shapley plot for ATM Russell 2000 put, Oct



(c) Shapley plot for ATM S&P500 call, Nov

(d) Shapley plot for ATM Nasdaq call, Nov

Figure 2.22: Shapley plots for feature impact for a selection of contracts

The main difference here is that the Black Scholes valuation seems to have a slightly larger impact for the at-the-money contracts, than feature importances alone would suggest. One intuition behind this, is that the nature of the relationship between the price and the Black Scholes prediction is simple in the cases of the at-the-money contracts, so the decision tree structure requires less complexity and less branches to approximate it. On the other hand, the GEX and LSTM features where included in the specifications, also have a large magnitude of impact, as well as feature importance. This may suggest that the nature of the relationship of these features to the final price is less simple, and it takes more decision tree iterations to approximate it. The impact of the liquidity features is observed to some extent in all contracts, with the most prominent effect for the deep out-the-money contract, which also sees almost no impact from the Black Scholes valuation feature. These findings echo the observations from the LSTM models, that for contracts further away from the money, the relative importance of liquidity features rises, and the relative importance of the Black Scholes valuation diminishes.

2.5 Summary

In this paper compelling empirical evidence is presented for the advantages of using an ex-ante high frequency measure of illiquidity, the Quote Volatility (QV) ratio to observe and detect different patterns of market behaviour. Some speculation about the underlying mechanism at work can also be provided, with similarities observed to the findings of Gabaix et al. (2006a). In the process, several promising new directions for future research were identified. We then performed a multi stage pricing experiment using machine learning models to value European equity index options. We note patterns in our results which confirm the usefulness of our measures of illiquidity in the context of the option illiquidity premium. We also demonstrate a useful strategy for specifying, tuning and estimating a range of machine learning models in this computationally challenging context. We demonstrate their out of sample superiority over the parametric Black Scholes benchmark in contracts with a wide range of moneyness and time-to-expiry observations. We also make an attempt at attribution analysis using various a combination of techniques, observing patterns that fit the overall empirical narrative.

2.6 Future research

The patterns presented earlier, which depict the relationship between the ex-ante measure of liquidity in a high frequency setting and price impact can serve as a starting point for several new directions of research.

The approach to examine this within a large universe of assets cross-sectionally and historically could illustrate further the potential ability of the measures to depict liquidity events and changes in market conditions due to systematic or idiosyncratic factors. Also, while the 3 indices examined earlier clearly have differences in their usual liquidity levels, it would be interesting to see how the patterns behave in the settings of assets with very different typical liquidity levles, perhaps by orders of magnitude. Does the relationship observed here hold for small cap illiquid stocks?

If it is different, in what ways, and what inferences can be made from that.

Another direction to investigate would be the ability of the metrics to capture other types of market behaviour which may share some similarities with the end of day gamma rebalancing effects discussed earlier. For instance, there are other types of time contingent position rebalancing activities observed in various markets. It is sometimes suspected that end of month, end of quarter and end of year portfolio rebalancing activities may take place and impact markets. It is also known that similar activities may take place in the FX market for the daily fix, or in some commodity markets such as energy for end of day rebalancing. While not necessarily related to gamma exposures, such rebalancing events could be hypothesized to impact markets in similar ways. Research could focus specifically on the markets and assets involved, and attempt to depict the pattern of activity using the QV measure. Such evidence would serve to reinforce the empirical versatility of the measure.

Finally, an additional, more in depth investigation could be performed into the mechanisms and patterns of behaviour connecting the QV measure with activity within the depth of the order book. Perhaps such observations could add information to the patterns of market conditions captured, and help provide clues about the correct positioning of the metric within existing theoretical frameworks.

2.A Relationship between QV, intraday Amihud and VWAP

Examining the formula for QV and the intraday specifications of Amihud proposed, it is possible to identify a mathematical connection, pending certain conditions, which could proxy for liquidity. One such condition would be that:

$$P_t = \frac{Ask_t + Bid_t}{2} \tag{2.6}$$

That is, if trades tend to on average occur exactly at, or close the midpoint Another similar condition would be that the Bid-Ask spreads narrow towards 0. In this case traded price, bid prices and ask prices would begin to converge together.

The intraday Amihud specifications proposed also have an obvious connection with the popular VWAP benchmark:

$$VWAP = \frac{\sum (Price_t \cdot Volume_t)}{\sum Volume_t} \tag{2.7}$$

The presence of such a connection is interesting, since it is known that VWAP is a popular execution benchmark for institutional traders, including via execution algorithms. It is certainly possible to see how a strategy optimizing for performance with respect to this benchmark might exhibit a sensitivity to the QV and intraday Amihud metrics.

Chapter 3

Ethereum Urgency Score

3.1 Introduction

Ethereum is the leading blockchain settlement layer for decentralized financial activity, supporting a wide range of applications including decentralized exchanges (DEXs), decentralized finance (DeFi) protocols, and non-fungible token (NFT) marketplaces. As of January 2025, it processed approximately 1.17 million transactions per day and remained the largest settlement network for DEX trading volume in 2024. Institutional and government interest in Ethereum has also accelerated, with major financial institutions such as BlackRock and Fidelity pursuing Ethereum-based exchange-traded funds (ETFs), and several central banks piloting Ethereum-based digital currencies.

Despite Ethereum's transparent ledger and verifiable transaction records, significant challenges remain in understanding market participant behavior and informational dynamics. While blockchain technology enhances transactional transparency, it does not fully reveal trader intent, informational advantage, or execution strategies. This persistent behavioral opacity limits both academic insight and the design of effective trading and risk management systems. Recent research on crypto market microstructure highlights these limitations, emphasizing that while decentralized settlement and open data structures distinguish blockchain markets from traditional finance, identifying and measuring informationally motivated trading in real time remains an open challenge Biais et al. (2023a); Cong and He (2019).

This paper addresses these challenges by introducing the Urgency Score, a novel on-chain metric designed to extract behavioral signals from Ethereum transaction data. By classifying transactions based on execution urgency derived from gas pric-

ing, timing behavior, counterparty patterns, and rarity; we test whether this metric enables the detection of informationally rich or strategically motivated activity on public blockchains. Results show that urgency-typed transactions contribute unequally to price discovery and market impact. This study offers new pathways for behavioral market analytics in blockchain environments, with implications for execution strategy design, liquidity provision, and regulatory monitoring.

Using a unique dataset of Ethereum transactions from 2021 to 2023, we develop a transaction-level metric that scores individual wallet behavior based on six urgency signals: (1) high gas fees, (2) off-hour executions, (3) large transaction size, (4) interaction with unusual counterparties, (5) transfers involving rare assets, and (6) engagement with rare or specialized smart contracts. These dimensions capture not only fee-based bidding behavior but also timing and relational execution patterns, providing a more comprehensive characterization of transaction urgency. Wallet-level behavior—including gas fee strategies, temporal execution patterns, and counterparty selection—reflects the underlying adoption dynamics and strategic heterogeneity observed in decentralized ecosystems, as theorized by Cong et al. (2021b). By moving beyond gas price alone, our Urgency Score offers a holistic, multi-dimensional view of real-time execution behavior. This blockchain-native metric enables market participants, smart contract designers, academics, and regulators to identify transactions likely driven by informational urgency, liquidity needs, or strategic execution objectives. As such, the Urgency Score provides a novel behavioral lens on blockchain activity, advancing the empirical study of blockchain market microstructure and token-based economic coordination.

We apply both reduced-form price impact analysis and Hasbrouck (1995) Information Share methodology to assess the informational contribution of transactions characterized by their urgency. Our reduced-form estimates reveal that transactions with higher Urgency Scores exhibit greater price impact, consistent with informed trading behavior. Complementing these results, our Information Share analysis shows that high-urgency flows contribute disproportionately to price discovery relative to their frequency, suggesting that these transactions carry long-term informational content. Notably, this informational dominance persists even after the introduction of private pools that conceal transactions from the public mempool[1], indicating that urgency-based execution reflects strategic information revelation, not

[1]The operationalization of private transactions on Ethereum accelerated following the adoption of MEV-Boost after the Merge in September 2022. Validators increasingly outsourced block building to specialized relays, many of which support private transaction bundles submitted directly by traders seeking MEV protection or execution privacy Flashbots (2022); Foundation (2022).

merely visibility. We further show that urgency-driven trades continue to lead price formation, particularly during crisis periods such as the FTX collapse. These findings provide robust evidence that wallet-level urgency behavior plays a central role in shaping blockchain market efficiency.

Building on the descriptive and structural evidence, we further assess the predictive utility of urgency-typed transactions using rolling monthly XGBoost models. We adopt the rolling window approach as in Cao et al. (2024). We train tree-based gradient boosting models to predict next-day centralized exchange (CEX) returns and order flow, using lagged features derived from urgency score transaction volumes alongside past return and order flow features. The use of rolling windows ensures that model performance reflects realistic, out-of-sample forecasting conditions that account for potential regime shifts and evolving market dynamics. Higher urgency scores consistently emerge as key predictive features in periods of market stress or heightened volatility, supporting their role as informationally rich signals. Meanwhile, low urgency transactions show more stable but weaker predictive relevance, consistent with their role in liquidity provision or routine trading. These machine learning results complement our structural VAR and Information Share findings by showing that urgency-typed transactions not only reflect past price impacts but also contain forward-looking information about short-term market dynamics.

To demonstrate the practical implications of the Urgency Score, we design a signal-based trading strategy that exploits short-term price momentum following urgency spikes. Our findings show that high-urgency momentum strategies generate statistically significant returns, outperforming naive execution benchmarks. This extends the literature on blockchain trading strategies and execution algorithms Aune et al. (2017); Easley et al. (2019) by providing a behaviorally grounded signal that is directly observable on-chain and robust to market regime shifts.

Our paper is primarily related to the literature on the strategic behavior of informed traders and its role in price discovery across different market structures. Prior work shows that informed traders often strategically time and size their trades to avoid detection, leading to prices that only partially reflect private information, as demonstrated by Collin-Dufresne and Fos (2015) in equity markets. Price discovery accelerates when multiple informed traders compete, as shown by Holden and Subrahmanyam (1992) and Foster and Viswanathan (1996). We extend this literature by showing that wallet-level behaviors on Ethereum, as captured by our Urgency Score, provide a real-time proxy for execution urgency and significantly influence price discovery in decentralized blockchain-based markets.

Our findings also contribute to the market design literature, particularly the challenges identified by Biais et al. (2023a), who emphasize the frictions, governance shortcomings, and informational asymmetries inherent in blockchain ecosystems. Despite Ethereum's transparent architecture, these structural limitations create behavioral blind spots that hinder effective market monitoring and regulatory oversight. By quantifying wallet-level urgency, our Urgency Score provides a practical tool to make latent strategic behavior observable, offering actionable insights into liquidity dynamics, information flow, and ecosystem stability.

Recent studies have highlighted various frictions and risks in DEX environments, such as sandwich attacks (Park, 2023), comparative transaction costs between DEXs and centralized exchanges (Barbon and Ranaldo, 2021), and the role of automated market makers like Uniswap in shaping liquidity and execution costs (Lehar and Parlour, 2021). Lehar et al. (2024) show how priority fee bidding leads to liquidity fragmentation across DEX pools, while Hasbrouck et al. (2022) find that higher DEX fees can stimulate trading volume. Closest to our work, Capponi et al. (2024) show that high-fee transactions on DEXs carry more private information, reflecting trader urgency. We complement this literature by incorporating timing, trade size, and counterparty selection as additional dimensions of price-relevant behavior, particularly during periods of network congestion and market stress.

Finally, our work contributes to the growing intersection of blockchain analytics, AI, and market stability. We show that real-time monitoring of wallet urgency provides early-warning signals of price movements, particularly during crises such as the FTX collapse. By integrating the Urgency Score into machine learning models, we demonstrate its predictive value for short-term price direction and centralized exchange activity. These findings suggest that urgency-based analytics can enhance trading algorithms, arbitrage detection, and risk management for both on-chain and off-chain market participants.

The paper proceeds as follows. Section 2 introduces the Ethereum network and presents the construction of the Urgency Score, a novel transaction-level metric for measuring execution urgency on the blockchain. Section 3 describes the data sources and preprocessing methods used in our empirical analysis. Section 4 presents the results of our structural analysis, showing that higher-urgency transactions are associated with the revelation of private information and enabling the design of urgency-based trading strategies. Section 5 evaluates the predictive power of the Urgency Score using XGBoost models, demonstrating its value in forecasting short-term price movements and liquidity conditions. Section 6 concludes and discusses

implications for market participants, regulators, and future research.

3.2 Ethereum Blockchain and Urgency Score

In this section, we describe Ethereum institutional details, from network structure, consensus mechanisms, fee market design and smart contract execution. We discuss how these institutional details dictate how transactions are prioritized, executed and the economics behind transaction urgency.

3.2.1 Institutional Details

Ethereum Blockchain is a public ledger that keeps records of all Ethereum related transactions. It is shared between all participants and is based on a reward mechanism as an incentive to run the transaction networks. The network relies heavily on cryptography to secure a transaction as a consensus mechanism. Each account consists of a public and private key duo: the private key is used to sign each account's transaction, whereas the public key is used by all participants to verify the transaction's validity in a transparent and decentralized way. Ethereum Blockchain's ability to process transactions in a trust-less environment, in addition to trading its official cryptocurrency "Ether" (ETH) presents the framework of execution of smart contracts.

Consensus Mechanism: Proof-of-Stake PoS

Ethereum's transition from Proof-of-Work (PoW) to Proof-of-Stake (PoS) - a process known as The Merge, was driven by the need for a more energy-efficient and scalable blockchain system, in line with Saleh (2021). Ethereum uses PoS validators who include transactions in blocks. Transactions by higher gas fees are prioritized. Unlike the previous Proof-of-Work (PoW), where miners competed for block rewards, PoS incentives validators through priority gas fees, reinforcing the economic impact of urgency in Ethereum blockchain ecosystem. We use gas fees as one of the components in determining our urgency score.

Ethereum Mempool priority system

The Ethereum mempool functions as a public but transient order book, where transactions queue before they are included in a block. The validators in PoS select transactions based on priority, typically determined by gas fees. Simple transactions such as ETH transfers are prioritized over more complex transactions such as DeFi if

they offer similar gas fees. This is because ETH transfers require less computational work and have a lower gas cost compared to transactions that execute smart contracts. Validators can however reorder, include, or exclude transactions to maximize profit Park (2023). For instance, they may prioritize transactions that make profits by frontrunning them, or arbitrage bots may be prioritized even if they are complex. Users have incentive to increase their priority fee or use private transactions to bypass the mempool, or send transactions through flash bots to avoid front-running.

DeFi protocols built on Ethereum provide financial services without intermediaries, using smart contract to execute trades, loans and asset management automatically. Unlike traditional finance, where clearing houses, brokers, and regulated exchanges act as counterparties, DeFi operates on peer-to-peer (P2P) and automated market making (AMM) models. Without clearinghouses, the risk is mitigated by overcollateralization, or automatic smart contract liquidations. For instance, DEXs relying on blockchain validators to execute trades, creates a fundamentally different market structure compared to centralized exchanges (CEXs). Unlike CEXs, where orders are processed continuously based on time priority, DEXs operate in discrete intervals determined by the blockchain's block time. Since validators prioritize transactions offering higher fees, traders effectively engage in a fee-based competition rather than a speed-based one as in CEX. The competition for block space may lead to a pike in gas fees especially during network congestion. This can make DEX trading costly for smaller traders while benefiting large traders and arbitrageurs.

Beyond ETH, stablecoins like USDC and USDT, and major DeFi tokens such as UNI and AAVE, Ethereum also supports niche and rare digital assets—including NFTs like CryptoPunks and Art Blocks, low-liquidity governance tokens like INDEX (Index Coop), experimental assets like AMPL (Ampleforth), and bespoke smart contracts powering DAOs, prediction markets, or custom derivatives.

Various types of transactions queue in the mempool, competing for block space based on priority fees. These transactions include:

- *Transfer*: ETH transfers from one address to another and they are generally at lower gas cost and higher priority compared to other more complex executions.
- *TransferFrom*: ERC-20 tokens transfers (such as USDT, LINK) which require interactions with smart contracts, making them more complex than ETH transfers. Users bid medium gas fees based on the contract logic.
- *Swap*: exchange of one token for another via a liquidity pool. These types of

transactions require higher gas fees competing with swaps on other DEXs.

- *Withdraw*: the removal of assets out of the smart contract, e.g. liquidity removal or DeFi withdrawals which can be time-sensitive. The gas fees can be medium to high depending on users' urgency and the demand. During market downturns, users rush to withdraw funds, increasing gas fees.

- *Mint*: the creation of new tokens or NFTs such as DeFi yield farming and NFT minting. This can cause congestion during high-demand events such as large drops of NFTs, when users bid high gas fees.

Maximal Extractable Value (MEV) refers to the profit a trader can extract by reordering, inserting, or censoring transactions within a block. One common MEV strategy is the sandwich attack, where a bot detects a large swap about to happen (e.g., a user buying a token on a DEX), then places a buy order just before (front-run) and a sell order just after (back-run) the user's transaction. This artificially inflates the price for the user and allows the bot to profit from the price movement. To execute this, MEV bots often pay higher gas fees to miners or validators to prioritize their transactions, leading to network congestion and increasing gas costs for all users.

Lehar and Parlour (2023) examines how transparent and batched settlement on decentralized ledgers enables certain actors—namely settlement agents (e.g., miners or validators) — to extract value from arbitrage trades, often through MEV-related strategies. This extraction can make some forms of arbitrage socially inefficient by turning otherwise beneficial activities into a competition over transaction ordering. To explore potential improvements, the authors model a private (off-chain) settlement mechanism as an alternative to public, transparent settlement. They find that this model could impact the overall efficiency of arbitrage activity. One striking empirical finding is that arbitrageurs pay over $1 million per day to private settlers, suggesting a high cost of accessing priority or private execution channels and a potentially distorted market for fair access.

3.2.2 Urgency Score

Wallets are the base layers of the crypto market. Any process in the higher layers will originate and eventually trace bake the wallet's holder behavior. Unlike Bitcoin, Ethereum wallet transaction data still unexplored in research presenting an opportunity to analyze it. We analyze the competition between DEXs and other transaction

types around gas fees and block space availability. We propose a new metric, namely Urgency Score, to rank each wallet transaction by urgency as discussed previously. We further discuss how wallet behavior in the Ethereum Blockchain based on gas fees, network congestion and 24 hours functioning provide insights into price movements and order flows.

For each wallet transaction, we define six binary indicators that capture how urgent is:

1. *higher gas fee transaction* to secure priority as discussed in the previous section.

2. *out-of-hours transaction* such as during weekends, holidays, or late-night periods in major financial hubs, can be time-sensitive or urgent requiring high-priority for several reasons: detecting arbitrage opportunities, avoiding MEV front-running, reacting to network congestion and gas fee fluctuations, and avoiding DeFi liquidation risks.

3. *large transaction* such as high-value swaps, whale transfers, large DeFi liquidations and arbitrage trades are all time-sensitive and require strategic execution to ensure fast validation due to their potential price impact on the markets.

4. *unusual counterparty* requires an immediate and urgent action depending on competition, liquidity risks and market impact. Smart traders act fast to seize opportunities and minimize losses.

5. *unusual smart contract* requires an immediate action to protect funds, seize opportunities or avoid risks due to changes in market conditions for instance.

6. *rare asset* requires an immediate action such as withdrawal when market conditions signal urgency due to several factors such as illiquidity and missing arbitrage opportunities.

We compute the number of criteria met by each wallet and rate it between 1 and 6. For example, if a large transaction occurred out-of usual hours of trading - which is determined outside the hours a wallet used to transact - and at higher gas fee; this transaction is scored 3. The closer to 6 the score is for a given wallet, the more urgent its behavior is considered. We also compute an exponential urgency score by considering the first 3 criteria as one factor, and the last three criteria as the second factor: $e^{out-of-hours+large+expensive} + e^{rareasset+rarecounterparty+rarecontract}$.

3.3 Data and Summary Statistics

Our sample period consists of two years from January 2022 - December 2023. This sample period allows us to examine the wallet behavior particularly at volatile times including ETH price crash in January 2022, Terra Luna collapse in May 2022, FTX collapse in November 2022, transition to PoS in September 2022, and the rise of private pools, USDC depeg and SVB collapse in March 2023. We collect high-frequency price and trade volume of ETH/USD executed on Binance exchange that we bought from Kaiko. Kaiko obtains the data by querying APIs provided by centralized exchanges. The variables contained in the data are timestamped to the millisecond: the price at which the trade happened, the amount of the trade, and an indicator whether the trade was buy or sell-initiated. We next detail how we collect Ethereum Blockchain wallet data which we match later with Binance data in block time to compute returns, and traded volume on CEX.

3.3.1 Ethereum wallet data

We here detail the construction of Ethereum Wallet transactions data for our analysis. Specifically, we analyze wallet behavior by categorizing transactions into Swaps, Transfers, TransferFrom, Withdraw, and Mint functions.

Collecting the data

We employ an automated data retrieval methodology via Etherscan API. This approach enables the extraction of raw transactional data directly from the Blockchain. Users' addresses can function as wallets, assets such as ERC20 tokens, or smart contacts such as DEX protocols. There is an overwhelming diversity of traded assets, including exotic and unknown coins.

Each transaction is characterized by a set of attributes including: the addresses of the sender and the receiver, the gas price expressed in ETH, the quantity of gas utilized, the timestamp, the block number, the transaction value denominated in ETH, and the input - a hexadecimal string encoding function calls within smart contracts. To further our understanding of transactions involving value transfers not denominated in ETH, we construct a stablecoin dictionary. This repository includes detailed information on all stablecoins transacted on the Ethereum platform, encompassing the address, symbol, number of decimals, and the asset type (e.g. fiat currency, cryptocurrency, precious medal, or physical good) to which the stablecoin is pegged. In total we identify 265 stablecoins, anchored to 27 different

fiat currencies, ETH, BTC, silver, gold and other assets. The compilation of this stablecoin dictionary was informed by data extracted from Etherscan and official documentation associated with each stablecoin.

Decoding the data

Transactions encode various smart contract interactions which can be challenging to understand and study. We use the Web3.py library and a vast collection of Application Binary Interface (ABIs) to decode the hex input into readable formats, which were then consolidated for comprehensive data analysis. After decoding, we indentified 14 functions: pause, unpause, swapExactTokensForTokens, state, getReward, addLiquidity, approve, withdraw, burn, mint, repay, transferFrom, transferOwnership, and transfer.

We inspect all the functions we identify. For the swapExactTokenForTokens function, we identify the addresses of stablecoins dictionary which pegged with USD, ETH, and BTC, then we can derive the economic value of any transaction they are involved with. For example, the majority of swap transactions will involve a stablecoin or ETH as one leg. It is really unusual to swap one ERC20 token for another unless it is a stablecoin. One plausible reason is that one can hold her free invested funds in ETH to pay the transaction fee in gas, to buy an ERC20 token or selling ERC20 token for ETH.

We only focus on transactions with economic value. For example, "approve" function is used when one first connect her wallet to a DEX, and authorize it to be traded in exchange for a small gas fee. This is simply means that one approve the contract to interact with her wallet without any economic exchange. We do not focus on transactions that use the approve function in their inputs, as it has no direct economic value. Another type of contract function is withdraw. As we inspect more, we find out that the majority of withdraw transactions are transactions where the holder of wrapped ETH convert it back to a normal ETH. Wrapped ETH are ETH in tokenised form so that it conforms with an ECR20 tokens protocols. The process is the following: one can send ETH to the wrapped ETH smart contract to receive wrapped ETH at a 1 to 1 ratio. The withdraw transactions are interesting in this case. On the one hand, they may not have an immediate economic impact since the asset is essentially transformed from one form to another at a 1 to 1 ratio incurring transaction cost. On the other hand, they might signal some trading intentions. Wrapped ETH is used for various transactions with smart contracts and DeFi systems. An account converting from wrapped ETH to plain ETH may signal the intention to shift from an active to a passive behavior.

We build a search algorithm as follows: (1) we identify the type of transaction by decoding the input via ABI. In particular, we construct template ABIs for decoding common interactions such as Transfer, TransferFrom, Swap, Withdraw and Mint; (2) we recognize the type of assets involved such as stablecoins, wrapped ETH or any other of which we could convert the token value to USD; (3) if we do not recognize the asset, we look for recent cases where at least one of the assets were involved in a transaction of which the price is known. This allows us to assign a value to the new transaction.

Building the data

After processing the data, we extract variables of the following transactions:

- Mint: Mint_Value refers to the calculated value in USD, Mint_Stabcoin refers to boolean value equals to 1 if it is a stablecoin and 0 otherwise.
- Withdraw: Withdraw_Value refers to the calculated value in USD, Withdraw_Wrapped is equal to 1 if this a wrapped ETH withdrawal, Withdraw_Stabcoin is equal to 1 if this a stablecoin withdrawal.
- TransferFrom: Transform_Value refers to the calculated Value in USD, Transfer_Burn is equal to 1 if it is sent to a burn address, Transfer_Sender is the actual sender address, Transform_Receiver is the actual receiver address, Transform_Amount is the amount transferred.
- Transfer: Transfer_Value is the calculated value in USD, Transfer_Burn is equal to 1 if it is sent to a burn address, Transfer_Receiver is the actual receiver address.
- Swap: Swap_value is the calculated value in USD, Swap_Failed is equal to 1 if the swap fails, Swap_Stabcoin is equal to 1 if the Swap involves a stablecoin or 0 otherwise, Swap_Asseta refers to Asset A, Swap_Assetb refers to Asset B, Swap_Sender is who initiated the Swap, Swap_Receiver is who is on the other side of the Swap, and Swap_Contract is what was the smart contract (DEX).

3.3.2 Summary statistics

The full dataset includes all transactions recorded on Ethereum between January 2022 to December 2023. This period yields a comprehensive collection comprising

791,053,919 transactions, which contributed to the creation of 4,942,231 blocks and involved 93,930,642 unique addresses. We were able to derive information from a diverse array of transactions and interactions on a day-by-day basis as detailed previously. We manage to comprehend 65% of the whole dataset that we use in this analysis.

To be able to compute the Urgency Score for each wallet as detailed in section 3.2.2, we include the history of transactions initiated by each wallet over the past 30 days. We remove inactive wallets for over 30 days. For active wallets, we have the following variables: the gas fees, the total transaction value, total unique counterparties (receivers), and the timezone in which they tend to be active based on the 75% activity rule. We also determine the top 20 most frequent counterparties for each active wallet. This will allow us to define six binary variables for each active wallet: (1) out-of-hours transaction which is determined outside hours a given wallet used to transact; (2) significantly larger value defined as 2 times more than the average USD transactions value of this wallet; (3) higher gas fees than the average for the same recent transactions; (4) interact with unusual counterparty, i.e. not in the top 20 of the list; (5) involving a rare asset for this wallet; and finally (6) interacting with unusual smart contract.

We aggregate wallet transaction data by block, and compute the average urgency score by block time over the sample period. Figure 3.1 depicts the variation of Urgency Score over the sample period highlighting the significant events that could influence transaction urgency and the distribution of transactions across major DeFi pools, such as Uniswap, Aave and Compound. During the sample period, the Ethereum blockchain and its DeFi ecosystem experienced a series of major events. In early January, Ethereum's price crashed, leading to reduced swap transactions. This is depicted by the highest level of transaction urgency score during January 2022.

Urgency score peaks again at times of Terra Luna collapse in May 2022 which caused market instability. This has led to a surge in swap activities. Post Merge in September 2022, urgency score increases again with DeFi transactions picking up due to improvement in network efficiency. The collapse of the centralized exchange FTX in November 2022 underscored the risks associated with centralized platforms. This particular event shifts user preference towards decentralized platforms, emphasizing the importance of transparency and decentralization in crypto markets. This is depicted by an increase in urgency score and percentage of DeFi transactions. Again in March 2023, the collapse of Sillicon Valley Bank (SVB) and USDC depeg has

led to an increase in transaction urgency as depicted by Urgency Score and a big increase in DeFi's activity. Throughout these events, transaction urgency as depicted by our Urgency Score seems to be fluctuating based on market conditions and user sentiment. Periods of significant uncertainty has led to increased network congestion, prompting users to urgently transact as depicted by our Urgency Score metric.

We match transactions with quote data and trading data by block time. We use manily the quotes and trades reported on Binance. Table 3.1 reports the relative distribution of transactions by urgency score across different market regimes, including high and low return periods, high and low order flow conditions, and high and low volatility environments. Across all regimes, low-urgency transactions (U_0 and U_1) consistently dominate transaction activity, accounting for approximately 76% of all transactions. This stability suggests that routine or baseline transactional activity remains the primary driver of network usage under normal market conditions. In contrast, high-urgency transactions (U_5 and U_6) represent less than 1% of all transactions, but their shares remain stable across regimes, indicating that these transactions are persistent but rare expressions of informational urgency or strategic execution behavior. Notably, the share of U_2 transactions slightly increases in high-return and high-volatility periods, potentially reflecting an increase in moderately informed or opportunistic trading during market stress. These patterns support the hypothesis that urgency-typed activity varies systematically with market conditions, providing additional structure to blockchain transaction data that is otherwise opaque. While these descriptive patterns highlight systematic shifts in urgency-based transaction composition across market regimes, they do not directly reveal how such transactions influence price formation or informational efficiency. To address this, the next section examines the dynamic price impact and long-term informational contributions of urgency-typed transactions using structural vector autoregressions and Hasbrouck (1995) information share decomposition.

3.4 Measuring the Informativeness of Ethereum transactions via Urgency Score

To estimate the informativeness of Ethereum transactions segmented by Urgency Score, we follow Hasbrouck (1991) and specify a structural VAR model:

$$Ay_t = \alpha + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \tag{3.1}$$

where $\Phi_1, \ldots, \Phi_p$ are coefficient matrices, α is a constant vector, and ε_t is a vector of structural shocks. The structural shocks satisfy the standard assumptions:

$$\mathbb{E}[\varepsilon_t] = 0, \quad \mathbb{E}[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon, \quad \mathbb{E}[\varepsilon_t \varepsilon_s'] = 0 \quad \text{for } s \neq t$$

The covariance matrix Σ_ε is assumed diagonal due to orthogonal structural innovations. The contemporaneous relations among variables are captured by matrix A.

3.4.1 Specification with Urgency-Scored Transactions

We define the endogenous variable vector to include midquote returns on a centralized exchange and blockchain transaction flows classified by Urgency Score from 0 to 6:

$$y_t = \begin{pmatrix} r_t^{\text{CEX}} \\ x_t^{\text{CEX}} \\ x_t^{(0)} \\ x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ x_t^{(4)} \\ x_t^{(5)} \\ x_t^{(6)} \end{pmatrix} \tag{3.2}$$

where:

- r_t^{CEX} is the midquote return on a centralized exchange (e.g., Binance) from $t-1$ to t,
- x_t^{CEX} is the trade flow volume on the centralized exchange over the same interval,
- $x_t^{(i)}$ is the Ethereum transactions at Urgency Score level $i \in \{0, 1, \ldots, 6\}$ in block t.

This structure accounts for different classes of trading behavior, where urgency is a proxy for information asymmetry or strategic execution.

3.4.2 Permanent Price Impact (PPI)

We compute the informativeness of each transaction type via its *Permanent Price Impact* (PPI), defined as the cumulative impulse response of returns to a unit structural shock in each urgency score group:

$$\text{PPI}_i = \sum_{j=0}^{\infty} \frac{\partial r_{t+j}}{\partial \varepsilon_{i,t}} = [\Theta(1)]_{1,i+2}, \quad i = 0, 1, \ldots, 6 \tag{3.3}$$

where:

- $\Theta(1) = \sum_{j=0}^{\infty} \Theta_j$ is the long-run moving average matrix from the VMA representation of the SVAR,
- $[\Theta(1)]_{1,i+2}$ denotes the entry corresponding to the effect of a shock in urgency level i on the return (position 1),
- Index shift $i + 2$ reflects the ordering: position 1 is r^{CEX}, 2 is x^{CEX}, and positions 3–9 correspond to $x^{(0)}$ to $x^{(6)}$.

A higher PPI_i implies that trades with Urgency Score i convey more private information and are associated with long-lasting price effects, rather than temporary inventory-driven price changes.

Bounds on Information Share using Cholesky Decomposition

Following Hasbrouck (1995), we use the Cholesky decomposition to orthogonalize the structural shocks ε_t. Let Σ_ε be the estimated covariance matrix of the residuals from the VAR model:

$$\Sigma_\varepsilon = PP'$$

where P is a lower-triangular matrix from the Cholesky decomposition. The long-run impact of shocks is captured by:

$$\Psi = C(1)P$$

The variance of the permanent component (efficient price innovation) is:

$$\mathrm{Var}(\mu_t) = \Psi_1 \Psi_1'$$

where Ψ_1 is the first row of Ψ, corresponding to the midquote return. The Information Share of variable i is computed as the squared contribution of Ψ_{1i} to the total variance:

$$\mathrm{IS}_i = \frac{\Psi_{1i}^2}{\sum_j \Psi_{1j}^2}$$

Because the Cholesky decomposition depends on the ordering of the variables, Hasbrouck proposes calculating upper and lower bounds for the information share of each variable by varying its position in the ordering.

For each urgency score $x^{(i)}$, we compute:

- $\mathrm{IS}_i^{\mathrm{lower}}$: When $x^{(i)}$ is ordered last (i.e., all others contemporaneously explain it),
- $\mathrm{IS}_i^{\mathrm{upper}}$: When $x^{(i)}$ is ordered first (i.e., it contemporaneously explains all others).

This yields an interval estimate of each flow's contribution to price discovery:

$$\mathrm{IS}_i \in [\mathrm{IS}_i^{\mathrm{lower}}, \mathrm{IS}_i^{\mathrm{upper}}]$$

We report the midpoint of this interval as a summary measure:

$$\mathrm{IS}_i^{\mathrm{mid}} = \frac{1}{2}\left(\mathrm{IS}_i^{\mathrm{lower}} + \mathrm{IS}_i^{\mathrm{upper}}\right)$$

These measures provide insight into the extent to which trades classified by each Urgency Score level contribute to the incorporation of information into prices. In the reduced price impact model, a significant positive coefficient for Urgency (e.g., U_6) indicates that higher urgency is associated with larger price movements, implying that urgent traders (perhaps whale-like actors) have a higher impact on the price formation process.

In this study, we apply two methodologies to assess the impact of transaction volume by Urgency Score on CEX price changes. First, the IS methodology is used to quantify the contribution of each Urgency Score (U_0 to U_6) to the long-term price discovery process. By applying Hasbrouck's variance decomposition, we

isolate the permanent component of price innovations attributable to each type of transaction. Second, a reduced price impact model is used to estimate the short-term price effects driven by transaction volume and urgency. This model posits that higher Urgency Scores correlate with larger price movements, indicating that urgent transactions likely reflect private or strategically important information. The combination of these methods offers both a long-term and short-term perspective on how urgency in trading behavior influences CEX price formation.

3.4.3 Interpretation of Results

To better understand the informational role of urgency-typed transactions on Ethereum, we analyze both impulse response functions (IRFs) derived from a structural VAR and Hasbrouck (1995) Information Share (IS) estimates. This dual framework enables us to disentangle short-term price pressure from long-term contributions to price discovery. Figure 3.2 plots the IRFs of transactions by urgency to CEX return. The Impulse response functions of U_0 as opposed to U_6 show a larger magnitude response with a higher peak and a more gradual decay. U_6 to return shows instead a smaller magnitude response with a lower peak and a quicker decay. U_0 to U_4 generally show simpler patterns with fewer peaks, indicating more stable or less complex relationships between the respective variables and returns; whereas U_5 and U_6 show some complexity with multiple peaks, suggesting a more intricate and possibly non-linear relationship.

Figure 3.3 plots the SVAR coefficients on the left panel and IS on the right panel for CEX order flow and transactions across urgency scores. U_1 has the highest short-term effect (VAR t-stats) but lower informational share, this means it is more about temporary impact or reaction-based trading. U_0 and CEX Orderflow dominate both in IS and SVAR suggesting long-term informational content, suggesting they are key drivers of efficient price discovery. Interestingly U_5 and U_6 have surprisingly high IS despite weak t-stat, maybe they carry private info or rare but impactful trades. This is consistent with Cong and He (2019), who argue that blockchain transactions, even when sparse or subtle in immediate impact, can contribute meaningfully to price discovery by publicly revealing verifiable, information-rich records.

The IRFs indicate that U_1 transactions, representing low urgency transactions, generate the most pronounced immediate impact on returns. The response peaks within the first few lags and subsequently decays. Despite this strong contempo-

raneous effect, the IS estimates for U_1 are relatively low. This suggests that such transactions, while influential in the short run, do not significantly shape the long-run efficient price. These patterns might be consistent with liquidity-driven or uninformed trading (e.g., Hasbrouck (1991).

Transactions categorized as U_3 and U_4 exhibit moderate price impacts in the IRFs, coupled with modest yet non-trivial information shares. This profile may be reflective of semi-informed agents—participants reacting to public signals with delay or aggregating dispersed signals. The persistence in the IRF supports the notion of gradual information diffusion, in line with theories of slow-moving capital Duffie (2010).

While U_5 and U_6 display weaker and more volatile IRFs, they account for disproportionately large shares of information in the Hasbrouck decomposition. This asymmetry highlights that these high-urgency trades—though less frequent and noisier—likely embed material private information that the market incorporates over time. This echoes the behavior of informed traders in microstructure models with asymmetric information (Easley and O'Hara (1987); Kyle (1985b)).

The divergence between IRF magnitudes and IS values across urgency scores underscores the value of distinguishing between price pressure and informativeness. Overall, our results suggest that high-urgency transactions (U_5, U_6) are associated with greater information asymmetry and contribute more significantly to price discovery, despite their lower short-term impact. In contrast, low-urgency flows (U_1) appear to exert strong transitory price pressure but offer limited information about future prices. Taken together, these results support the integration of urgency-based metrics in models of blockchain market microstructure and informed trading.

3.4.4 Trading strategy and performance analysis

To translate the informational content of urgency-typed transactions into practical execution strategies, we develop a signal-driven trading framework that dynamically adjusts position sizing based on the relative intensity of high-urgency activity (U_5 and U_6). The strategy scales exposure proportionally to the share of high-urgency transactions relative to total activity, capturing the intuition that larger urgency surges reflect stronger market signals. To avoid directional bias, we incorporate a momentum-based overlay, taking long or short positions depending on the sign of recent returns, which is consistent with price continuation behaviors observed

in crypto markets. To align execution timing with the observed impulse response decay patterns, we systematically optimize the holding period by evaluating rolling performance across multiple horizons.

Results are reported in Table 3.2. Our block-based strategy evaluation reveals that short-term execution aligned with immediate urgency signals remains optimal. Specifically, a 1-block holding period achieves the highest Sharpe ratio (0.0256) and the lowest drawdown (0.32%), suggesting that the market absorbs urgency-typed information almost instantaneously. As the holding period extends to 5, 10, and 20 blocks, both return and Sharpe ratio decline significantly, while drawdown increases. This performance decay mirrors the impulse response decay patterns identified earlier and suggests that urgency-typed signals carry short-lived informational content that is quickly incorporated into prices. Importantly, by structuring the holding periods in block units, we ensure that both the information share estimation (based on 10-block VARs) and the strategy evaluation are aligned on a common event-based horizon, avoiding any mismatch between blockchain-native informational flows and execution timing. These findings emphasize the need for high-frequency, event-driven execution frameworks when trading on blockchain-based signals.

To validate the robustness of these findings, we extend the analysis by testing the same urgency-weighted strategy under alternative directional regimes, including long-only, short-only. The results consistently show that momentum-aligned long/short positioning yields superior performance relative to static directional exposures. Both long-only and short-only implementations produce negligible returns with higher drawdowns, reinforcing the importance of directionally adaptive execution. These results suggest that urgency-typed blockchain transactions primarily generate short-term directional signals that require timely and adaptive trading strategies to realize their informational value.

3.5 Machine learning of Urgency Score informational impact

Building on structural evidence that urgency-typed Ethereum transactions carry heterogeneous informational content, we adopt a machine learning framework to assess their predictive value for future market behavior. Specifically, we implement XGBoost, a scalable and interpretable gradient boosting algorithm, to model two key prediction tasks: (1) short-term centralized exchange (CEX) price movements,

and (2) abnormal outflows from CEXs, which serve as a proxy for systemic stress.

Our feature set includes the six components of the Urgency Score for each block, the contemporaneous number of wallet-level transactions, a DeFi transaction flag, and the urgency imbalance, defined as the difference in transaction volume between urgency types U_1 and U_6. This approach enables us to evaluate whether urgency scores not only correlate with informational impact but also serve as forward-looking signals within the broader Ethereum trading ecosystem.

To evaluate model performance, we adopt a time-series-aware cross-validation strategy in which training strictly precedes testing, thereby eliminating look-ahead bias. Out-of-sample tests are conducted on a month-by-month basis, rather than as a single pooled forecast. This design reflects the reality of financial markets, where the relevance of predictive signals—especially urgency-typed transactions—evolves dynamically across time. Monthly testing enables us to track model performance and feature importance under different market regimes, volatility conditions, and liquidity cycles. It also allows for the real-time monitoring of signal degradation.

We use mean absolute error (MAE) and directional accuracy to capture both magnitude and sign of return forecasts. We also extract feature importance scores from the trained XGBoost models to identify which urgency components contribute most to each task, thereby uncovering the dimension-specific informativeness of urgency-typed transactions in explaining different market behaviors.

3.5.1 CEX Return predictions

The monthly out-of-sample evaluation reveals that urgency-typed Ethereum transactions carry significant predictive power for both short-term CEX returns and systemic stress events. Table 3.3 report the results. For the return prediction task, the model achieved stable performance across most months, with mean absolute error (MAE) decreasing over time and directional accuracy averaging around 45–48%, consistently outperforming a random walk benchmark.

The reduction in MAE and standard deviation over time suggests that the features used helped the model produce increasingly stable and precise predictions. The simultaneous decline in directional accuracy—particularly in 2023—implies that while the features captured average price behavior well, they may lack sensitivity to directional or regime shifts. The modest percentage of correct directional predictions (ranging around 45–48%, and dropping as low as 32% in some months) indicates that the feature set does not contain strong predictive signals for the sign of returns. This

could be due to noisy features, weak leading indicators, or a lack of high-frequency sentiment/macro inputs. The near-zero mean errors across most months suggest that the model does not systematically over- or under-predict returns, pointing to good calibration and feature scaling.

The model showed strong classification performance, particularly in months surrounding major market volatility (e.g., DeFi protocol hacks or large asset depeggings). The urgency imbalance (U_1 minus U_6) was consistently ranked among the top predictors, suggesting that shifts in urgency type composition offer early signals of pending liquidity stress. Feature importance analysis confirmed this pattern: the block transaction signal, urgency types 5 and 6, and the urgency imbalance were among the most frequently selected top features across months. Figure 3.6 visualizes the frequency of top-3 feature appearances over the 24-month period, revealing a clear dominance of urgency-derived variables. This supports the hypothesis that urgency scores encode latent private information not fully captured by traditional transaction volume or order flow. Notably, urgency types U_5 and U_6 exhibited dynamic predictive relevance. They featured prominently in high-frequency market adjustments.

These findings highlight the temporal heterogeneity of urgency informativeness: during stable periods, urgency scores primarily help refine return directionality, while during stressed periods, they act as early-warning indicators of broader market dislocations. Moreover, the XGBoost model's ability to dynamically re-weight these features each month reinforces the value of monthly evaluation.

To complement the individual feature importance analysis, we also examine the most frequent combinations of top features selected by the XGBoost model across the 24-month evaluation window. This analysis provides further insight into how joint patterns of urgency-related signals contribute to predictive power during distinct market phases. The results in Figures 3.4 and 3.5, summarized below, indicate a rotation in dominant feature combinations across time:

From January to August 2023 and again from October to December 2023, the model most frequently selected a combination of block transaction urgency 5, block transaction urgency 6, and the block transaction signal. This suggests that in later periods, the joint interaction between moderate and extreme urgency levels gained predictive relevance, likely reflecting the growing informativeness of wallet behavior under post-crisis volatility.

From February to September 2022, a different combination dominated: urgency balance, block transaction signal, and urgency 5. This blend reflects a more imbal-

anced signal framework, where differences in urgency types (U_1 vs. U_6) contributed alongside intermediate urgency types (U_5), possibly due to fragmented information diffusion during relatively stable periods.

From September to December 2023, the combination of urgency balance, block transaction signal, and urgency 6 rose in frequency. The increasing importance of U_6 in conjunction with imbalance metrics suggests an elevated market sensitivity to rare, potentially whale-driven or crisis-response transactions in late 2023.

The earliest period, January 2022, was characterized by a unique combination of urgency 2, urgency 5, and urgency balance, hinting at early signs of urgency imbalance mattering even when U_6 was largely inactive.

These temporal patterns highlight that predictive informativeness is not static. The market interprets urgency signals differently across time, and effective models must dynamically adapt. The ability of the XGBoost framework to shift its feature weighting accordingly justifies the month-by-month out-of-sample testing strategy and supports the idea that urgency-typed behavior contains both stable and regime-dependent informational components.

3.5.2 Abnormal CEX order flow predictions

To identify the most informative predictors of CEX order flow imbalance, we analyze the frequency with which various urgency-related features appear in the top three ranked positions across monthly XGBoost models spanning January 2022 to December 2023. The results, presented in Table 3.4 and visualized in Figure 3.7, offer clear evidence of the differential predictive value of Ethereum transaction features.

The urgency imbalance—defined emerges as the most consistently important feature, appearing in the top three across all 24 months. This consistent prominence highlights the central role of asymmetric urgency behavior in forecasting directional wallet imbalances, suggesting that divergences between low-urgency and high-urgency wallet activity carry systematic information about short-term shifts in liquidity flows.

Closely following urgency balance, the block transaction signal and urgency type 6 transactions also demonstrate high relevance, appearing in 23 and 22 months respectively. The signal, likely capturing transaction clustering or non-random timing within blocks, appears to be a robust proxy for information intensity or coordinated activity. Meanwhile, the prominence of urgency type 6 (characterized by rare, high-

fee, and potentially informed transactions) underscores its importance in signaling unusual market behavior, possibly linked to risk-sensitive flows.

In contrast, features such as urgency types 2 and 4, as well as block-level transaction count, rank in the top three in only a single month each. This infrequency suggests that these features either contain minimal predictive value or their relevance is highly context-specific and not generalizable across different market regimes.

These findings reinforce the notion that urgency-typed transactions are not homogeneous in their informativeness. Instead, a small subset—most notably the urgency balance and high-urgency transaction signals—consistently dominate the predictive landscape. This insight has practical implications for real-time monitoring systems and risk dashboards, which can prioritize these high-signal features to anticipate imbalances and stress in Ethereum-based trading environments.

The temporal analysis of top-3 feature combinations, as summarized in Figure 3.7, reveals distinct shifts in the dominant predictors of flow imbalance over the 24-month period from January 2022 to December 2023. These patterns provide insight into how the Ethereum market's informational structure evolves under varying market regimes.

The most persistent and recent combination—comprising urgency 5, urgency 6, and the block transaction signal—appeared in 11 of the 24 months, with a pronounced concentration. The co-occurrence of medium- and high-urgency transaction types (U_5 and U_6) alongside the signal feature suggests that in the current market environment, flow imbalances are increasingly driven by clusters of informed, high-priority transactions, possibly reflecting heightened speculative positioning.

In contrast, the earlier part of the sample (notably February to September 2022) was characterized by a different cluster: urgency balance, signal, and urgency 5. This combination, present in 8 months, underscores the relevance of asymmetric urgency dynamics during that period, with U_5 still playing a central role but complemented by the net urgency signal (U_1 minus U_6). This may reflect a regime where retail-driven urgency asymmetries carried stronger predictive power than they did in later periods.

A notable transition occurs in the final third of the sample (September–December 2023), where the combination of urgency balance, signal, and urgency 6 becomes prevalent. The replacement of U_5 with U_6 in these months may reflect a market increasingly responsive to rare, extreme, or potentially informed activity, possibly in response to greater macroeconomic uncertainty or the emergence of new

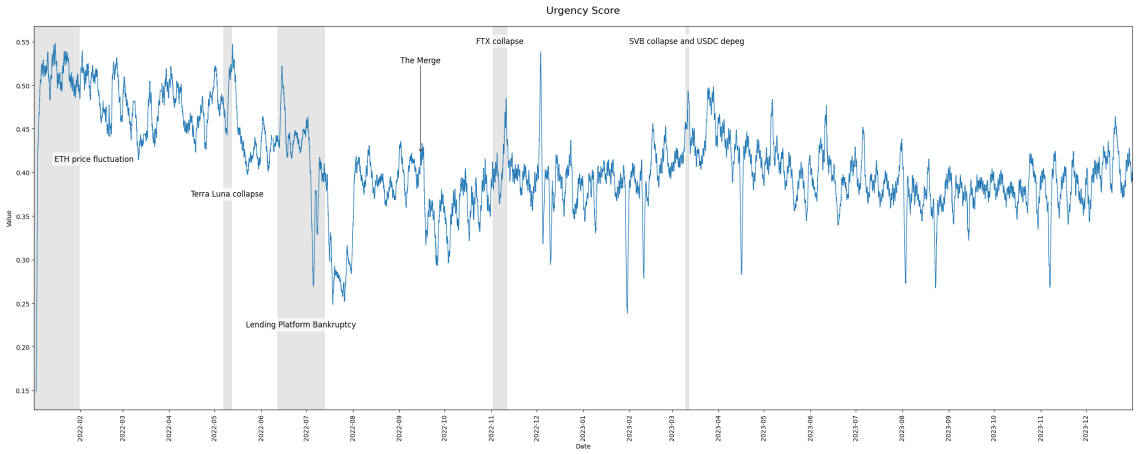
trading strategies exploiting high-urgency information. Lastly, the unique pattern observed in January 2022—which included urgency type 2—did not recur in subsequent months, suggesting it was either idiosyncratic or linked to transient structural conditions at the beginning of the sample period.

Overall, these evolving combinations indicate that the relative informativeness of urgency-typed features is non-stationary, reflecting shifts in trader behavior, market microstructure, and risk appetite. This underscores the value of a time-varying, feature-adaptive modeling approach, such as the monthly XGBoost framework employed in this study.

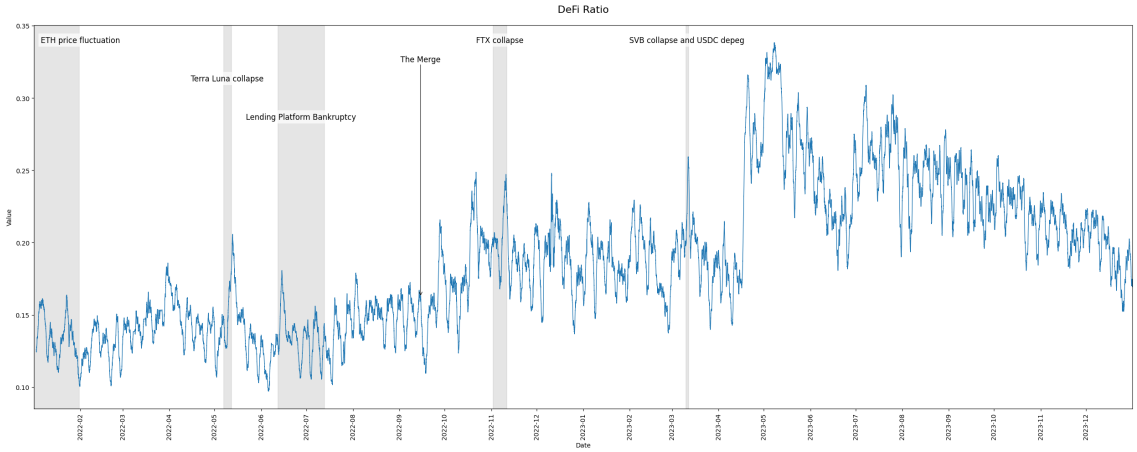
3.6 Conclusion

The inherent transparency of blockchain technology enables the collection and analysis of granular transaction-level data from the Ethereum network. In this study, we leverage this transparency to decode, structure, and build a novel dataset capturing the behavior of the most active wallets on-chain. Our primary objective is to link informational signals embedded in decentralized wallet activity to trading dynamics observed on centralized exchanges. Our findings show that the Urgency Score—a blockchain-native metric derived from transactional behavior—is time-varying and amplifies during periods of market stress. This suggests that on-chain urgency not only reflects shifts in trader execution behavior but also serves as a leading indicator of rising informational asymmetries and latent market fragility across decentralized and centralized venues. These insights have significant implications for both market practitioners and regulators. For market designers and liquidity providers, urgency-based signals offer a novel tool for anticipating liquidity shocks and enhancing real-time risk management systems. For regulators, our findings demonstrate the potential of blockchain analytics as a market surveillance tool, capable of improving transparency, monitoring systemic risk, and bridging the informational gap between decentralized networks and centralized financial infrastructures.

3.7 Figures



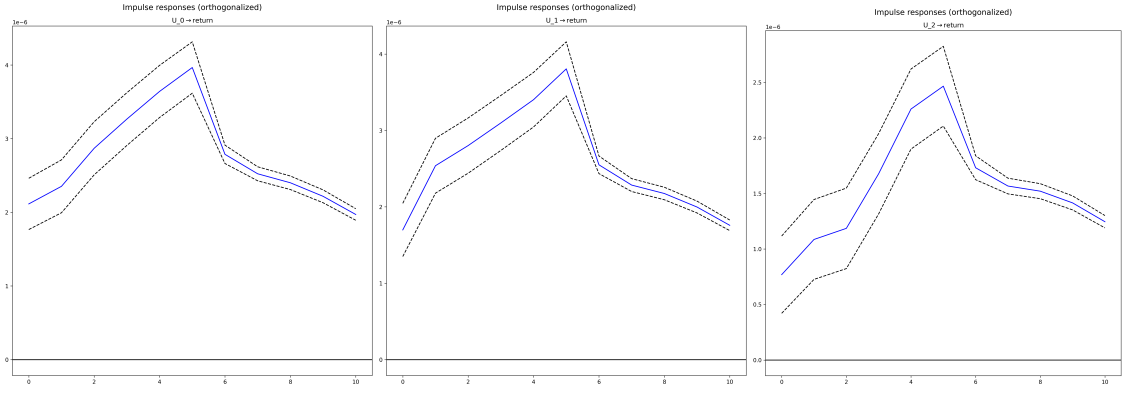
(a) Ethereum Urgency Score



(b) Percentage of DeFi big pools

Figure 3.1: (a) Ethereum Urgency Score; (b) Percentage of transactions originating from the major DeFi pools (Uniswap, Aave, Compound).

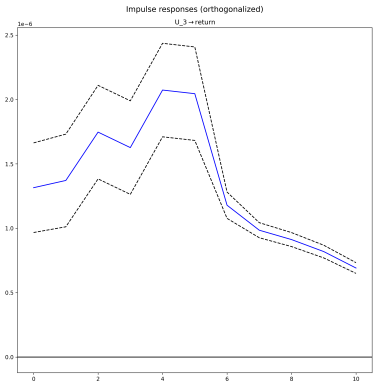
Notes: Panel (a) plots the time-series of our Urgency Score computed from Ethereum wallet data, 2022-2023. Panel (b) shows the DeFi-pool share over the same horizon.



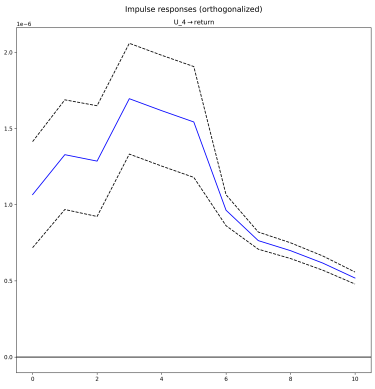
(a) U_0

(b) U_1

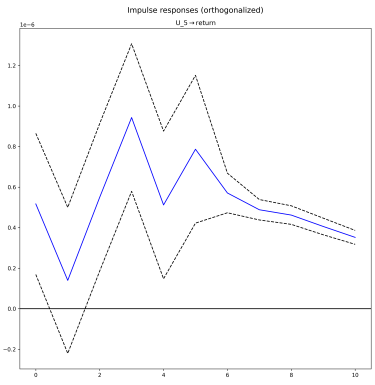
(c) U_2



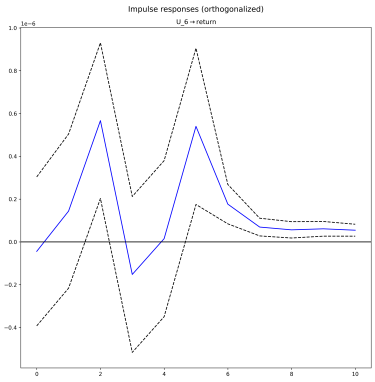
(d) U_3



(e) U_4



(f) U_5



(g) U_6

Figure 3.2: Impulse-response functions of CEX ETH price changes to one-standard-deviation shocks in transaction volumes by Urgency Score (U_0–U_6), estimated from a 10-lag Structural VAR.

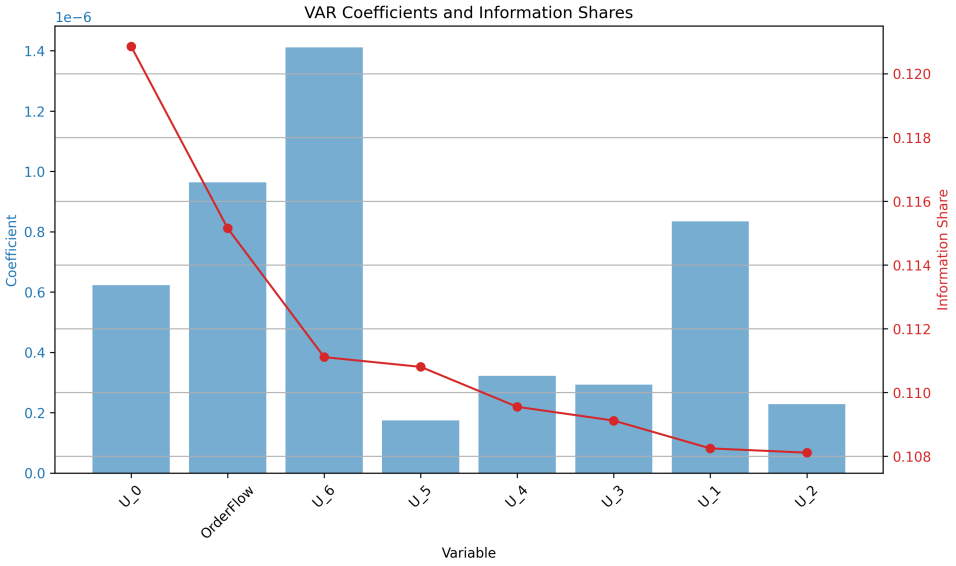
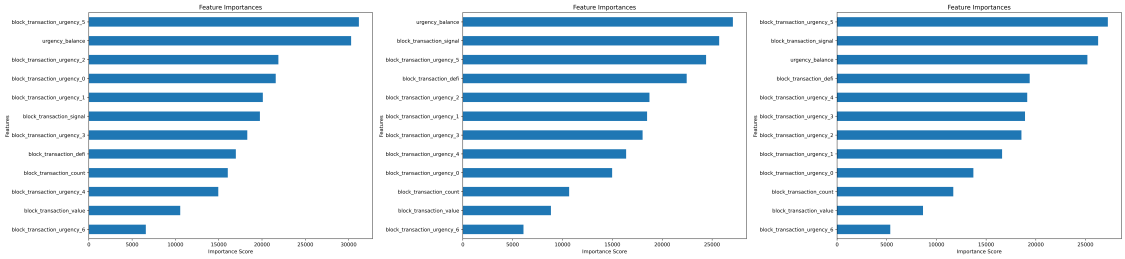


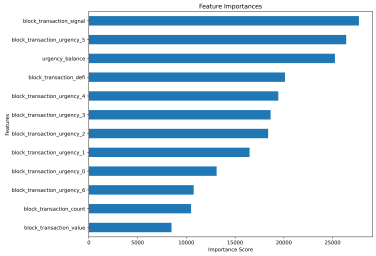
Figure 3.3: Short-term (SVAR coefficient) vs. long-term (Information Share) effects of transaction volume by Urgency Score on CEX price changes. Details in text.



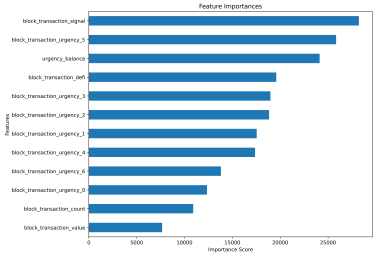
(a) Jan ’22

(b) Feb ’22

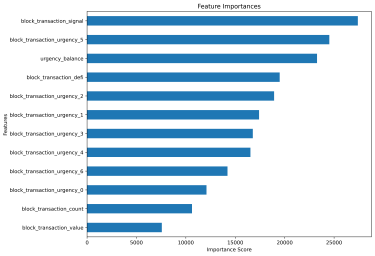
(c) Mar ’22



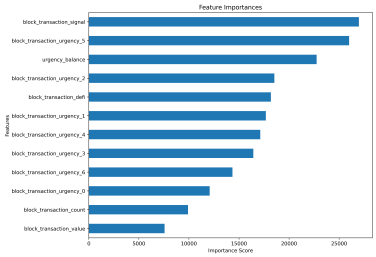
(d) Apr ’22



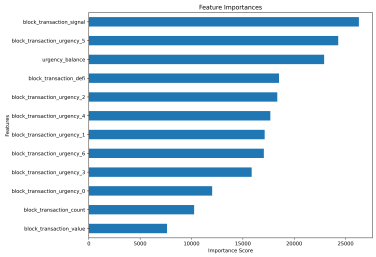
(e) May ’22



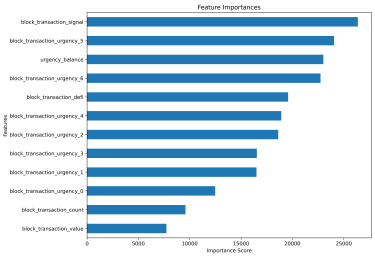
(f) Jun ’22



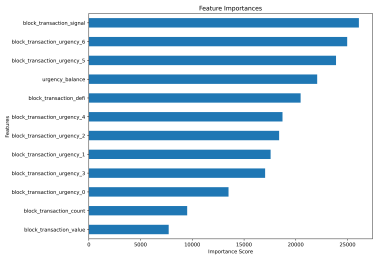
(g) Jul ’22



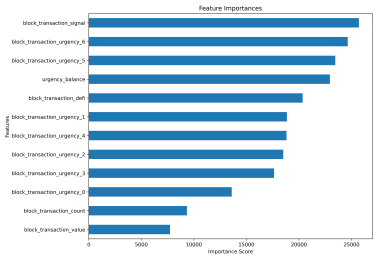
(h) Aug ’22



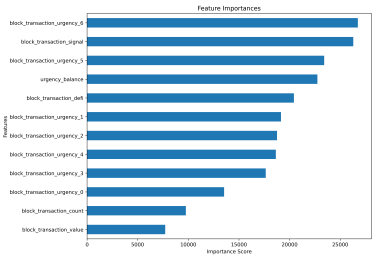
(i) Sep ’22



(j) Oct ’22

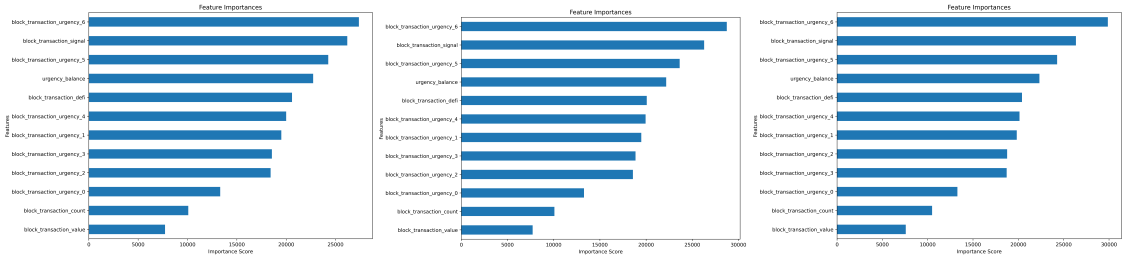


(k) Nov ’22

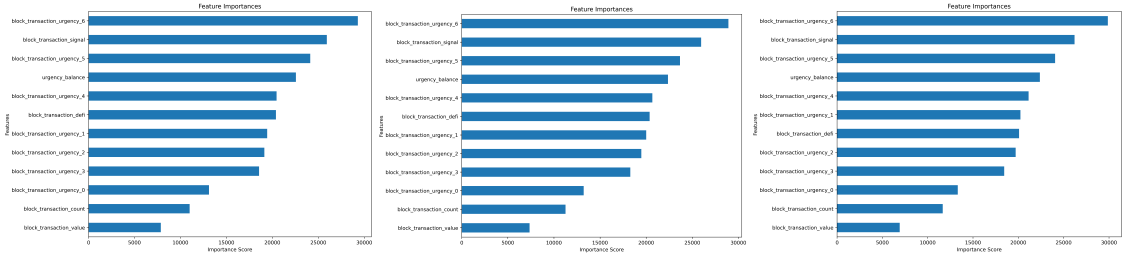


(l) Dec ’22

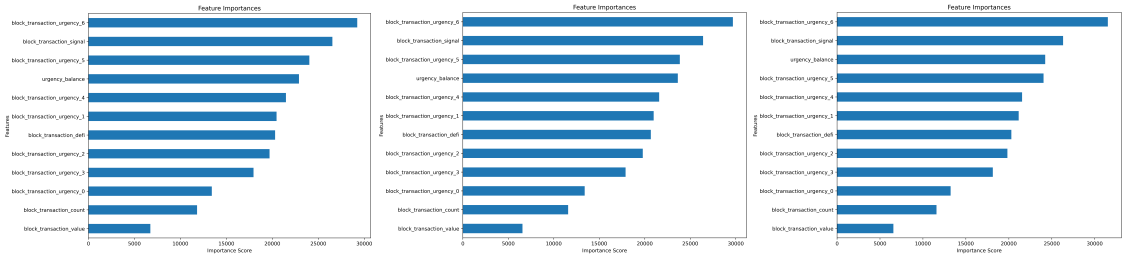
Figure 3.4: Feature-importance bar charts for CEX-return predictions (out-of- sample) in 2022.



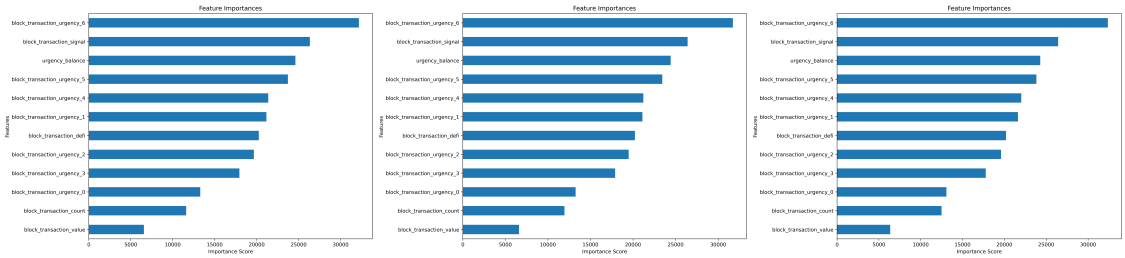
(a) Jan ’23 (b) Feb ’23 (c) Mar ’23



(d) Apr ’23 (e) May ’23 (f) Jun ’23



(g) Jul ’23 (h) Aug ’23 (i) Sep ’23



(j) Oct ’23 (k) Nov ’23 (l) Dec ’23

Figure 3.5: Feature-importance bar charts for CEX-return predictions (out-of- sample) in 2023.

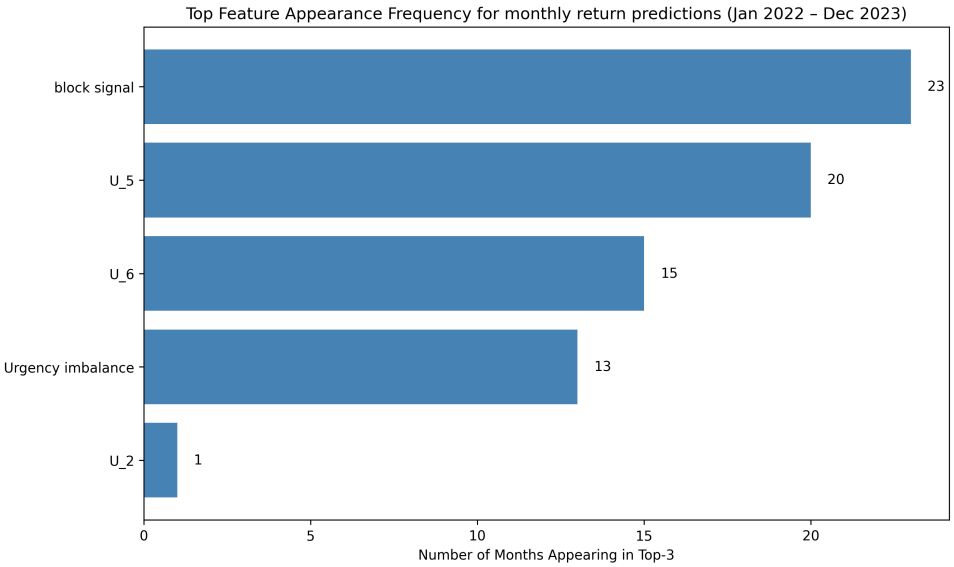


Figure 3.6: Frequency with which each feature appeared among the top-3 predictors of CEX returns (monthly XGBoost models).

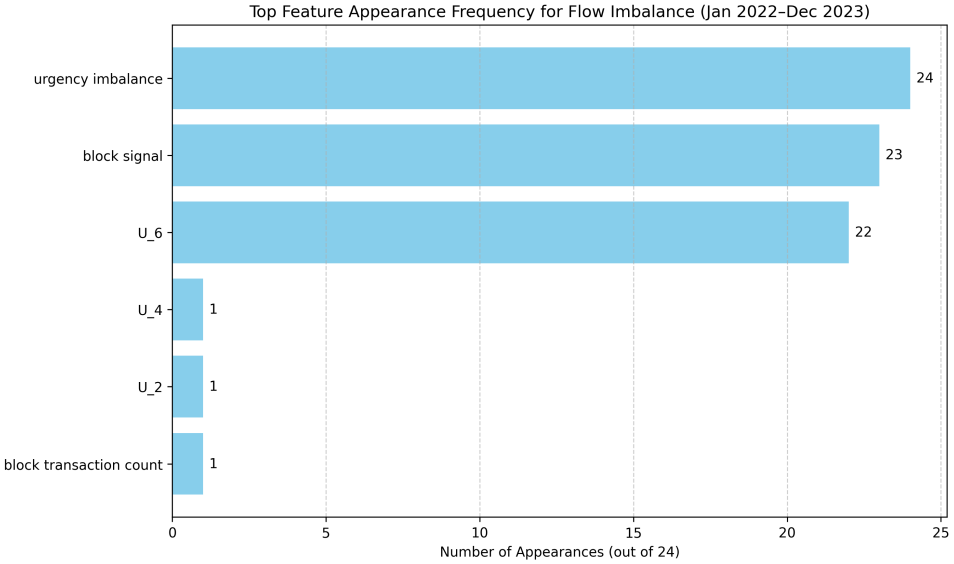


Figure 3.7: Frequency with which each feature appeared among the top-3 predictors of CEX flow imbalance (monthly XGBoost models).

3.8 Tables

Table 3.1: Percentage of transactions by urgency score across market regimes. Relative distribution of urgency-typed transactions (U_0 to U_6) across different market regimes, illustrating changes in transaction composition under varying return, order flow, and volatility conditions.

Urgency Score	**High Return**	**Low Return**	**High Flow**	**Low Flow**	**High Volatility**	**Low Volatility**
U_0	42.40%	42.40%	42.40%	42.40%	41.62%	43.13%
U_1	33.85%	33.85%	33.85%	33.85%	34.03%	33.68%
U_2	18.51%	18.53%	18.51%	18.53%	19.05%	18.00%
U_3	3.37%	3.39%	3.37%	3.39%	3.46%	3.30%
U_4	1.39%	1.38%	1.39%	1.38%	1.42%	1.35%
U_5	0.53%	0.52%	0.53%	0.52%	0.52%	0.52%
U_6	0.0028%	0.0025%	0.0028%	0.0025%	0.0028%	0.0025%

Table 3.2: Performance Across Block-Based Holding Periods for the Weighted Urgency-Momentum Strategy. Holding periods are defined in blockchain blocks. A holding period of 1 corresponds to the next block, 5 to the next 5 blocks. Metrics include Mean Return, Standard Deviation (Std Dev), Sharpe Ratio, and Maximum Drawdown.

Holding (Blocks)	**Mean Return (%)**	**Std Dev (%)**	**Sharpe Ratio**	**Max Drawdown (%)**
1	0.00267	0.1044	0.0256	0.32
5	0.00006	0.1070	0.0005	1.81
10	0.00010	0.1040	0.0009	1.24
20	0.00002	0.1133	0.0002	2.23

Table 3.3: Out-of-Sample Performance Metrics for CEX Return Predictions (January 2022 – December 2023). This table reports monthly evaluation metrics for return predictions on centralized exchanges (CEX). MAE denotes the Mean Absolute Error, measuring average prediction error magnitude. STDev represents the standard deviation of prediction errors, indicating volatility or consistency in errors. Mean is the average signed error, capturing any directional bias. % Correct refers to the percentage of times the model correctly predicted the direction of returns (positive or negative). Lower MAE and STDev indicate better precision, while higher % Correct suggests better directional accuracy.

Period	**MAE**	**STDev**	**Mean**	**% Correct**
January 2022	0.7976	1.3517	0.0073	47.21
February 2022	1.1179	1.7512	0.0064	48.34
March 2022	0.8535	1.3092	-0.0059	46.13
April 2022	0.7516	1.1807	-0.0083	45.31
May 2022	0.5730	0.9125	0.0081	46.84
June 2022	0.5817	0.8059	-0.0082	48.50
July 2022	0.7047	1.0755	-0.0043	47.50
August 2022	0.6667	1.0041	0.0038	47.87
September 2022	0.5323	0.7068	0.0022	48.68
October 2022	0.5113	0.6912	0.0006	47.92
November 2022	0.3851	0.4954	0.0047	45.91
December 2022	0.2032	0.2121	0.0004	42.59
January 2023	0.4582	0.5671	0.0008	47.74
February 2023	0.3740	0.4379	0.0002	45.57
March 2023	0.4867	0.5901	0.0023	47.94
April 2023	0.4132	0.4849	-0.0019	41.32
May 2023	0.4157	0.4140	0.0008	39.49
June 2023	0.4508	0.5472	0.0025	42.43
July 2023	0.3092	0.1994	-0.0009	32.16
August 2023	0.4043	0.5014	-0.0001	42.04
September 2023	0.3754	0.3597	0.0034	42.27
October 2023	0.4661	0.4279	0.0015	44.24
November 2023	0.6119	0.4644	0.0012	46.52
December 2023	0.6780	0.6712	-0.0033	47.15

Table 3.4: Out-of-Sample Performance Metrics for CEX Flow Imbalance Predictions (January 2022– December 2023) This table reports monthly evaluation metrics for flow imbalance predictions on centralized exchanges (CEX). MAE denotes the Mean Absolute Error, measuring average prediction error magnitude. STDev represents the standard deviation of prediction errors, indicating volatility or consistency in errors. Mean is the average signed error, capturing any directional bias. % Correct refers to the percentage of times the model correctly predicted the direction of returns (positive or negative). Lower MAE and STDev indicate better precision, while higher % Correct suggests better directional accuracy.

Period	**MAE**	**STDev**	**Mean**	**% Correct**
January 2022	110.3820	296.4784	5.5559	50.51
February 2022	156.2792	335.5266	-1.3975	50.03
March 2022	127.5332	273.1083	-4.8579	49.90
April 2022	116.9995	281.0211	-2.1198	49.97
May 2022	166.6903	409.9546	5.5058	49.38
June 2022	418.7761	890.4353	8.8763	49.80
July 2022	314.0130	669.8919	-10.8163	49.85
August 2022	502.3797	1031.4967	16.9519	50.07
September 2022	470.6890	906.4836	14.0598	49.88
October 2022	343.1570	653.5091	2.9111	49.78
November 2022	335.8033	604.1494	10.7677	49.77
December 2022	196.9775	370.3486	4.0795	50.14
January 2023	306.2022	512.4641	-6.3321	50.40
February 2023	269.7933	450.4441	-3.6311	50.16
March 2023	284.9438	441.8273	-0.0375	49.86
April 2023	232.9702	338.5729	-2.2751	49.74
May 2023	186.7331	80.7879	-0.8865	50.01
June 2023	137.8632	73.5100	-1.7614	49.45
July 2023	188.8987	238.7971	-1.9543	49.98
August 2023	211.5925	316.1794	0.1209	50.22
September 2023	200.0702	275.2573	1.0686	50.39
October 2023	266.2099	264.4734	-1.8563	50.04
November 2023	314.0047	297.1592	-1.6917	49.88
December 2023	333.7654	315.7043	-2.9726	50.10

Chapter 4

Intraday Momentum and Liquidity Crises

4.1 Introduction

In the market microstructure literature, liquidity is typically analyzed along two key dimensions: its level and its resilience. A low level of liquidity can signal a market's vulnerability to external shocks, making it less resistant to sudden changes. In contrast, liquid markets enable assets to be bought and sold quickly without causing significant price disruptions, thereby absorbing shocks and maintaining price stability (Brunnermeier and Pedersen, 2009b; Gabaix et al., 2006b). This function is particularly critical during periods of financial stress, ensuring that market participants can continue to trade and meet their liquidity needs.

The dash for cash during COVID-19 pandemic[1] and recent geopolitical tensions have highlighted how exogenous shocks can quickly overwhelm liquidity levels, pushing markets into distress and amplifying volatility. The theories of Brunnermeier and Pedersen (2009b); Gromb and Vayanos (2010) predict that higher volatility tightens funding constraints of market makers and thereby reduces their liquidity supply. Liquidity withdrawals by liquidity suppliers amplifies price swings and fuels momentum-driven reversals, as documented by Nagel (2012). Monitoring liquidity risk and illiquidity has thus become increasingly important from a financial stability perspective[2], echoing past flash crash episodes where sudden liquidity withdrawals

[1] https://www.wsj.com/articles/the-day-coronavirus-nearly-broke-the-financial-markets-11589982288.

[2] Reflecting on extreme events during the COVID-19 crisis[3], the Financial Stability Board's April 2024 report emphasizes the need for policies that address liquidity strains during heightened

exacerbated volatility in modern markets. Liquidity provision has become more selective, with market participants particularly high-frequency traders and hedge funds operating at different frequency, and more likely to withdraw from the markets at first sign of distress.

Measuring illiquidity in modern high-frequency markets remains a formidable challenge. Traditional metrics, such as the price impact of informed trades (Kyle, 1985b)[4] and the probability of informed trading, have been challenged in recent literature. Several studies argue that these measures fall short in capturing the nuanced dynamics of realized informed trading[5]—and even less so the broader systemic forces that contribute to liquidity crises[6]. On one hand, short-term liquidity needs are persistent at high-frequency level and can induce market freezes, as explored by Cespa and Vives (2015) in the context of modern short-termism. On the other hand, during periods of market stress, liquidity providers and arbitrageurs often withdraw from the market instead of absorbing large price deviations.

In this paper, we argue that a robust illiquidity measure during crises should reflect prevailing market conditions and the effect demand pressure has on liquidity displayed in limit order books by validating the limited risk-bearing capacity of intermediaries. Illiquidity should capture the short-term and rapid fluctuations in liquidity supply and demand, which are defining features of modern electronic markets[7]. During periods of market stress, one-sided order flow—driven by hedgers or momentum traders—can trigger aggressive trading in the same direction (e.g., sustained selling pressure). In response, liquidity providers adjust their quoting behavior to manage inventory risk (Amihud and Mendelson, 1980; Ho and Stoll, 1981). We introduce Quote Volatility (QV) to capture both the frequency and magnitude of bid-ask quote adjustments in short-term interval. The strategic quote updates generate frequent and asymmetric bid-ask oscillations, raising the costs associated with monitoring and responding to rapid quotes changes, as in Foucault et al. (2013),

margin and collateral calls in times of market stress.

[4]In Kyle's model, traders with private information attempt to execute trades without significantly impacting prices, as large price movements can reveal their informational advantage to others. However, excessive volatility often becomes detached from fundamental information and explained by risk premia at times of crises. Kondor and Vayanos (2019) predict excessive volatility is more reflective of market conditions and liquidity pressures rather than underlying fundamentals.

[5]See Augustin et al. (2019); Bogousslavsky et al. (2024); Collin-Dufresne and Fos (2015); Kacperczyk and Pagnotta (2019).

[6]See Andersen and Bondarenko (2014, 2015); Kyle and Obizhaeva (2023).

[7]Biais et al. (2015) offers a nuanced view on speed competition, emphasizing its role in generating negative externalities through strategic liquidity supply. Empirically, Hendershott and Menkveld (2014) study limit orders by high-frequency traders (HFTs) on the Canadian stock exchange, showing that the high frequency of submissions and cancellations reflects strategic liquidity provision.

which our QV metric captures in real time as indicators of transient illiquidity.

One can argue that QV may reflect noise or benign market making activity, not fragility. In fact, prior work including Conrad et al. (2015); Hasbrouck (2018)[8] shows that frequent quote updates can reflect healthy competition in high-frequency setting. In our framework, we distinguish contextual QV spikes — those coinciding with one-sided order flow and low book resiliency — as indicators of fragility. It's not only the frequency of quotes per se, but also their interaction with demand pressure and momentum trading that signals risk. In such instances, QV can function as an early-warning signal for liquidity spirals, especially during stress periods when persistent order flow imbalances exert downward pressure on prices. To complement this picture, we propose a second indicator, the Price Momentum ratio (PM), which captures the sequencing of aggressive trades on one side of the market—often associated with intraday selloffs—and the subsequent mean-reversion as liquidity recovers in the book. Both these metrics jointly develop and increase in the wake of wide selloffs[9]. Our metrics are detailed in the following section.

We further adopt a data-driven approach to evaluate the performance of these high-frequency metrics. Specifically, we train a Gradient Boosted Trees (GBT) algorithm to identify the persistence in liquidity needs at times of stress as captured by our ratios, on historical systemic liquidity events, including the Covid-19 deleveraging shock. The algorithm detects periods of rapid quote oscillation alongside price momentum, which is a key indicator of sharp price declines driven by selling pressure in illiquid markets. This method yields a real-time illiquidity measure capable of anticipating short-term liquidity disruptions and providing early warnings of potential liquidity spirals. We control for structural differences across asset classes by evaluating the out-of-sample predictive power and model explainability of our high-frequency metrics. Using SHAP values, we assess whether the importance of QV and PM remains consistent when the XGBoost model, trained on E-mini S&P 500 futures during the COVID-19 crisis, is applied to other asset classes such as US Treasury note futures and ETF Oil futures and to other stressful periods, such as geopolitical events of early January 2022, and April 2025. This cross-asset validation ensures that our framework captures universal liquidity stress patterns rather than asset-specific features.

[8]The authors employ variance ratios to suggest that frequent quote updates among quote setters in high-frequency markets typically reflect robust liquidity competition. In line with their findings, we observe elevated QV during stable market conditions. However, during periods of stress, QV peaks align with momentum-driven price dislocations, indicating a strategic retreat by liquidity providers in response to persistent selling pressure.

[9]We thank an anonymous referee for this suggestion.

This paper contributes to four strands of literature. First, it complements research on excessive and transitory price pressures in high-frequency markets. Bogousslavsky and Collin-Dufresne (2023) provide both a theoretical and empirical framework suggesting that high-frequency order imbalances are primarily driven by inventory risk rather than information risk at high-frequency setting. This goes in line with the evidence that algorithmic traders, as the new market makers, typically carry short inventory (Menkveld, 2013). Hendershott and Menkveld (2014) interpret price pressures as temporary deviations from fundamentals due to inventory constraints. In line with Baltussen et al. (2021b), we document strong intraday time-series momentum across futures contracts suggesting that hedging demand pressure results in persistent price pressure.

On market stress, Brogaard et al. (2018) analyze extreme price movements using high-frequency data by identifying large jumps associated with order imbalances, high-frequency trader (HFT) activity. Their framework captures realized dislocations, i.e. price jumps, rather than the latent fragility leading up to such events. In contrast, our proposed metrics identify environments prone to sudden liquidity-driven price dislocations before these jumps materialize. Related, Kyle and Obizhaeva (2023) demonstrate that large directional bets can drain liquidity, triggering large price deviations to the point of crashes. By comparing the 2010 Flash Crash to earlier market crashes, the authors highlight the amplifying role of trading speed in aggravating short-term price effects of large liquidations. Nagel (2012) finds that liquidity-driven price distortions can generate excessive momentum, followed by sharp mean reversion once liquidity returns. Our approach captures this phenomenon intraday of liquidity-driven price distortions. Intraday QV quantifies order book instability, serving as a high-frequency liquidity stress indicator that reflects the potential impact of large liquidations on market liquidity. By integrating QV with PM, our framework helps monitoring the dynamics that set the stage to liquidity crises.

Second, this paper contributes to the literature by demonstrating that a data-driven approach effectively detects liquidity crises in today's markets. We show that our proposed ratios, trained on systemic liquidity events, generalize well in out-of-sample tests across multiple markets, events and asset classes. To ensure robustness, we compare our approach with established illiquidity measures, including market depth, Volume-Synchronized Probability of Informed Trading (VPIN), and Amihud's illiquidity measure. Using feature importance analysis, we find that QV emerges as a key predictor of liquidity crisis dynamics, outperforming traditional measures. Market depth and VIX also contribute significantly, underscoring the

role of order book instability and volatility risk in market-wide liquidity stress. These findings reinforce the value of QV as a real-time indicator for identifying and monitoring liquidity disruptions.

Third, we contribute to financial stability policy and the interpretability of complex machine learning models in market regulation. Our results suggest that machine learning-based liquidity monitoring could help regulators distinguish between momentum trading in normal market conditions and periods where momentum trading amplifies liquidity crises. Price-level-driven circuit breakers—commonly used in markets—may exacerbate panic selling and market dysfunctions, as documented in Chen et al. (2024)[10]. Instead of rigid price thresholds, we argue that regulators should promote resilient market structures, enhance incentives for liquidity provision, and implement mechanisms that facilitate rapid order book recovery. An illiquidity-based mechanism would allow liquidity to recover organically, supporting natural price discovery and could mitigate the risk of self-reinforcing selloffs. In such resilient markets, traders are more likely to supply liquidity, thereby fostering a more cooperative and stable market equilibrium (Bessembinder et al., 2016).

Fourth, this paper contributes to the intersection of market microstructure, machine learning, and big data. We build on the discussions in Goldstein et al. (2021) on the application of machine learning in market microstructure research, distinguishing it from broader financial applications. Our approach aligns with machine learning methodologies in empirical asset pricing, as in Gu et al. (2020). Studies using reinforcement learning include Guéant and Manziuk (2019), which examines market-making under inventory constraints, and Kwan et al. (2021), which explores price discovery. Additionally, Q-learning algorithms have been employed in market-making strategies facing adverse selection risks Colliard et al. (2022), while AI-driven models have been used to analyze informed trading behavior Dou et al. (2024). Easley et al. (2021) combine random forest models with various adverse selection measures to predict liquidity and volatility changes throughout the trading day. Finally, Bogousslavsky et al. (2024) use machine learning to develop a new measure of informed trading based on insider transaction data. We apply machine learning to episodes of systemic liquidity crises to propose novel measures of illiquidity suitable in high-frequency setting.

The rest of the paper is organized as follows. Section 2 introduces the high-

[10]In the US markets, market circuit breakers are triggered when the S&P 500 index declines by at least 7% from the previous day's closing price, and trading will be halted for 150 minutes. The US market wide circuit breaker has been triggered on October 27, 1997 and four times during Covid-19 selloffs in March 2020.

frequency metrics. Section 3 describes the dataset, outlines the timeline of major liquidity events, and presents empirical results on time-varying illiquidity during crises. Section 4 explains how our data-driven approach can shed light on liquidity-induced crises. Section 5 presents the main results and their implications. Section 6 concludes. The Appendix provides additional details on the machine learning algorithms and includes further robustness checks.

4.2 High Frequency Quote Volatility and Price Momentum

In this section, we develop the two metrics to capture illiquidity and the associated price momentum at high frequency time scales. Both our proxies can be constructed from public data, such as Refinitiv, because they require as inputs only the best bid and ask quote updates, and the corresponding price and volume of the asset respectively.

Quote Volatility

Quote Volatility (QV) ratio is constructed from the order books and simply defined as the rate of oscillation between the best ask and the best bid in a short period of time, i.e. matter of seconds. An extreme selling pressure in the wake of sell-offs might lead to large price moves if the order book is one-sided. We propose to summarise the information from the limit order book using top-of-book prices as follows:

$$QV = \left(\frac{\sum_{i=1}^{T} |Ask_{i+1} - Ask_i|}{\max\left(|Ask_T - Ask_1| , Tick\ Size \right)} \right) + \left(\frac{\sum_{i=1}^{T} |Bid_{i+1} - Bid_i|}{\max\left(|Bid_T - Bid_1| , Tick\ Size \right)} \right), \tag{4.1}$$

where $\sum_{i=1}^{T} |Ask_{i+1} - Ask_i|$ is the sum of absolute incremental (instant-by-instant) changes in the ask price over the window T; $|Ask_T - Ask_1|$ is the absolute change in the ask price between the starting and the ending points of the window T; $\sum_{i=2}^{T} |Bid_{i+1} - Bid_i|$ is the sum of absolute incremental (instant-by-instant) changes in the bid price over the window; $|Bid_T - Bid_1|$ is the absolute change in the bid price between the ending and the starting point of the window T. *Tick Size* is the tick size, ensuring numerical stability if net price movement is zero.

We also consider the other alternatives of this proxy to detect the rapid increase

in quote volatility on each side of the market separately. The *Ask* specification detects the quote volatility at the ask side while activity at the bid side remains relatively unchanged. The *Bid* specification detects quote volatility which occurs at the bid side of the market, while the ask side activity remains unchanged.

$$QV_{Ask} = \frac{\frac{\sum_{i=1}^{T} |Ask_{i+1} - Ask_i|}{\max\left(|Ask_T - Ask_1|, Tick\ Size\right)}}{\frac{\sum_{i=1}^{T} |Bid_{i+1} - Bid_i|}{\max\left(|Bid_T - Bid_1|, Tick\ Size\right)}}. \tag{4.2}$$

$$QV_{Bid} = \frac{\frac{\sum_{i=1}^{T} |Bid_{i+1} - Bid_i|}{\max\left(|Bid_T - Bid_1|, Tick\ Size\right)}}{\frac{\sum_{i=1}^{T} |Ask_{i+1} - Ask_i|}{\max\left(|Ask_T - Ask_1|, Tick\ Size\right)}}. \tag{4.3}$$

Compared to variance ratio, QV provides a more granular measure of intraday price dynamics by directly incorporating bid-ask oscillations. Unlike variance ratio tests, which focus on mean reversion over longer horizons, QV is designed to detect rapid liquidity stress in real time due to inventory risk. Similarly, while bid-ask bounce models estimate transaction price reversals, QV captures a broader set of liquidity dislocations driven by order book instability.

Existing studies argue that the rise of machine-based trading led to an increase in quote volatility. For instance, Hasbrouck (2018) documents that high-frequency markets exhibit more volatility resulting from the undercutting behavior of limit order traders, suggesting healthy quote competition. However, in our analysis, the rapid burst in quote volatility suggests there is substantial noise in prices due to illiquidity. That is, because the demand pressure would have depressed prices, which, coupled with deleveraging, amplifies downside risk. Large price deviations should then signal potential large price reversals. To capture these episodes around liquidity crises, we next propose a price momentum ratio from the trading data.

Price Momentum

Price momentum is computed by taking the price difference between successive transactions for a fixed time interval in relation to the price at the start and at the end of that interval respectively. Then, we define the maximum difference which provides information on the large price moves downward or upward within that interval.

This involves estimating two "distances" for each trade (n) in a given window: $PM1_t = |P_t - \textit{Start Price}|$ and $PM2_t = |P_t - \textit{End Price}|$, and then deriving the largest sum $TPM_t = PM1_t + PM2_t$. We normalize this sum by the difference between prices at the start and at the end of that interval. Using PM ratio has an intuitive appeal, as this ratio would indicate cases in which price drops (increases) sharply and subsequently reverses.

$$PMR_w = \frac{\text{MaxTPM}_w}{\max\left(|\text{End Price} - \text{Start Price}|, \textit{Tick Size}\right)} \tag{4.4}$$

where

$$\text{MaxTPM}_w = \max_{t \in w} \left(|P_t - \text{Start Price}| + |P_t - \text{End Price}|\right)$$

To illustrate with an example on PM computation, consider the trading price of the WTI Oil futures was 34.05$ (start price) and reverted back to 34.04$ (end price) of a given 1-minute interval in the morning of 9th of March 2020. Within that one-minute interval, the price moved down to 33.8$ from 34.05$ before reverting back to 34.04$. This largest move defines the maximum price difference in that interval due to a sequence of sell-initiated transactions. We take the difference between 33.8$ and the start price and the end price respectively. The sum of these two differences which is 0.49 will be the numerator of the PM ratio. This yields a PM ratio of almost 168 for that minute interval. Figure 4.1 shows a sequence of episodes of large liquidations on the morning of March 9, 2020 for the WTI Oil futures. It depicts the variation of our ratios QV and PM for the WTI Oil futures during ten minutes between 10:30:34 to 10:41:31 on the 9th of March 2020. The top Figure shows the updates of the best bid and the best ask on the left and the corresponding QV on the right; the bottom Figure shows the trading price on the left and the corresponding PM on the right. Several observations are noted. First, large price moves seem to be followed by changes in the distribution of the bid/ask quote changes, suggesting that there is a shift of the bid and ask away from the pre-execution level with worse prices on future executions. Second, the increase in trading activity at these times, in intensity and size, has led to an increase in quote volatility. Third, price momentum increase seems to be persistent as QV develops, reflecting the impact of trading activity on liquidity. Finally, the shift in the ask and the bid combined with the large price moves downwards seem to perfectly coincide with the peaks of our two ratios that persist for several minutes.

4.3 Data and Systemic Liquidity Events

4.3.1 Data

We use intraday quote and trade data for the S&P 500 E-mini futures contract, US 10-year Note futures and the West Texas Intermediate (WTI) Oil futures collected from Refinitiv Tick History[11]. The data set contains a sequence of trades and Level 1 order book quote updates of each contract as reported by Chicago Mercantile Exchange (CME). The data feed is very detailed as each change in the order books or trade is recorded. It is worth mentioning that these futures contracts only trade on the Globex electronic trading platforms. Hence, there is no concern about unobserved trades occurring on other exchanges. We only look at the front-month contract for each month - the contract with the nearest expiration date which often has the highest volume and open interests of all open contracts. In Refinitiv, the identifiers for the front month futures are ESc1 for the E-mini, TYc1 for the US 10-year Note, and CLc1 for the WTI Oil. We also consider USO the Exchange-traded fund (ETF) that tracks the price of WTI crude Oil.

The level 1 quote data contain the best bid and the best ask quote updates with the corresponding quantities timestamped at millisecond frequency. The trade data contains common fields such as the price and the number of contracts traded timestamped in milliseconds. The identities of traders submitting, or changing their bids and offers, as well as those whose orders get executed are not disclosed publicly. Hence we could not identify whether it was Getco or Citadel which placed or executed a particular order. Still, the granular data on quote updates and the executed trades provide insights into the changes in order placement and the intensity of trading in low-latency environments on aggregate. Estimates suggest that more than 70% of trades are computer-driven in the US futures markets, of which high-frequency traders account for 60% of all futures volume in 2012 on CME, according to New York industry research Tabb Group.

4.3.2 Timeline of Main Liquidity Events

The volatility that resulted from the Covid-19 pandemic in March and April 2020 caused the sharpest price drops that major markets have ever experienced. Over a period of six weeks, volatility coupled with extreme liquidity conditions led credit investors to exit the market in numbers which surpassed those experienced during

[11]The successor to the Thomson Reuters Tick History database.

the global financial crisis. Indeed, liquidity spirals amplified and led to spread the initial shock when selling leads to more selling, higher margin requirements and large withdrawals of capital. These events were certainly tail risk events which we examine to propose a framework to monitor liquidity risk using our measures throughout this unprecedented market stress. The following are the main events:

On March 9 known by black Monday, the first Covid-19 real crash occurs. Black Monday was a combination of the Oil price war, a crash in equities and escalation of the virus. Three days later, on the 12 known by the black Thursday, E-mini futures dropped more than 200 points in less than one hour. On March 16 known by the black Monday II, another large price drops occur due to wide market selloffs across markets and assets. The S&P 500 triggered a Level 1 circuit breakers during the opening hour on March 9, 12 and 16 which results in trading being halted for 15 minutes for the E-mini S&P 500, and S&P 500 futures and options. The circuit breakers tripped midday on March 18 [12]. Clearing houses increased initial margin requirements. For instance, E-mini margins were increased from $ 6,300 per contract until March 2, 2020 to reach $ 12,000 per contract by March 23, 2020.

To illustrate the significant impact of panic selling, we plot price and price volatility and a time line reviewing the main events for each contract during the period from March 2019 till April 2023. Figure 4.2 shows how prices dropped and rebounded from beginning of March 2020 till the end of April 2020. It also documents an amazing drop in prices driven by the run to exit for the E-mini S&P 500 futures. Similar patterns are observed for NYMEX WTI Oil futures for Covid-19 sell-offs. The price path is similar to the theoretical price path when everyone runs for the exit, as predicted by Brunnermeier and Pedersen (2005).

Figure 4.2 also depicts the highest peak of volatility on WTI futures occurred at times Oil prices went to negative territory on the 20th of April 2020 in a historical event. On that day, the May contract for crude oil benchmark WTI prices fell below zero to -$37.63. The US Energy Information Agency claims the two main reasons for this to happen: limited available storage and low liquidity[13]. On the one hand, the impact of Covid-2019 on economic activity and the consumption of petroleum products led to an extreme demand shock on crude oil volumes that has been placed into storage. On the other hand, the WTI front-month futures contract was for May 2020 delivery, and the contract was set to expire on April 21,

[12]Trading also halts on the DJIA and the Nasdaq Composite when a circuit breaker is triggered on the S&P 500.

[13]https://www.eia.gov/petroleum/weekly/archive/2020/200422/includes/analysis_print.php.

2020. Market participants that held WTI futures contracts to expiration must take physical delivery of WTI crude oil in Oklahoma. Typically, most futures traders close contracts ahead of expiration through cash settlement in order to avoid taking physical delivery. This led to more selling at lower prices, even negative prices in very illiquid markets.

The figures illustrate that markets became extremely volatile on the 26 November 2021 known by the Black Monday of 2021 due to the growing concerns over the new Omicron Covid-19 variant. On the 26th of January 2022, another market selloffs picked up due to the geopolitical uncertainty. Amid all the uncertainty, between November 2021 and February 2022, Figure 4.2 shows that E-mini price dropped and rebounded in thin markets as shown by low level of top-book-depth. This provides a clear signal of rush to exit with extreme liquidity needs. As for the WTI Oil futures, the price increased early 2022 as Oil and energy commodities in general have gathered more interest from investors in the wake of the geopolitical events.

Furthermore, The US 10-year Treasury price that trading became disordered due to lack of liquidity as of March 9 on the US Treasury futures markets. In the wake of Covid-19, market regulators were concerned about the functioning of the world's most liquid debt markets. Market analysts point to the unwinding of relative value trades as a contributing factor to the high volatility in the treasury markets. The claim is that leveraged investors that bought Treasuries in the cash market and hedged the interest rate risk with futures contracts started unwinding these positions as futures prices rose, leading to a feedback loop of lower prices and greater sales in the cash market, as described by Schrimpf et al. (2020). This caused price volatility and margins to further increase. The Federal Reserve responded aggressively by purchasing $775 billion in Treasury securities on March 15 in order to support the smooth functioning of the markets.

4.3.3 Deterioration of market liquidity

Figure 4.3 shows the VIX index which is the volatility of the S&P 500 equity index as implied by the options markets. VIX may be related to funding liquidity as many institutions are exposed to VIX directly or indirectly. The 16th of March peak of 82.69 was a historical high for the VIX. Covid-19 volatility remained higher for longer. It has been associated with a deterioration in liquidity, indicating a link between market liquidity, funding liquidity and volatility as explained by the theory.

As volatility picked up and margins widen during Covid-19 crises, market liquid-

ity dried up in one market after another. Market liquidity, as measured by the depth of the market at the best quotes, collapsed at times of Covid-19 crises. Top-of-book depth refers to the quantities offered at the best ask and best bid in the order book available for immediate execution. This measure is another way to capture price impact of large liquidations as eventually large orders walk down the books to get executed. The patterns of low top-of-book depth and extreme high volume explain the violent price moves during moment of crises as the quantities at the best quotes are quickly and frequently depleted[14].

We report the distribution of properties of liquidity measured by the top-of-book depth for the Covid-19 sample as opposed to quiet days. We include statistics for the May 2010 flash crash for comparison. Table 4.1 shows the median values top-of-book depth for each futures contract, trading volume measured in contracts and the number of trades, based on one-minute intraday snapshots. We aggregate all orders executed at the same time, same millisecond, and at the same price, to be part of the same trade and are aggregated accordingly.

During normal times such as in March - April 2019, results suggest that the E-mini trades usually about 193 contracts for a volume of 1,591 and for 224 trades per minute as shown in the Table 4.1. That number fell to 30 in March 2020 for a much higher volume and higher trading intensity, 2,811 contracts and 846 number of trades. We observe the same pattern for the other two contracts. Given the statistics of the top-of-book depth, the volume and the number of trades in Table 4.1, it is unreasonable to think that the market will match those trades at the published best bid/ask prices for each contract without violent market moves. As uncertainty increases and market sentiment turned to negative due to the virus outbreak, forced liquidations of margin accounts over several days with little depth exerted a permanent downward pressure on prices that led to the collapse in prices.

Time-varying illiquidity and intraday momentum

Theories that emphasize the equilibrium risk-bearing capacity of the market as a whole imply that illiquidity should rise with exogenous risk. In such circumstances, they suggest that demand pressure generates intraday momentum trading which makes markets fragile and prone to crashes. Figure 1A1 shows evidence of intraday momentum trading by extending the results of Baltussen et al. (2021b) on more recent period including Covid sample. Results on scatterplots and regressions are

[14]It is also worth noting that the daily mean E-mini futures top-of-book depth has already been declining over the years with concerns of violent market moves. Such declining top-of-book depth over the years shows weakening equity futures liquidity, as highlighted further by the exhibit 7 from Goldman Sachs Global Investment Research in February 2022, and reported by CME.

reported in Figure IA1 and Table IA1 respectively. Two aspects are to examine in the context of our study using our measures: time-varying illiquidity associated with an increasing intraday price momentum.

We compute QV and PM time series for each contract, and match them by specific time point. Each data point represents episodes of quote volatility and the associated price momentum. The rationale is to capture the impact of selling pressure on liquidity. That is the price momentum that fluctuates within one minute, and with it the impact that quote changes and a sequence of transactions have on prices. Two consecutive data points have a fixed time difference of 1 minute.

To illustrate illiquidity at times of flash crash, Figure IA2 shows unique data points for the Emini S&P 500 that develop and increase throughout the day of 2010 flash crash on the 6th of May 2010, that are not comparable to the adjacent days to the crash. The May 2010 was accompanied by a record of high frequency trading activity, extreme volatility and very low levels of market liquidity. Kirilenko et al. (2017a) document that buyers and sellers executed large trades relatively to those posted by liquidity providers in the order book, leading to an increase in price impact. The breakdown in provision of liquidity in the period leading to the crash is depicted by the high values of QV, and usual large realization of PM. Similar episodes have been observed during the wide market selloffs originated by the virus outbreak in March 2020, and the day of Oil crash on the 20th of April 2020. Detecting short-term liquidity evaporation is relevant across different market crashes. Liquidity evaporation is a common feature in both oil crashes and the 2010 Flash Crash, even though their causes differ. These patterns are observed during Covid-19 selloffs, and during the war periods.

We rely on these observations to formally analyze these patterns in the next section.

4.3.4 Features and labels

We analyze several measures used as warnings in the literature such as VIX and VPIN. VPIN is grounded in adverse selection theory and has been recognized as a warning signal for liquidity crises. Easley et al. (2021) demonstrate that VPIN spiked to record levels in the hour preceding the 2010 flash crash, proposing order-flow toxicity-induced liquidity crises as a hypothesis. In the context of our study, the assumption that liquidity traders are equally likely to trade in either direction may not hold, as suggested by the structural estimation of the VPIN model. For example,

distressed liquidity traders may continue to trade in the direction of price movements during crises. Furthermore, we examine rapid quote changes as natural responses of market makers reacting to inventory risk due to demand pressure. Therefore, it is crucial to evaluate the performance of VPIN using our proposed metrics. We incorporate volatility indices for US adn Japanese front-month futures, utilizing tick-level trade data from Refinitiv for the sample period.

4.4 Detecting illiquidity in high frequency setting

Detecting illiquidity during times of crisis presents unique challenges that undermine the effectiveness of traditional liquidity measures. Conventional metrics built to detect adverse selection costs, often fail to reflect the fragility of liquidity supply in stressed environments. At these times, liquidity providers may strategically reduce quote sizes or rapidly withdraw quotes in order to manage their exposure to inventory risk. Moreover, the nonlinear and self-reinforcing nature of market dynamics during crises—characterized by abrupt shifts in order flow, volatility spikes, and forced deleveraging—makes it difficult to rely on static or linear indicators. Cross-asset contagion, the dominance of latency-sensitive strategies further complicate real-time assessment. These limitations call for alternative metrics that can capture the high-frequency instability in liquidity provision.

Our proposed QV metric addresses this gap by tracking the frequency and magnitude of bid-ask quote updates, a defining feature in modern markets. QV is particularly sensitive to microstructural stress, offering a forward-looking signal of potential liquidity withdrawal and market dysfunction. When combined with price momentum, QV enables early detection of liquidity spirals—providing regulators and risk managers with a more responsive and interpretable indicator during periods of systemic stress. We further use machine learning models to account for nonlinearities and interactions between variables. After the model is estimated on a training sample of E-mini during Covid-19, we test it across assets and markets and out-of-sample. We also use a placebo sample from 2019 as non-crises periods.

4.4.1 Crises signals

We find that our two proposed metrics—Quote Volatility (QV) and Price Momentum (PM)—tend to jointly spike at the onset of systemic liquidity events, including the 2010 Flash Crash, the Covid-19 market meltdown, and geopolitical tensions such

as wars and trade disruptions. To systematically detect these episodes, we define a crisis signal whenever both QV and PM exceed their 90th percentile thresholds within a rolling window. This thresholding captures the extreme co-movement of illiquidity and directional pressure that typically precedes market breakdowns.

As shown in Figure 4.4, the share of periods classified as crash episodes based on this definition spikes during early 2020, coinciding with the onset of the Covid-19 pandemic. These episodes are more frequent and persistent than in normal times, reflecting both the initial liquidity shock and the prolonged volatility as the crisis evolved. In contrast, during pre-pandemic periods marked by economic stability, the joint exceedance of QV and PM is rare, and thus crash signals are infrequent. Importantly, we observe that Federal Reserve interventions in mid-March 2020, aimed at stabilizing the US Treasury market, coincide with a reduction in the percentage of crisis signals, despite elevated volatility levels. This suggests that macro-prudential interventions may suppress illiquidity spirals, even if the underlying uncertainty remains.

Formally, we define the crisis signal for each futures contract using a binary indicator function:

$$\mathbb{I}_t = \begin{cases} 1, \text{ if QV and PM} > \text{90th percentile at time } t \\ 0, \text{ otherwise} \end{cases}$$

This framework enables real-time monitoring of market fragility across contracts and can serve as an early warning system for flash crashes or liquidity-driven market stress.

Our approach to crisis predictability is rooted in a binary classification framework that maps high-frequency market behavior into two distinct states: crisis (1) and non-crisis (0). We frame this as a supervised learning problem where models are trained to distinguish episodes of market fragility based on our two metrics—Quote Volatility (QV) and Price Momentum (PM).

As a baseline, we implement logistic regression, which is widely used in financial applications for binary classification. To capture potential non-linear patterns in the data, we also incorporate advanced machine learning techniques, including Artificial Neural Networks (ANNs) and tree-based models. Prior research has shown that neural networks are particularly effective in predicting financial crises in low-frequency data, such as currency and sovereign debt crises (Fioramanti (2008), among others). We extend this logic to high-frequency financial markets.

Our modeling includes several state-of-the-art algorithms: the Long Short-Term Memory (LSTM) network, its Bidirectional variant (BiLSTM), the Gated Recurrent Unit (GRU), and the Gradient Boosting Tree model (XGBoost). These models are selected to explore the capacity of both sequential and tree-based learning frameworks in forecasting high-frequency liquidity crises. Training is performed on crisis episodes identified during known stress events, particularly the Covid-19 turmoil of March 2020, using a labeled dataset constructed from the joint exceedance of QV and PM thresholds. To enhance generalizability and reduce overfitting, we apply a dropout rate of 20% across the deep learning models. Each model is trained for 100 epochs, with architecture specifications as follows: one input layer with 64 neurons, one hidden layer with 64 neurons, and a single output neuron for classification. The models are fed 3D input tensors representing samples, time steps, and features. Feature normalization and reshaping are handled using custom preprocessing pipelines (see Appendix C for implementation details).

We ultimately use XGBoost as our main modeling framework, owing to its strong balance between predictive performance and interpretability. As a gradient-boosted ensemble of decision trees, XGBoost produces structured and rule-based outputs that are more transparent than the layered transformations used in deep learning models. It also offers built-in feature importance metrics and integrates seamlessly with post-hoc explainability tools like Shapley Additive Explanations (SHAP), enabling us to attribute predictions to specific input features with precision. This is particularly critical in financial applications where regulatory transparency and academic interpretability are paramount. While XGBoost is the model for reporting and analysis in this paper given its best-in-class status for structured, tabular data—our main findings on liquidity crisis predictability are robust across models.[15]

4.5 Results on intraday crash probabilities

4.5.1 Cross-Asset Out-of-Sample Predictions

The model is trained exclusively on E-mini S&P 500 (ES) and then tested out-of-sample on other assets such as USO (Crude Oil ETF) and Treasury futures. This tests cross-asset transferability i.e., whether the model generalizes beyond equities

[15]XGBoost is widely regarded as a top-tier algorithm for supervised learning on structured datasets. It implements regularized gradient boosting, supports parallel training, and handles missing data efficiently. Its scalability and predictive accuracy have made it the preferred algorithm in benchmark environments such as Kaggle competitions.

to other asset classes with different fundamentals. Figure 4.5 evaluates the out-of-sample performance of a crash probability model trained solely on E-mini S&P 500 (ES) data when applied to distinct asset classes: USO (a crude oil ETF) and U.S. Treasury futures. Despite being trained on equity index data, the model successfully identifies rising crash risk ahead of major price declines in both test assets during the COVID-19 crisis of early 2020. In the USO panel, predicted probabilities exhibit frequent and sharp spikes in March–April 2020, aligning closely with the observed collapse in oil prices. Similarly, in the Treasury panel, elevated crash probabilities emerge in mid-March—just prior to the dislocation in Treasury markets documented during the dash-for-cash episode.

Despite structural and market microstructure differences across equities, commodities, and sovereign bonds, the model identifies elevated crash probabilities ahead of major price dislocations in both test assets during the COVID-19 crisis. These results highlight the portability of liquidity-driven signals across asset classes, especially in times of systemic stress.

Quantitatively, the model performs robustly well on the test set in detecting the majority class (non-alert periods). It achieves a specificity of 97.5%, correctly classifying 2,660 out of 2,728 normal (non-alert) instances. For the minority class—potential crash periods—it identifies 422 out of 634 cases, yielding a recall of 66.5%. The precision for the positive class stands at 86.1%, reflecting relatively few false alarms. This trade-off results in an F1-score of 75.1%, demonstrating a strong balance between sensitivity and precision. In financial settings where false positives can incur unnecessary hedging or trading costs, this level of model stability is desirable. The model's ability to maintain high specificity while still capturing meaningful risk signals supports its practical use as an early-warning tool in risk management and trading systems.

Explaining why ML methods work especially in financial markets is a well-known challenge. As Nagel (2021) highlights, machine learning models often provide strong predictive power but lack clear economic interpretation, making it difficult to link their success to underlying market mechanisms. We address this issue when applying ML to detect liquidity evaporation and price crashes. While ML can identify patterns in order book instability QV and price momentum PM, explaining why these features matter requires grounding the model in established theories such as Nagel (2012) on mean-reverting price distortions. We quantify the relative importance of different characteristics for the prediction of crash probabilities. We follow recent advances in computer science and estimate SHAP values **?**, which approxi-

mate changes in the model predictions had we excluded certain characteristics in its estimation.

To evaluate the relative importance and timeliness of different signal families, we combine a first-occurrence analysis with Shapley value decompositions. Shapley values, grounded in cooperative game theory, quantify each feature's marginal contribution to model performance by averaging over all possible feature permutations. Figure 4.6 demonstrates that signals from the qv and pm families consistently emerge earlier across asset classes, suggesting their potential role as leading indicators. This early activation is corroborated by the Shapley-based importance rankings in Figure 4.7, where qv features dominate the top-5 positions, accounting for 60% to 80% of the explanatory power across assets. In contrast, features from the vix family appear uniformly late in the signal sequence and contribute nothing to the top-5 Shapley ranks, indicating limited incremental explanatory value. These results underscore that early-appearing features, particularly from the qv family, tend to offer the most predictive value, consistent with the notion that features delivering timely and distinct information are most influential in high-frequency settings.

To summarize, the dual approach combining Shapley-based importance with signal activation timing enables us to assess both the relevance and responsiveness of each feature family. Results confirm that qv and pm signals are not only the most impactful but also among the earliest to activate, providing early warning of liquidity stress and aligning with theoretical expectations about order book instability and price distortions.

4.5.2 Liquidity Consequences of Crash Probability Shocks

The preceding sections show that machine learning models with our qv and pm ratios, when grounded in market microstructure theory and interpreted through Shapley values, can detect early warning signals of crash risk with high predictive values. To establish the real-world relevance of these signals, it is crucial to assess how predicted crash risk maps onto actual market conditions. Specifically, we examine whether exogenous increases in crash probability correspond to measurable signs of liquidity stress namely, widening bid-ask spreads, thinning order book depth, and surging trade volume. These market responses are essential for validating that the crash probability signal does not merely reflect statistical correlations but captures economically significant risk dynamics.

Figures 4.9 present impulse-response estimates quantifying the effect of a shock

to the predicted crash probability on three core liquidity dimensions. The results reveal a pronounced divergence between crisis and non-crisis regimes. During crisis periods including the 2010 Flash Crash, COVID-19 in 2020, the 2022 inflationary tightening cycle, and recent instability in 2024–25 a rise in crash probability leads to immediate and persistent liquidity deterioration: spreads widen top panel), depth collapses (middle), and trade volume spikes (bottom). These effects unfold within 5–15 minutes, aligning with the intraday mechanics of market dislocation.

Importantly, these effects are largely absent during relatively quiet periods (2019 and 2021), where the same crash probability signal produces no meaningful shifts in liquidity metrics. This regime-dependence serves as a key robustness check: it rules out the possibility that our model simply captures noise or generic volatility. Instead, it suggests that the crash probability signal encodes latent fragility that is amplified in stressed environments but muted under normal conditions a result consistent with the theoretical predictions of Nagel (2012) and other models of endogenous liquidity dry-ups.

From a methodological standpoint, this impulse-response approach also addresses concerns about post-model interpretability: it demonstrates that predicted crash risk is not only statistically significant but economically actionable, manifesting in observable microstructure changes that are costly to liquidity providers and risk managers. The consistency of results across multiple crisis episodes further strengthens their generalizability.

4.5.3 Real-Time Performance During Market Stress

Having demonstrated the predictive salience of different signal families across a broad cross-asset panel, we now examine how these insights translate to out-of-sample performance in real-world crash episodes. The goal is to assess whether the machine learning model not only generalizes across asset classes but also anticipates known episodes of market distress. To that end, we focus on test windows from two distinct events: the 2010 Flash Crash in U.S. equity futures (ES), and the August 2024 drawdown in Nikkei futures (NKF). These windows were not used in model training and thus serve as a stringent test of the model's ability to forecast sharp price corrections in a live setting.

Figure 4.8 displays the predicted crash probabilities (blue) alongside realized price action (red) for ES and NKF. In the ES panel (top), the model generates a steady build-up in predicted probability beginning well before the May 6, 2010 Flash

Crash. This anticipatory signal strengthens sharply in the hours leading up to the crash, aligning closely with the onset of disorderly price action. In the NKF panel (bottom), the model captures a series of elevated crash probabilities throughout the August 2024 stress window, preceding each major leg down in the price series. Importantly, these signals are not reactive, i.e., they do not merely spike after price drops but rather exhibit forward-looking behavior, consistent with early-warning dynamics.

Taken together, these results reinforce the model's capacity to translate microstructural signals into actionable crash probabilities. They provide further evidence that ML-based crash forecasting, when grounded in theory informed features and interpreted through Shapley-based methods, can successfully identify periods of latent fragility prior to observable price dislocations[16].

4.6 Daily aggregation of crash probabilities

While the previous section focused on the model's performance in identifying specific intraday crash events, this section shifts the perspective toward a more systematic, time-aggregated approach. Rather than analyzing model accuracy around known events, we examine whether elevated crash probabilities emerge in a diffuse, real-time setting—capturing periods of heightened market stress that may not be tied to a single identifiable shock.

To this end, we aggregate model predictions on a daily basis. Specifically, we evaluate each intraday prediction and identify instances where the predicted crash probability exceeds a given threshold (e.g., 10%, 25%, 50% or 75%). For each trading day, we compute the total number of such exceedances across all observations, providing a count of how often the model signaled elevated risk throughout that day. This count serves as a proxy for the intensity and breadth of crash-related signals in the market at different levels of confidence.

This approach is conceptually related to systemic risk indicators such as SRISK (Acharya et al., 2017), which quantifies the expected capital shortfall of financial institutions during market downturns, and CATFIN (Allen et al., 2012), which aggregates extreme quantile tail risks across institutions. Similarly, our framework draws parallels to forward-looking market-based indicators such as the VIX and high-frequency measures of tail risk (Kelly and Jiang, 2014), but adapts them to the

[16]Out-of-sample ROC-AUC scores exceed 80% across both ES and NKF test sets, indicating strong classification performance in forecasting crash events.

electronic market microstructure setting by leveraging prediction outputs directly. By visualizing the time series of daily signal intensities across varying thresholds, we distinguish between ambient low-confidence warnings and concentrated high-confidence alerts thus offering a flexible and scalable early warning framework for monitoring systemic risk buildup.

Figure 4.10 plots the intensity of predicted crash signals in the U.S. Treasury market across different confidence levels. Each subplot corresponds to a specific threshold for the model's predicted probability of a crash—greater than 10%, 25%, 50%, and 75% respectively. For each day, we count the number of instances where individual predictions exceeded the given threshold. This allows us to track how widespread elevated crash risks were across the market on any given day. Distinct market regimes are highlighted in the background: a quiet pre-crisis period (light blue), the COVID crash (light orange), and the recent tariff-induced volatility (light yellow). The results show that during crisis periods, not only does the number of high-probability predictions increase, but more predictions also exceed stricter thresholds, indicating both broader and deeper model-detected distress signals. The increase in counts during the yellow region, which corresponds to the period starting around March 2020, suggests that there were more days with higher predicted probabilities during the early months of the COVID-19 pandemic and during the volatile days induced by the tariffs. There is a noticeable increase in the frequency and magnitude of bars, indicating more days with higher predicted probabilities during these two time periods.

Figure 4.11 plots the results on the crash signal intensity in the E-mini market by prediction threshold. Distinct market regimes are highlighted in the background of each plot. The results show that during crises periods, including Covid-19 sell-offs (light orange), and the onset of war in 2022 (pink) and the recent tariff-induced volatility (light yellow), not only does the number of high-probability predictions increase, but more predictions also exceed stricter thresholds, indicating both broader and deeper model detected distress signals.

4.7 Conclusion

In this paper, we develop a new measure of illiquidity by integrating a high-frequency order book–based indicator designed to capture short-term oscillations in bid-ask quotes. We employ a machine learning algorithm to identify episodes of order book thinning and rapid quote reversals, incorporating additional variables related to

market depth, trading volume, and the VIX. The model is trained on systemic liquidity events — including the COVID-19 market selloff, the onset of the war in 2022, and recent tariff-induced volatility in 2025 — and then applied out-of-sample to major U.S. futures contracts such as 10-Year Treasury Note futures, WTI crude oil, and the USO oil ETF, as well as to the Nikkei futures during the August 2024 crash.

We demonstrate that the machine learning model accurately predicts illiquidity out-of-sample and serves as a strong predictor of systemic liquidity stress. Furthermore, we show that periods of rising intraday momentum trading coincide with our measure, consistent with the intuition that the price impact of large liquidations persists under illiquid conditions, often precipitating market crashes. We also aggregate the model's high-frequency outputs to a daily level, enabling comparisons with forward-looking indicators such as the VIX. This aggregation offers a flexible and scalable framework for real-time monitoring of systemic risk buildup.

The central implication of this study is that our data-driven, machine learning–based approach combined with a novel quote volatility measure provides a reliable, high-frequency signal of illiquidity in falling markets where intraday momentum persists, supporting the liquidity-based hypothesis for market crashes and offering practical value for market surveillance and risk management. During the recent trade tariffs–induced volatility in April 2025 in particular, markets experienced a sharp and sudden deterioration in liquidity conditions across key asset classes. The escalation of tariffs targeting semiconductor equipment and rare earth exports between major economies triggered broad risk-off behavior, leading to flight-to-safety flows and dislocations in the U.S. Treasury market. Despite an increase in trading volumes, Treasury market depth at the best quotes thinned significantly, and bid-ask spreads widened across the curve — especially in the 10-Year Note futures, where our model detected a spike in quote volatility and order book instability. These disruptions occurred even as post-2020 reforms in the U.S. Treasury market, including enhanced central clearing and improved transparency mandates, were meant to enhance market resilience. Our illiquidity measure flagged these stress signals in real time, suggesting that structural reforms, while helpful, may not be sufficient under sudden macroeconomic shocks. This episode underscores the model's ability to serve as a high-frequency complement to existing regulatory surveillance tools, with implications for both monetary policy execution and market stability assessments.

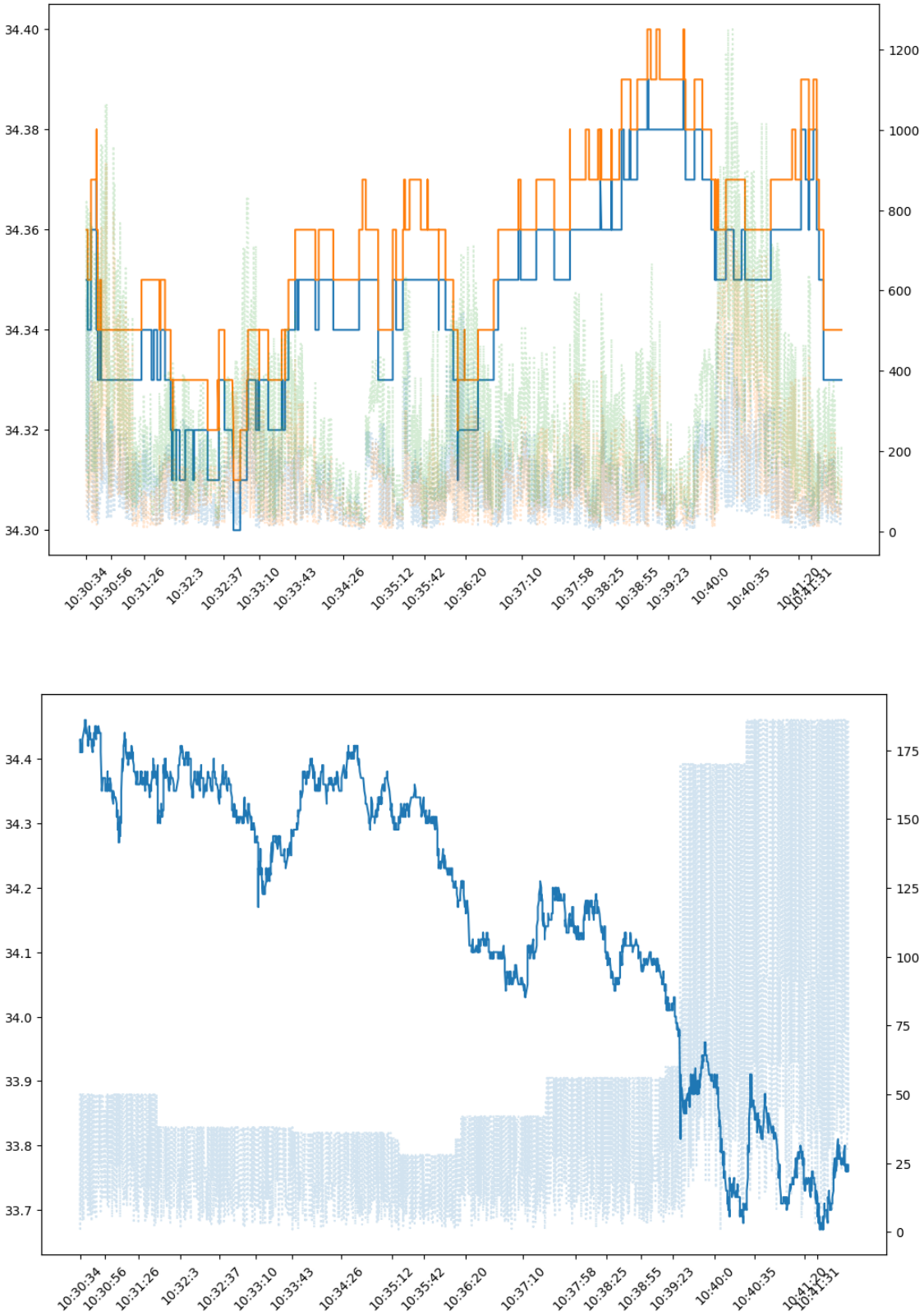


Figure 4.1: Episodes of QV (top panel) and PM (bottom panel) during liquidations of WTI Oil futures on the morning of 9th of April 2020.

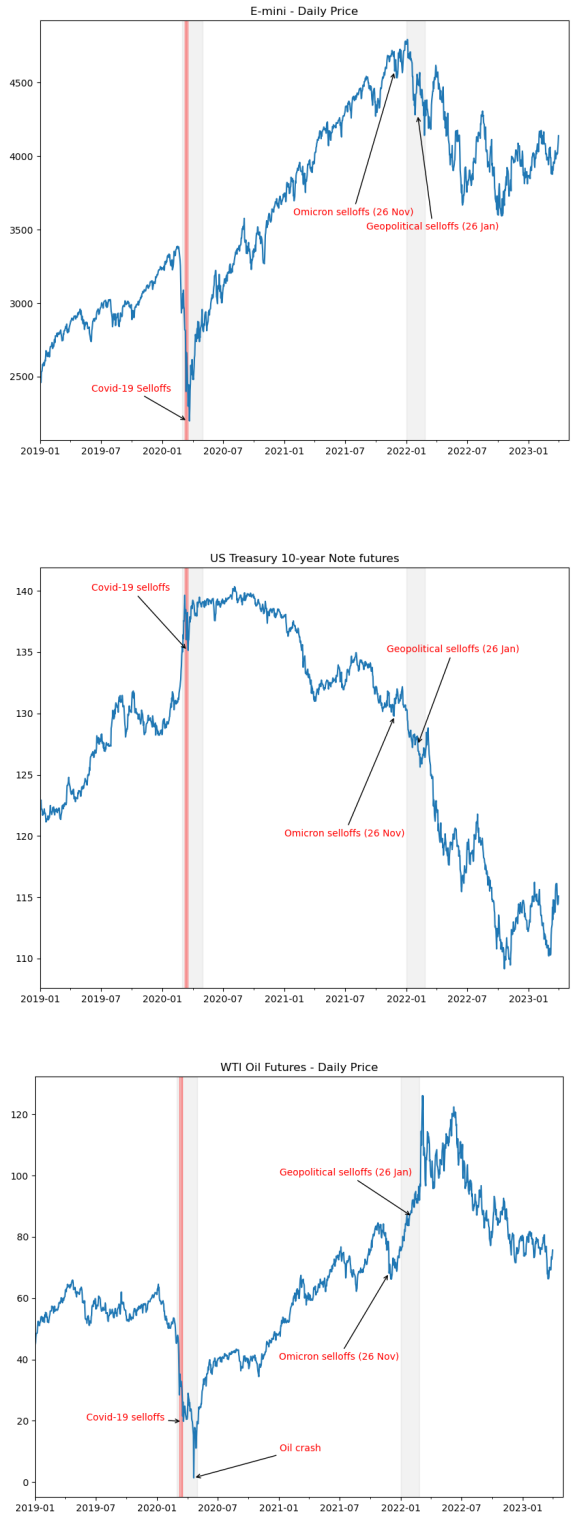


Figure 4.2: Daily transaction prices of CME Emini S&P 500, CME 10-year Note futures, and Light Crude Oil futures for March 2019 - April 2023. Price is the closing price of the day.

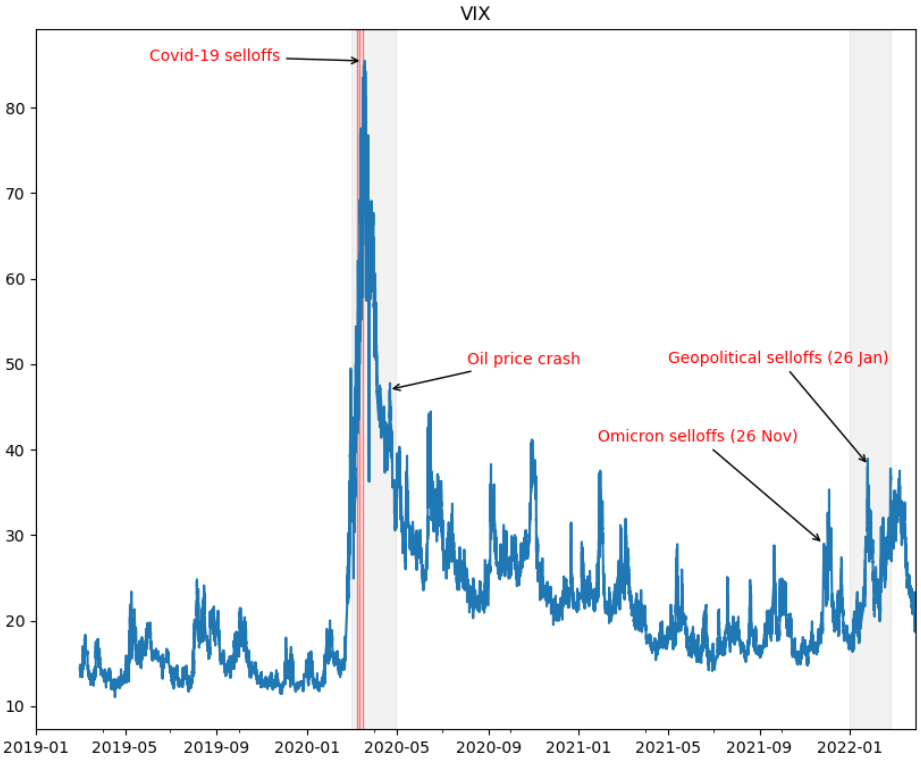


Figure 4.3: This figure presents VIX prices during March 2019 - April 2022.

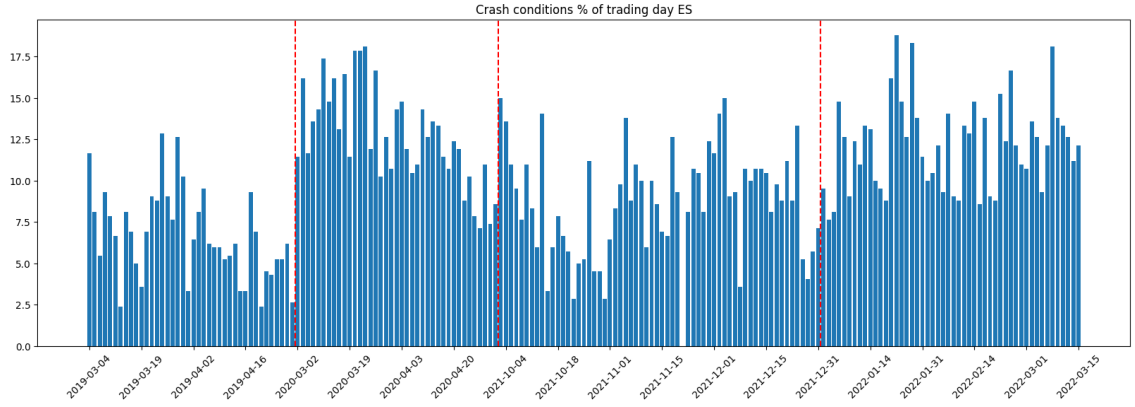


Figure 4.4: This Figure plots the percentage of crash episodes in the E-mini futures during Quiet periods from March-April 2019, Covid-19 period from March-April 2020 and geopolitical events during January 2022. We identify episodes of upcoming crashes when QV and PM are jointly at 90 percentile in rolling windows. There is an obvious increase in crises periods during Covid-19 sample till January 2022, as opposed to relatively quiet periods in 2019. The data is extracted from Refinitiv (TRTH).

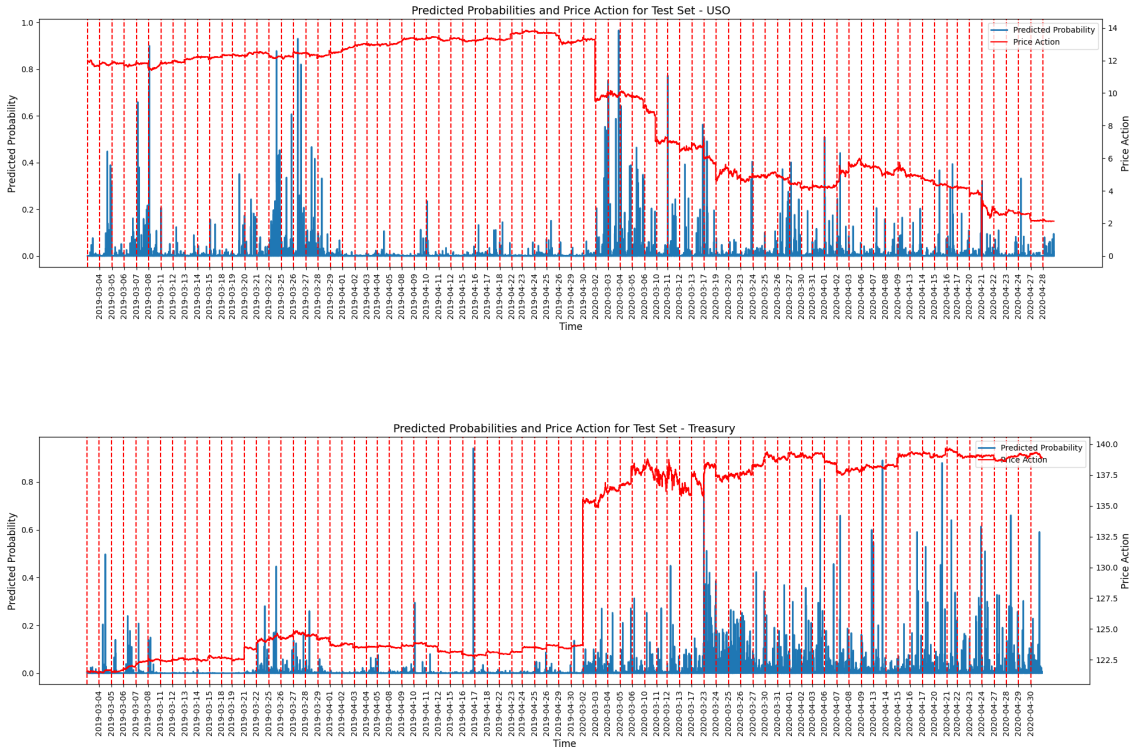


Figure 4.5: Cross-asset predictions of liquidity stress using QV and PM metrics in an XGBoost framework trained on E-mini futures. The model is tested on USO ETF Oil and the U.S. Treasury note futures during historical episodes of COVID-19 crisis, early January 2022 war tensions. Predicted probabilities of crisis (in blue) rise sharply prior to observed dislocations (in red), validating the model's capacity to anticipate cross-asset liquidity fragility.

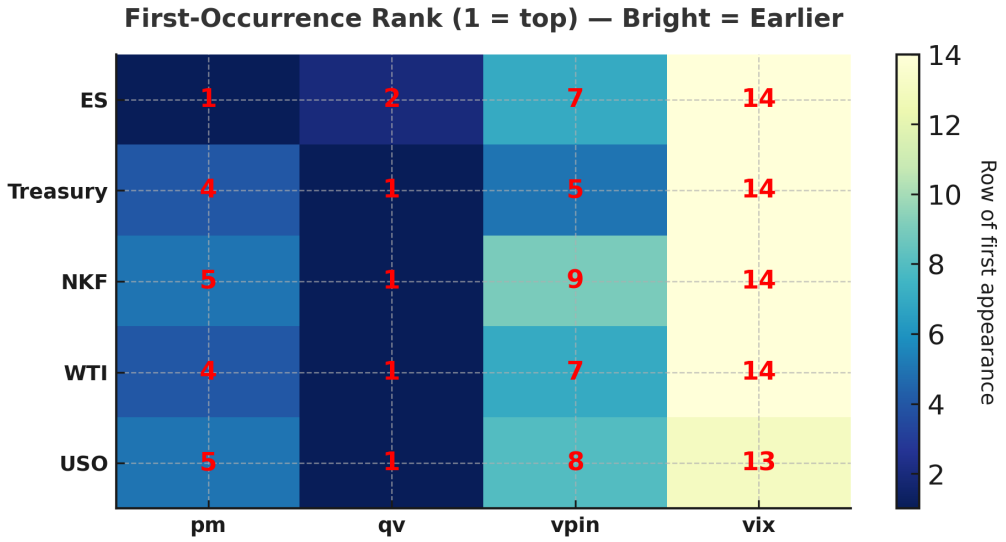


Figure 4.6: First-occurrence rank of signal families across assets. Each cell represents the earliest rank position (1 = earliest) at which a feature from the corresponding signal family appears in the feature set for a given asset. Brighter colors indicate earlier occurrence. qv and pm families consistently appear earlier across assets, while vix signals tend to emerge last, suggesting lower timeliness in predictive contexts.

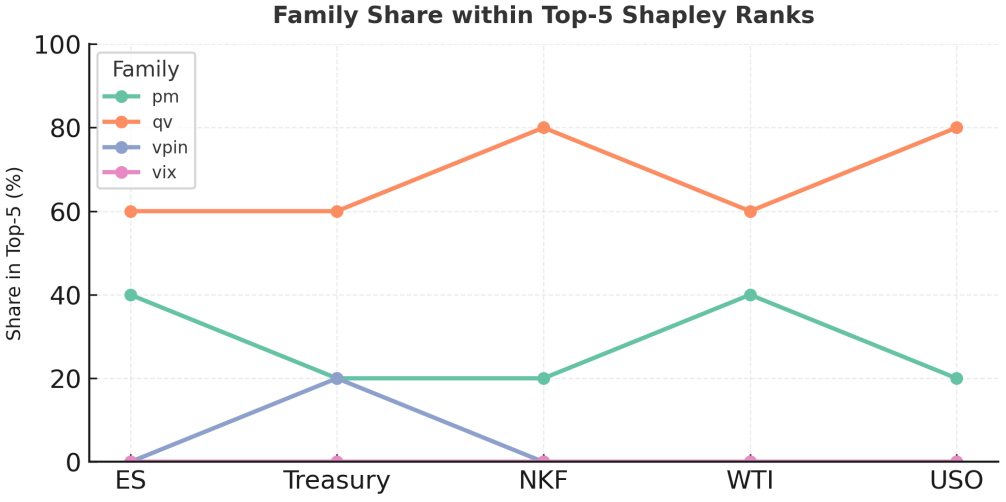


Figure 4.7: Share of each signal family among top-5 Shapley value ranks by asset. For each asset, the proportion of the top five most important features (by Shapley value) belonging to each signal family is shown. The qv family dominates across most assets, reflecting high marginal contribution to predictive performance. In contrast, vix signals are absent from the top ranks, reinforcing their limited explanatory power.

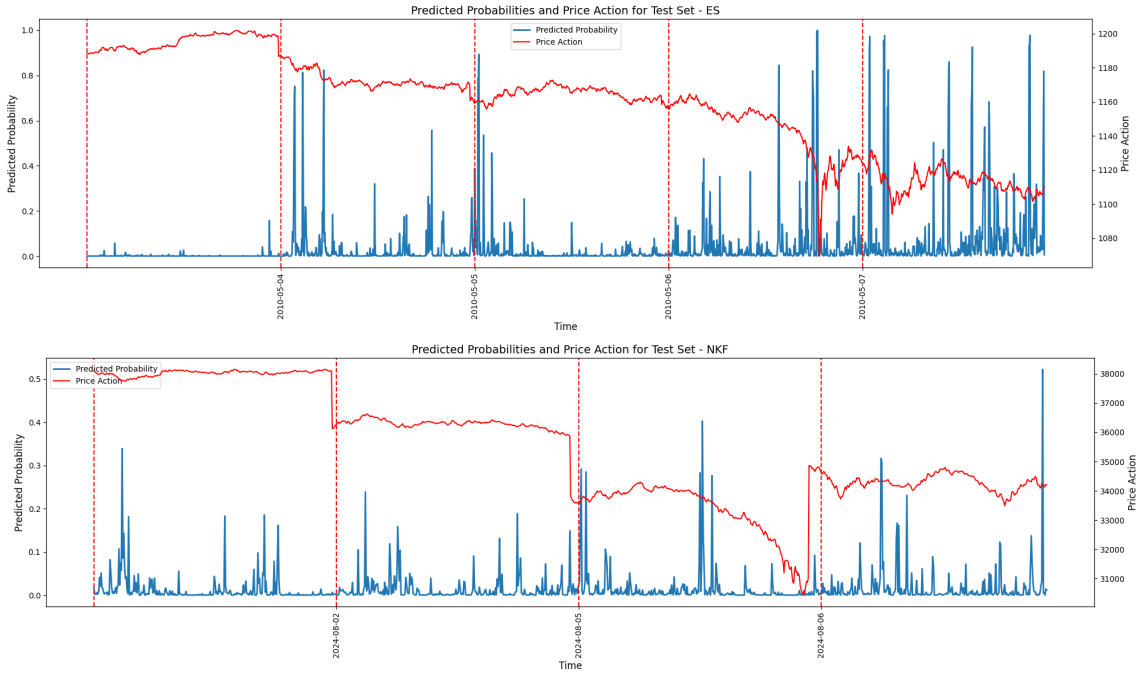


Figure 4.8: Out-of-sample predictions of liquidity stress using QV and PM metrics in an XGBoost framework trained on E-mini futures during the COVID-19 period. The top panel shows model predictions for E-mini S&P 500 futures around the time of the 2010 Flash Crash, while the bottom panel depicts predictions for Nikkei futures during the Japanese market dislocation in August 2024. Predicted crisis probabilities (blue) increase sharply ahead of actual price dislocations (red), demonstrating the model's ability to anticipate out-of-sample liquidity fragility.

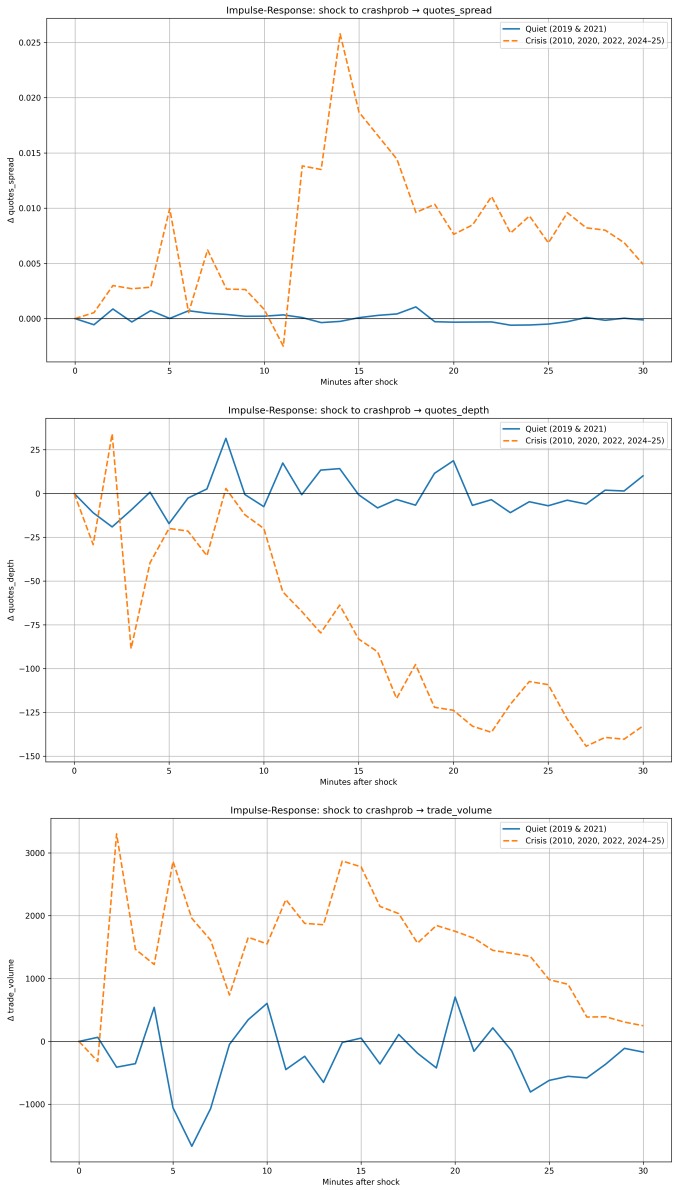


Figure 4.9: Impulse responses of market liquidity variables to a one-standard-deviation shock in predicted crash probability. Panels show responses for bid-ask spread (top), order book depth (middle), and trade volume (bottom). We divide the sample into crisis (red) and non-crisis (blue) regimes. During crisis regimes, crash risk leads to wider spreads, reduced depth, and increased volume, while quiet periods show relatively no responses.

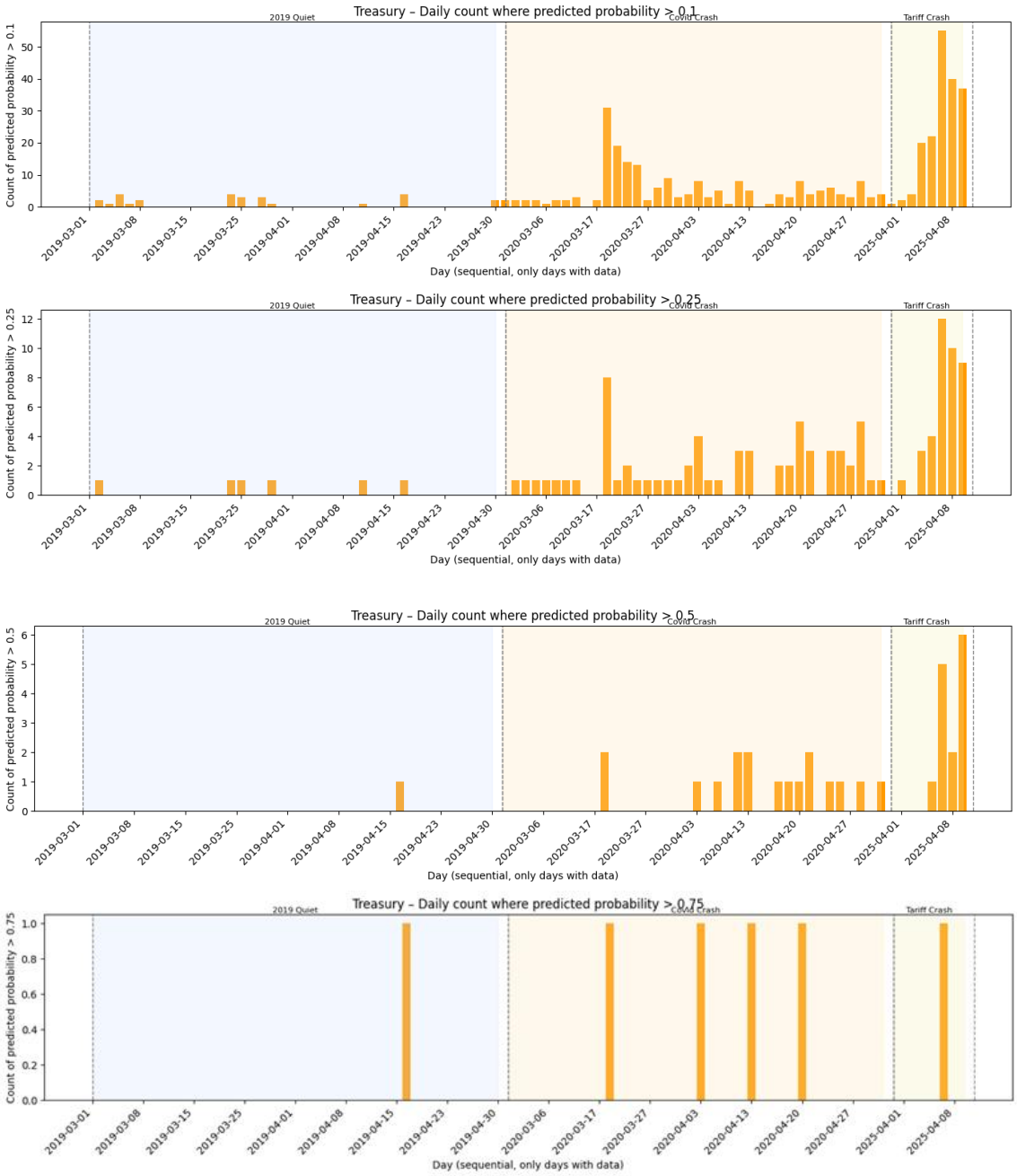


Figure 4.10: Crash Signal Intensity in the Treasury Market by Prediction Threshold. This panel chart visualizes the intensity of predicted crash signals in the U.S. Treasury market across different confidence levels. Each subplot corresponds to a specific threshold for the model's predicted probability of a crash—greater than 10%, 25%, 50% and 75%, respectively. For each day, we count the number of instances where individual predictions exceeded the given threshold. This allows us to track how widespread elevated crash risks were across the market on any given day. Distinct market regimes are highlighted in the background: a quiet pre-crisis period (light blue), the COVID crash (light orange), and the recent tariff-induced volatility (light yellow).

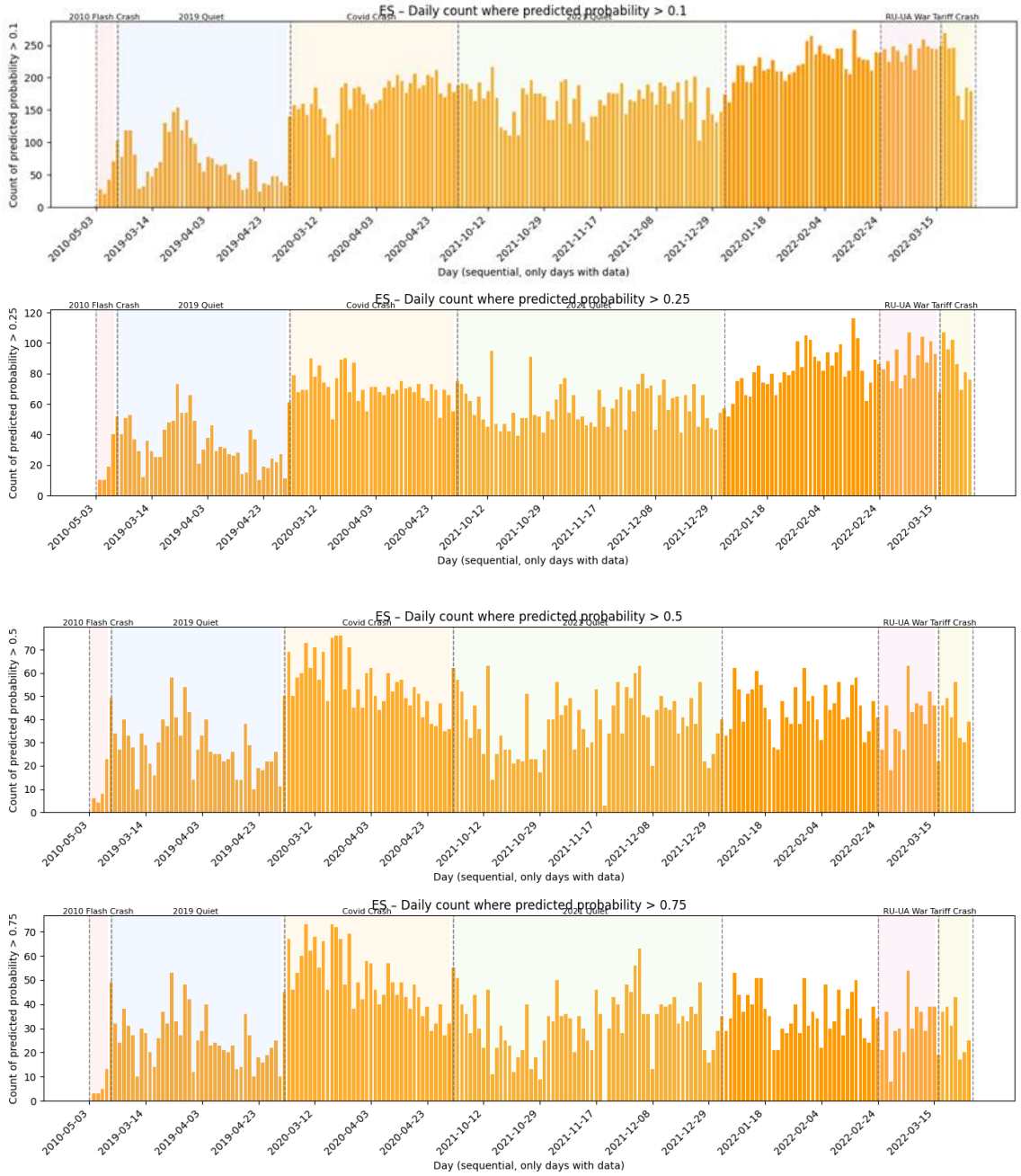


Figure 4.11: Crash Signal Intensity in the E-mini Market by Prediction Threshold. This panel chart visualizes the intensity of predicted crash signals in the E-mini market across different confidence levels. Each subplot corresponds to a specific threshold for the model's predicted probability of a crash—greater than 10%, 25%, 50% and 75%, respectively. For each day, we count the number of instances where individual predictions exceeded the given threshold. This allows us to track how widespread elevated crash risks were across the market on any given day. Distinct market regimes are highlighted in the background: 2010 flash crash (light pink), a quiet pre-crisis period in 2019 (light blue), the COVID crash (light orange) and post-Covid in 2021, the beginning of the geopolitical tensions in 2022 (pink), and the recent tariff-induced volatility (light yellow).

Table 4.1: Statistical Properties of Top-of-Book Depth and Trading Activity
The table shows one-minute median values calculated between the opening and closing times for each futures contract during each sample period. Top-of-book depth is the quantity of contracts available at the best ask and best bid. Volume is the number of contracts traded, and the number of trades. HL represents the difference between high and low prices. The sample periods include: the Covid-19 crisis (March 2, 2020 - April 30, 2020, including April 20, 2020, the oil price crash); the May 2010 Flash Crash; and quiet periods (March-April 2019). The three most liquid US futures contracts are analyzed: the E-mini S&P 500 (Panel A), US 10-year Note futures (Panel B), and WTI Crude Oil futures (Panel C).

	Top-of-Book Depth	Volume (Contracts)	Number of Trades	H-L
Panel A: E-mini S&P 500				
May 2010 Flash Crash	686	4,795	1,060	1.37
Covid (Mar-Apr 2020)	30	2,811	846	4.93
Quiet (Mar-Apr 2019)	193	1,591	224	1.18
Geopolitical (Jan-Feb 2022)	29	2,752	955	4.50
Panel B: US 10-year Note Futures				
Covid (Mar-Apr 2020)	629	1,128	65	0.02
Quiet (Mar-Apr 2019)	3,078	624	24	0.01
Panel C: WTI Crude Oil Futures				
Covid (Mar-Apr 2020)	18	725	339	0.12
Quiet (Mar-Apr 2019)	73	646	209	0.04

Table 4.2: Statistical Properties of QV and PM Ratios
The table shows the distribution of our ratios: Quote Volatility (QV) and Price Momentum (PM), calculated between opening and closing times for each futures contract across different sample periods.

	Quote Volatility (QV)						Price Momentum (PM)					
	Mean	Median	Stdv	P95	Max	Kurtosis	Mean	Median	Stdv	P95	Max	Kurtosis
Panel A: E-mini S&P 500												
May 2010 Flash Crash	50.68	28.40	78.68	161.02	1,404.00	81.64	4.01	3.00	4.28	10.00	45.00	19.07
Covid (Mar-Apr 2020)	170.35	76.00	347.32	914.00	9,932.00	103.03	4.57	2.33	6.91	15.00	205.00	58.27
Quiet (Mar-Apr 2019)	85.70	54.00	103.69	270.65	3,074.00	86.75	3.12	2.00	3.74	10.00	60.00	24.99
Geopolitical (Jan-Feb 2022)	251.99	123.89	633.73	825.02	12,387.00	95.79	5.54	2.55	9.65	20.00	165.00	49.95
Panel B: US 10-year Note Futures												
Covid (Mar-Apr 2020)	6,827.97	4,925.08	23,192.00	7,034.92	646,137.06	299.68	0.97	1.00	0.94	3.00	15.00	5.47
Quiet (Mar-Apr 2019)	1,609.58	1,312.05	1,869.48	2,853.29	25,356.25	80.43	0.73	1.00	0.72	3.00	9.00	5.88
Panel C: WTI Crude Oil Futures												
Covid (Mar-Apr 2020)	175.89	93.33	558.80	571.04	68,958.00	112.52	1.62	1.00	4.95	3.40	651.00	144.50
Quiet (Mar-Apr 2019)	130.72	88.00	126.49	382.00	1,834.00	12.15	1.46	1.00	1.50	3.00	55.00	195.73

.1 Appendix

The Appendix presents additional results. Most of the results here are discussed in the paper.

.1.1 Evidence of intraday momentum trading

Table IA1 - Intraday Momentum Trading Regressions - This table shows the pooled regression results of regressing the last half-hour return (rLH) on a constant and the first half hour return (rONFH), and the return until the last half hour (rROD) for S&P 500 Emini futures, US 10 Year Treasury Note futures, and WTI Oil futures. We extend results of Baltussen et al. (2021b) to a more recent sample from March 2019 to March 2023, including Covid-19 selloffs and other more recent turbulent times. Trading hours of Emini futures are based on the trading hours of their underlying markets, for the other two futures contracts trading hours are matched to their volume patterns. The intercept is not reported. T-statistics in parentheses are computed using standard errors that account for clustering on time and market (in case number of clusters exceeds ten), see Cameron et al. (2011). Significance at the 1%, 5%, and 10% level is denoted by , , or , respectively. Adjusted R2 and slope coefficients are multiplied by 100.

	rONFH	rROD	Adjusted R-square
Emini S&P500	0.031***	0.075***	0.111
	(0.006)	(0.007)	
US 10-year Note	-0.011*	0.070***	0.098
	(0.006)	(0.007)	
WTI Oil futures	-0.010	0.080***	0.069
	(0.007)	(0.01)	

Figure IA1 - Scatterplots of the rest of day 30 minutes returns (rROD) against the last 30 minutes returns (rLH). The rest of day 30 minutes returns are shown on the horizontal axis, and last 30 minutes returns on the vertical axis, are from the S&P 500 Emini futures market (Panel A), the 10-year US Treasury Futures (Panel B), and the Light Crude Oil futures NYMEX (Panel C). The sample period is from March 2020 until April 2023, including the period of Coronavirus. Similar to Baltussen et al. (2021b), we consider trading hours of the E-mini S&P500 futures based on the underlying trading hours, i.e. 9:30-16:00; US 10-year T-Note futures trading hours are 8:20-15:00; Light Crude Oil NYMEX futures' trading hours are 9:00-14:30. All hours are expressed in Eastern Standard Time (EST).

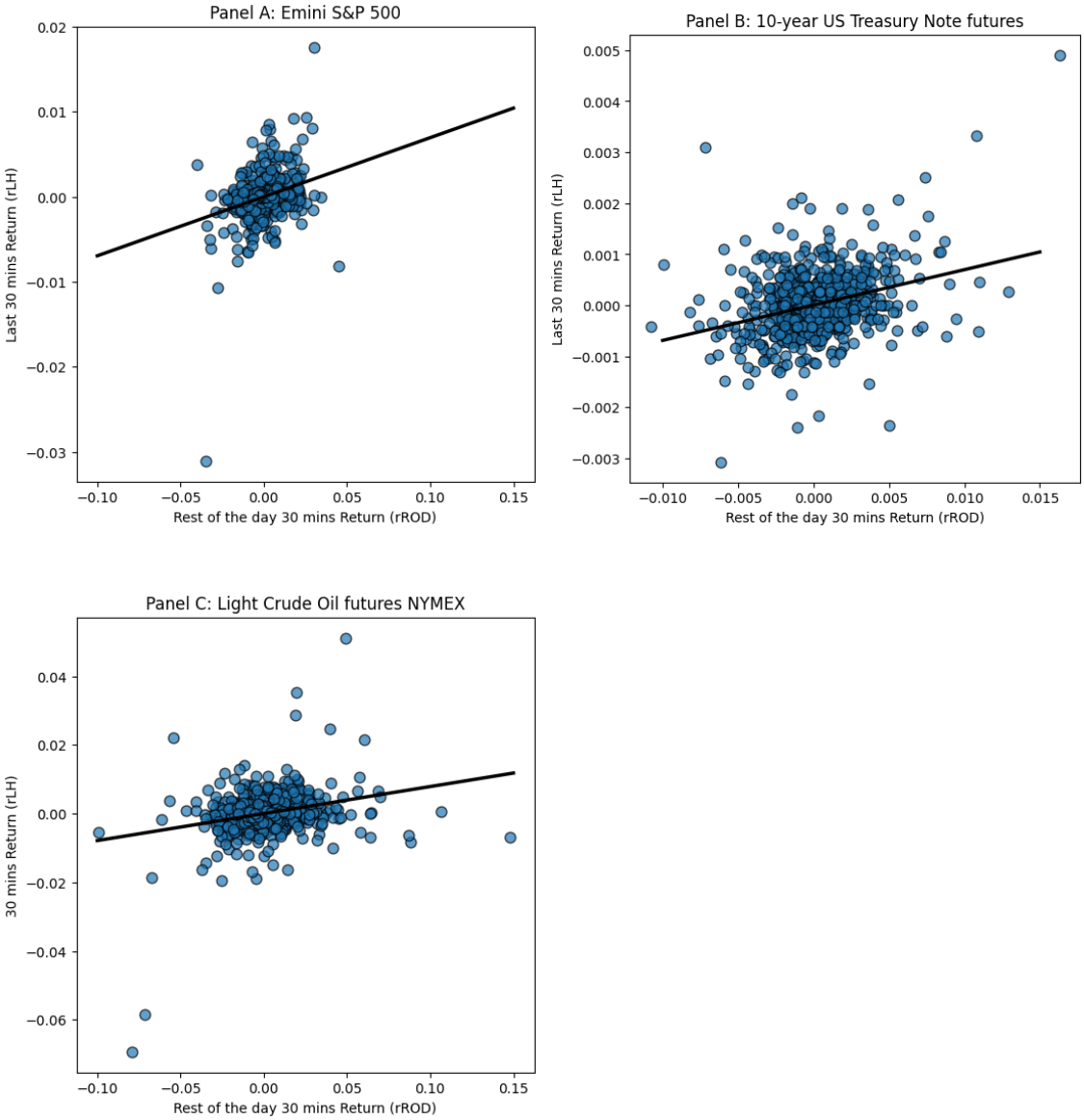
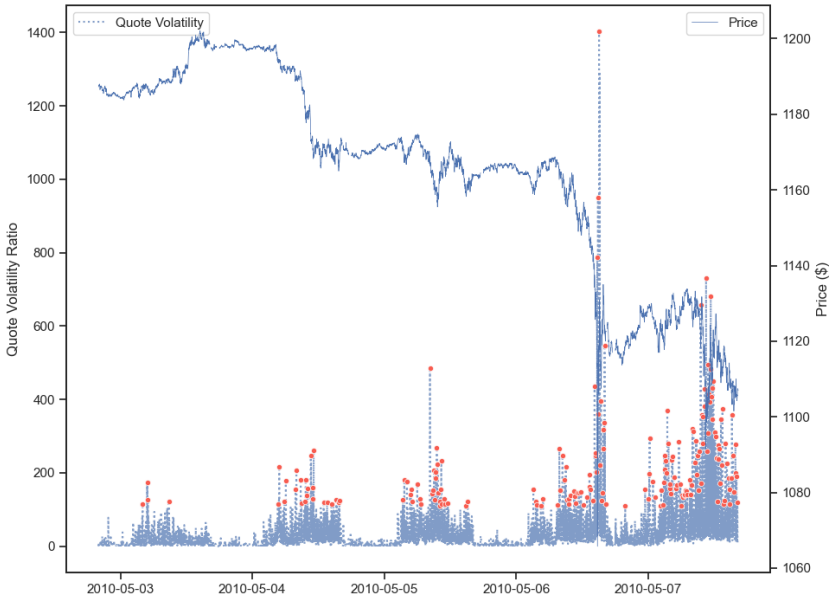
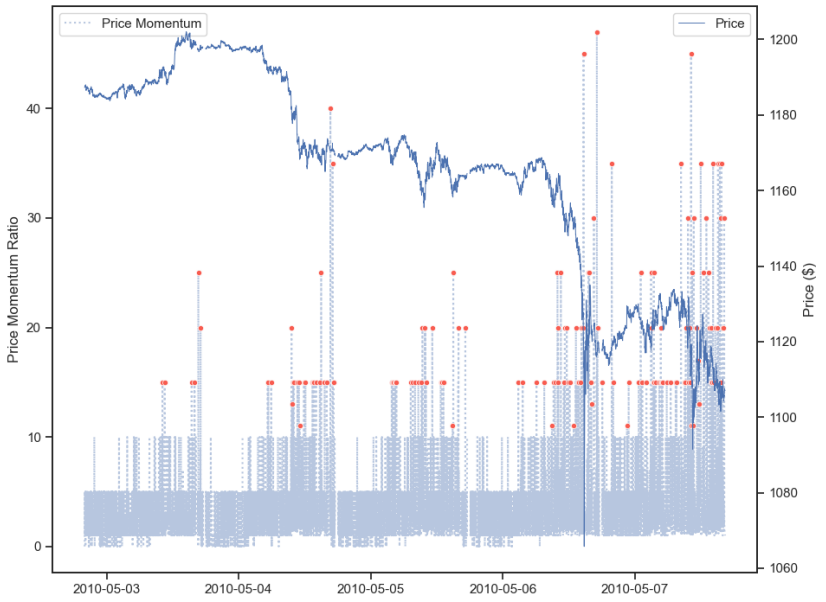


Figure IA2 - The Figure depicts the time series of the best ask and the best bid quotes alongside the Inask and Inbid QV ratios on CME of the E-mini futures (Top) before, during and after the May 6, 2010 flash crash; and the corresponding Trading price and the PM ratio (Bottom). The data is extracted from Refinitiv. CFTC and SECs' reports conclude that the afternoon 36-minutes price crash was initiated by a single algorithmic order that executed a large sale of 75,000 E-mini contracts in a very short period of time from fund management Waddell and Reed. They conclude that HFTs exacerbate the downturns. Menkveld and Yueshen (2019) suggest that losses occurred in the first five minutes of the crash.





Bibliography

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The review of financial studies*, 30(1):2–47.

Allen, L., Bali, T. G., and Tang, Y. (2012). Does systemic risk in the financial sector predict future economic downturns? *The Review of Financial Studies*, 25(10):3000–3036.

Amihud, Y. (2002a). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*.

Amihud, Y. (2002b). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Amihud, Y. and Mendelson, H. (1980). Dealership market: Market-making with inventory. *Journal of financial economics*, 8(1):31–53.

Andersen, T. G. and Bondarenko, O. (2014). Reflecting on the vpin dispute. *Journal of Financial Markets*, 17:53–64.

Andersen, T. G. and Bondarenko, O. (2015). Assessing measures of order flow toxicity and early warning signals for market turbulence. *Review of Finance*, 19(1):1–54.

Augustin, P., Brenner, M., and Subrahmanyam, M. G. (2019). Informed options trading prior to takeover announcements: Insider trading? *Management Science*, 65(12):5697–5720.

Aune, R. T., Krellenstein, A., O'Hara, M., and Slama, O. (2017). Footprints on a blockchain: Trading and information leakage in distributed ledgers. *The Journal of Trading*, 12(3):5–13.

Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, 27(2):12–23.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *ICLR 2015*.

Baltussen, G., Da, Z., Lammers, S., and Martens, M. (2021a). Hedging demand and market intraday momentum. *Journal of Financial Economics*.

Baltussen, G., Da, Z., Lammers, S., and Martens, M. (2021b). Hedging demand and market intraday momentum. *Journal of Financial Economics*, 142(1):377–403.

Baradehi, Y., Bernhardt, D., Ruchti, T. G., and Weidenmier, M. (2019). The night and day of amihuds 2002 liquidity measure. *The Review of Asset Pricing Studies*.

Barbon, A. and Buraschi, A. (2020). Gamma fragility. *University of St. Gallen*.

Barbon, A. and Ranaldo, A. (2021). On the quality of cryptocurrency markets: Centralized versus decentralized exchanges. *arXiv preprint arXiv:2112.07386*.

Barclay, M. and Hendershott, T. (2003). Price discovery and trading after hours. *The Review of Financial Studies*.

Beckmeyer, H. and Moerke, M. (2021). End-of-day momentum in the cross-section and option hedging. *European Financial Management Association*.

Bennel, J. and Sutcliffe, C. (2004). Black-scholes versus artificial neural networks in pricing ftse 100 options. *Intelligent systems in accounting, finance and management*.

Bessembinder, H., Carrion, A., Tuttle, L., and Venkataraman, K. (2016). Liquidity, resiliency and market quality around predictable trades: Theory and evidence. *Journal of Financial economics*, 121(1):142–166.

Biais, B., Capponi, A., Cong, L. W., Gaur, V., and Giesecke, K. (2023a). Advances in blockchain and crypto economics. *Management Science*, 69(11):6417–6426.

Biais, B., Declerck, F., and Moinas, S. (2016). Who supplies liquidity, how and when. *BIS Working Paper*.

Biais, B., Donier, J., and Riedel, S. (2023b). The microstructure of decentralised exchanges. *Finance Working Paper*.

Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial economics*, 116(2):292–313.

Board of Governors of the Federal Reserve System (2022). The collapse of terra-luna and implications for crypto-asset liquidity. Technical report, Federal Reserve Staff Notes.

Bogoev, D. and Karam, A. (2017). Detection of algorithmic trading. *Physica A: Statistical Mechanics and its applications*.

Bogousslavsky, V. and Collin-Dufresne, P. (2022). Liquidity, volume, and order imbalance volatility. *The Journal of Finance*.

Bogousslavsky, V. and Collin-Dufresne, P. (2023). Liquidity, volume, and order imbalance volatility. *The Journal of Finance*, 78(4):2189–2232.

Bogousslavsky, V., Fos, V., and Muravyev, D. (2024). Informed trading intensity. *The Journal of Finance*, 79(2):903–948.

Bollen, N. and Whaley, R. (2004). Does net buying pressure affect the shape of the implied volatility functions. *The Journal of Finance*.

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., and Sokolov, K. (2018). High frequency trading and extreme price movements. *Journal of Financial Economics*, 128(2):253–265.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*.

Brunnermeier, M. K. and Pedersen, L. H. (2005). Predatory trading. *The Journal of Finance*, 60(4):1825–1863.

Brunnermeier, M. K. and Pedersen, L. H. (2009a). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238.

Brunnermeier, M. K. and Pedersen, L. H. (2009b). Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238.

Cao and Wei (2010). Commonality in liquidity: evidence from the option market. *Journal of Financial Markets*.

Cao, S., Jiang, W., Wang, J., and Yang, B. (2024). From man vs. machine to man+ machine: The art and ai of stock analyses. *Journal of Financial Economics*, 160:103910.

Capponi, A., Jia, R., and Yu, S. (2024). Price discovery on decentralized exchanges. *Available at SSRN 4236993*.

Cespa, G. and Vives, X. (2015). The beauty contest and short-term trading. *The Journal of Finance*, 70(5):2099–2154.

Chen, H., Petukhov, A., Wang, J., and Xing, H. (2024). The dark side of circuit breakers. *The Journal of Finance*, 79(2):1405–1455.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *KDD '16*.

Chernobai, A., Longstaff, F., and Pan, J. (2016). Liquidity and credit risk in the cds market. *Journal of Financial Economics*.

Chou, R. K., Chung, S.-L., Hsiao, Y.-J., and Wang, Y.-H. (2011). The impact of liquidity on option prices. *The Journal of Futures markets*.

Christoffersen, P., Goyenko, R., Jacobs, K., and Karoui, M. (2018). Illiquidity premia in the equity options market. *The Review of Financial Studies*.

Colliard, J.-E., Foucault, T., and Lovo, S. (2022). Algorithmic pricing and liquidity in securities markets. *HEC Paris Research Paper No. FIN-2022-1459*.

Collin-Dufresne, P. and Fos, V. (2015). Do prices reveal the presence of informed trading? *The Journal of Finance*, 70(4):1555–1582.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., Li, Y., and Wang, N. (2021a). Tokenomics: Dynamic adoption and valuation. *Management Science*, 67(9):5087–5119.

Cong, L. W., Li, Y., and Wang, N. (2021b). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3):1105–1155.

Conrad, J., Wahal, S., and Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, 116(2):271–291.

Demsetz, H. (1968). The cost of transacting. *Quarterly Journal of Economics*, 82(1):33–53.

Deuskar, P., Gupta, A., and Subrahmanyam, M. G. (2004). Interest rate option markets: The role of liquidity in volatility smiles. *New York University working paper*.

Dou, W. W., Goldstein, I., and Ji, Y. (2024). Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Du, Q., Yin, F., and Li, Z. (2019). Base station traffic prediction using xgboost-lstm with feature enhancement. *The Institution of Engineering and Technology*.

Duffie, D. (2010). Asset pricing with heterogeneous agents. *Handbook of Financial Econometrics*, 1:639–678.

Easley, D., López de Prado, M., O'Hara, M., and Zhang, Z. (2021). Microstructure in the machine age. *The Review of Financial Studies*, 34(7):3316–3363.

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

Easley, D., O'Hara, M., and Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 134(1):91–109.

Financial Stability Board (2024). Lessons from the march 2020 market turmoil: Enhancing liquidity monitoring. Technical report, Financial Stability Board.

Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: A comparative approach. *Journal of Financial Stability*, 4(2):149–164.

Flashbots (2022). Mev-boost: Separation of block builders and validators. Technical Documentation. Available at https://boost.flashbots.net.

Fong, K. Y. L., Holden, C. W., and Trzcinka, C. A. (2017). What are the best liquidity proxies for global research? *Review of Finance*.

Foster, F. D. and Viswanathan, S. (1996). Strategic trading when agents forecast the forecasts of others. *The Journal of Finance*, 51(4):1437–1478.

Foucault, T., Kadan, O., and Kandel, E. (2013). Liquidity cycles and make/take fees in electronic markets. *The Journal of Finance*, 68(1):299–341.

Foucault, T., Kozhan, R., and Tham, W. W. (2017). Toxic arbitrage. *The Review of Financial Studies*.

Foundation, E. (2022). The merge. Online Documentation. Available at https://ethereum.org/en/upgrades/merge/.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2006a). Institutional investors and stock market volatility. *The Quarterly Journal of Economics*.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2006b). Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2):461–504.

Garleanu, N., Pedersen, L. H., and Poteshman, A. (2009). Demand-based option pricing. *The Review of Financial Studies*.

Glosten, L. and Milgrom, P. (1985a). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*.

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

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

Goldstein, I., Spatt, C. S., and Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, 34(7):3213–3225.

Goyal and Saretto (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*.

Goyenko, R., Holen, C., and Trzcinka, C. A. (2009). Do liquidity measures measure liquidity. *Journal of Financial Economics*.

Gromb, D. and Vayanos, D. (2010). Limits of arbitrage. *Annu. Rev. Financ. Econ.*, 2(1):251–275.

Gu, S., Kelly, B., and Xiu, D. (2019). Empirical asset pricing via machine learning. *The Review of Financial Studies*.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Guéant, O. and Manziuk, I. (2019). Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality. *Applied Mathematical Finance*, 26(5):387–452.

Haddad, V., Moreira, A., and Muir, T. (2021a). When sellers are few: The effect of large owners on market liquidity. *Journal of Financial Economics*, 139(2):355–384.

Haddad, V., Moreira, A., and Muir, T. (2021b). When selling becomes viral: Disruptions in debt markets in the covid-19 crisis and the fed's response. *The Review of Financial Studies*, 34(11):5309–5351.

Hagstromer, B. and Norden, L. (2018). The diversity of high-frequency traders. *Journal of Financial Markets*.

Harvard Law School Corporate Governance Forum (2023). Crypto runs and market freezes: Policy lessons from 2022. Technical report, Harvard Law School.

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

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The journal of Finance*, 50(4):1175–1199.

Hasbrouck, J. (2018). High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis*, 53(2):613–641.

Hasbrouck, J., Rivera, T. J., and Saleh, F. (2022). The need for fees at a dex: How increases in fees can increase dex trading volume. *Available at SSRN 4192925*.

Hendershott, T., Livdan, D., and Schurhoff, N. (2015). Are institutions informed about news. *Journal of Financial Economics*.

Hendershott, T. and Menkveld, A. J. (2014). Price pressures. *Journal of Financial economics*, 114(3):405–423.

Ho, T. and Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47–73.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*.

Holden, C. W. and Subrahmanyam, A. (1992). Long-lived private information and imperfect competition. *The Journal of Finance*, 47(1):247–270.

Hu, C. and Yuan, K. (2023). Crypto liquidity and market microstructure. *Journal of Financial Economics*. forthcoming.

Hu, J. (2013). Does option trading convey stock price information? *Journal of Financial Economics*.

Huang, R. and Stoll, H. (2015). The components of the bid-ask spread: A general approach. *The Review of Financial Studies*.

Hull, J. (2003). *Options, futures, and other derivatives*. Prentice Hall International, London.

Ivascu, C.-F. (2021). Option pricing using machine learning. *Expert Systems with Applications*.

Jabeur, S. B., Stef, N., and Carmona, P. (2023). Bankruptcy prediction using the xgboost algorithm and variable importance feature engineering. *Computational Economics*.

Kacperczyk, M. and Pagnotta, E. S. (2019). Chasing private information. *The Review of Financial Studies*, 32(12):4997–5047.

Karam, A. and Bogoev, D. (2023). Intraday momentum trading and liquidity crises. *upcoming*.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *The Review of Financial Studies*, 27(10):2841–2871.

Kirilenko, A., Kyle, A. S., Samadi, M., and Tuzun, T. (2017a). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998.

Kirilenko, A., Kyle, A. S., Samadi, M., and Tuzun, T. (2017b). The flash crash: The impact of high-frequency trading on an electronic market. *Journal of Finance*, 72(3):967–998.

Kondor, P. and Vayanos, D. (2019). Liquidity risk and the dynamics of arbitrage capital. *The Journal of Finance*, 74(3):1139–1173.

Kwan, A., Philip, R., and Shkilko, A. (2021). The conduits of price discovery: A machine learning approach.

Kyle, A. S. (1985a). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Kyle, A. S. (1985b). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Kyle, A. S. and Obizhaeva, A. A. (2023). Large bets and stock market crashes. *Review of Finance*, 27(6):2163–2203.

Lehar, A. and Parlour, C. (2021). Decentralized exchange: The uniswap automated market maker. *The Journal of Finance*.

Lehar, A. and Parlour, C. A. (2023). Battle of the bots: Flash loans, miner extractable value and efficient settlement. *Miner Extractable Value and Efficient Settlement (March 8, 2023)*.

Lehar, A., Parlour, C. A., and Zoican, M. (2024). Fragmentation and optimal liquidity supply on decentralized exchanges. *Available at SSRN 4267429*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.

Malinova, K., Park, A., and Riordan, R. (2018). Do retail investors suffer from high frequency traders? *Journal of Financial and Quantitative Analysis*.

Malliaris, M. and Salchenberger, L. (1993). Beating the best: A neural network challenges the black-scholes formula. *IEEE*.

Mayhew, S. (2002). Competition, market structure, and bid-ask spreads in stock option markets. *The Journal of Finance*.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of financial Markets*, 16(4):712–740.

Menkveld, A. J. and Yueshen, B. Z. (2019). The flash crash: A cautionary tale about highly fragmented markets. *Management Science*, 65(10):4470–4488.

Moyaert, T. (2013). The information content of volume price impact for intraday liquidity forecasting. *26th Australasian Finance and Banking Conference*.

Muravyev, D. (2016). Order flow and expected option returns. *Journal of Finance*.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

Nagel, S. (2021). *Machine learning in asset pricing*. Princeton University Press.

Ni, S. X., Pearson, N. D., Poteshman, A. M., and White, J. (2021). Does option trading have a pervasive impact on underlying stock prices? *Review of Financial Studies*.

Ondieki, K. M. (2022). Swaption pricing under libor market model using monte-carlo method with simulated annealing optimization. *Journal of Mathematical Finance*.

Park, A. (2023). The conceptual flaws of decentralized automated market making. *Management Science*, 69(11):6731–6751.

Park, J. and Hwang, E. (2021). A two-stage multistep-ahead electricity load forecasting scheme based on lightgbm and attention-bilstm. *Sensors*.

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

Parlour, C. A. and Lehar, A. (2024). Fragmentation of liquidity in decentralised exchanges. *Journal of Finance*. forthcoming.

Pastor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Saleh, F. (2021). Blockchain without waste: Proof-of-stake. *The Review of financial studies*, 34(3):1156–1190.

Schrimpf, A., Shin, H. S., and Sushko, V. (2020). Leverage and margin spirals in fixed income markets during the covid-19 crisis. *Available at SSRN 3761873*.

Semmelmann, L., Henni, S., and Weinhardt, C. (2022). Load forecasting for energy communities: a novel lstm-xgboost hybrid model based on smart meter data. *Energy Informatics*.

Soebhag, A. (2022). Option gamma and stock returns. *SSRN*.

Strumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*.

Sun, Y. and Tian, L. (2022). Research on stock prediction based on simulated annealing algorithm and ensemble neural learning. *International Conference on Computer Science and Communication Technology*.

Yates, C. (2021). The tail is wagging the stock market dog. *Web Article*.