Prague University of Economics and Business
Faculty of Informatics and Statistics

Implementation of a trading algorithm with a focus on volume profiles

BACHELOR'S THESIS

Study program: Data Analytics

Author: Vít Kohout
Supervisor: Ing. Václav Zeman, Ph.D.

Prague, June 2025

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Ing. Václav Zeman, Ph.D., for his guidanceand feedback throughout this research.

My sincere thanks are also extended to prof. Ing. Tomáš Kliegr, Ph.D., and the Department of Information and Knowledge Engineering (KIZI) for providing access to server and GPU resources that were helpful for the experiments and simulations conducted in this thesis.

Abstract

Thesis investigates the incremental value of integrating advanced analytics into classical intraday trading strategies. To systematically assess the impact of increasing analytical complexity, a tiered framework of nine strategies was developed spanning baseline (price-only), volume-enhanced, and deep learning-enhanced archetypes. Temporal Fusion Transformer model is implemented to provide 15-minute ahead volume forecasts for the most advanced strategies and everything is rigorously backtested on NASDAQ-100 securities. The empirical results demonstrate a key trade-off: while incorporating volume-based parameters did not significantly increase profitability, it offered a powerful, statistically significant advantage in risk management by consistently reducing maximum drawdowns. Conversely, the hypothesis that a sophisticated transformer-based forecast would further improve performance was rejected, as the strategies operating on predicted volume failed to outperform their simpler, heuristic-based counterparts. Analysis also validates that strategy performance is highly regime-dependent, particularly for mean-reversion and momentum approaches. It is concluded that the primary contribution of adding volume-based complexity in this context is defensive, serving to preserve capital rather than amplify return, and highlight a critical trade-off between the benefits and implementation costs of deploying advanced predictive models in intraday trading strategies on higher frequency data.

Keywords

algorithmic trading, quantitative finance, intraday trading strategies, volume forecasting, time series analysis, transformer models, machine learning, deep learning, market risk management

Contents

Introduction

0.1 Motivation

This research is, at its core, a personal exploration, born from my observations concerning the transformation of financial markets across my lifetime. Since the early 2000s, trading has evolved from a mainly discretionary craft, where intuition deciphered price charts and market rhythms, to a domain ruled by algorithms, where precision and automation reign supreme. The 2000s marked a turning point as electronic platforms surged, shifting reliance from manual decision-making to systematic strategies [1]. By the 2010s, traditional indicators, think moving averages or RSI, gave way to machine learning tools, a trend that has gotten even more momentum in the 2020s by the AI boom, exemplified by OpenAI's GPT-3 and the rise of transformer models for time-series analysis [13, 44].

This evolution isn't just technological, it's a competitive imperative, an arms-race. Staying ahead demands fresh approaches, repurposing cross-disciplinary methods, dusting off overlooked frameworks, or blending the two into something new. What has driven me to start this research was the stark contrast between yesterday's discretionary traders and today's algorithms. Back then, traders set thresholds and stop-losses partly from instinct, with their own subconscious pattern recognition abilities, often at levels more art than rigorous scientific method [89]. Discretionary traders get paralyzed by fear during crashes, overtrade on euphoria, or miss exits due to indecision. Yet, those very flaws make them and made the markets in the past more dynamic. Algorithms, by contrast, don't flinch, they execute with cold consistency, following predefined logic, a strength, but also a vulnerability. Predictability in complex systems invites exploitation. High-frequency traders, as O'Hara [82] illustrates, often converge on familiar signal zones, price levels like bids and asks, or bursts of volume, despite the transformation of order books, high-speed data feeds, and fragmented liquidity pools. This convergence, even amidst technological upheaval, reveals an opportunity. That's why I've chosen to revive a piece of the Market Profile framework, an old-school approach to mapping market structure, and pair it with modern confirmation methods and predictive tools, blending the adaptability of the past with the precision of the present.

0.2 Research Objectives

This study investigates the contemporary intraday trading landscape through three core objectives. First, it maps the current scope of intraday trading strategies, leveraging theoretical foundations that encompass short-term tactics, statistical methods, and advanced techniques from machine learning and deep learning, as outlined in the selected framework. Second, it develops and refines trading strategies tailored to intraday NASDAQ data, integrating volume-based enhancements and transformer-driven forecasts to optimize performance. Third, it assesses the efficacy of these strategies through rigorous backtesting, benchmarking, and possibly live incubation, evaluating their performance across diverse market conditions.

0.3 Research Hypotheses

The research is guided by three hypotheses that align with the proposed methods and reflect the interplay of market regimes, trading volumes, and advanced modeling techniques:

1. **Regime-Specific Performance Hypothesis:** Intraday trading strategies exhibit distinct performance profiles across market regimes.

H1a: Breakout strategies perform better in medium/high volatility than in low volatility.

$$H_0 : \mu_{SR}^{\text{breakout, highVol}} \leq \mu_{SR}^{\text{breakout, lowVol}}$$
$$H_1 : \mu_{SR}^{\text{breakout, highVol}} > \mu_{SR}^{\text{breakout, lowVol}}$$

H1b: Breakout strategies perform better in trending vs ranging markets.

$$H_0 : \mu_{SR}^{\text{breakout, trending}} \leq \mu_{SR}^{\text{breakout, range}}$$
$$H_1 : \mu_{SR}^{\text{breakout, trending}} > \mu_{SR}^{\text{breakout, range}}$$

H1c: Mean-reversion strategies perform better in low/medium volatility than in high volatility.

$$H_0 : \mu_{SR}^{\text{meanRev, lowVol}} \leq \mu_{SR}^{\text{meanRev, highVol}}$$
$$H_1 : \mu_{SR}^{\text{meanRev, lowVol}} > \mu_{SR}^{\text{meanRev, highVol}}$$

H1d: Mean-reversion strategies perform worse in uptrends than in ranging markets.

$$H_0 : \mu_{SR}^{\text{meanRev, uptrend}} \geq \mu_{SR}^{\text{meanRev, range}}$$
$$H_1 : \mu_{SR}^{\text{meanRev, uptrend}} < \mu_{SR}^{\text{meanRev, range}}$$

H1e: Momentum strategies perform better in trending/high-volatility markets than in low-volatility/ranging markets.

$$H_0 : \mu_{SR}^{\text{momentum, trendVol}} \leq \mu_{SR}^{\text{momentum, rangeLowVol}}$$
$$H_1 : \mu_{SR}^{\text{momentum, trendVol}} > \mu_{SR}^{\text{momentum, rangeLowVol}}$$

2. **Volume as a Confirmation Signal Hypothesis:** Incorporating volume-based parameters, such as Value Area breakouts, VWAP deviations, or relative volume surges improves the precision and profitability of intraday strategies compared to price-only baselines.

$$H_0 : \mu_{SR}^{\text{volume}} \leq \mu_{SR}^{\text{price}} \wedge \mu_{PF}^{\text{volume}} \leq \mu_{PF}^{\text{price}} \wedge \mu_{MDD}^{\text{volume}} \geq \mu_{MDD}^{\text{price}}$$
$$H_1 : \mu_{SR}^{\text{volume}} > \mu_{SR}^{\text{price}} \wedge \mu_{PF}^{\text{volume}} > \mu_{PF}^{\text{price}} \wedge \mu_{MDD}^{\text{volume}} < \mu_{MDD}^{\text{price}}$$

3. **Transformer-Based Volume Prediction Hypothesis:** Augmenting strategies with transformer-based forecasted short-term volume predictions, improves signal generation framework, achieving better metrics.

$$H_0 : \mu_{SR}^{\text{transformer}} \leq \mu_{SR}^{\text{volume}} \wedge \mu_{PF}^{\text{transformer}} \leq \mu_{PF}^{\text{volume}} \wedge \mu_{MDD}^{\text{transformer}} \geq \mu_{MDD}^{\text{volume}}$$
$$H_1 : \mu_{SR}^{\text{transformer}} > \mu_{SR}^{\text{volume}} \wedge \mu_{PF}^{\text{transformer}} > \mu_{PF}^{\text{volume}} \wedge \mu_{MDD}^{\text{transformer}} < \mu_{MDD}^{\text{volume}}$$

1. Related Works

1.1 Theoretical Foundations of Trading Strategies

1.1.1 Short-term and Intraday Trading Strategies

Short-term and intraday trading strategies exploit market inefficiencies within narrow windows, from microseconds to a single day, a practice tracing back to Charles Dow's emphasis on price patterns and volume as market signals [79]. Electronic trading platforms and high-frequency data have since refined these methods, enabling traders to capture rapid price shifts with precision [60, 76]. Unlike long-term strategies, which rely on prices aligning with fundamental values over extended periods, short-term trading targets temporary mispricings, favoring frequent, small profits over sustained growth [17, 61]. This contrast manifests across time horizons, risk management, and market focus: intraday trades close within a session to avoid overnight exposure, exploiting microstructure features like order flow imbalances, while long-term approaches endure volatility through diversification, guided by macroeconomic trends [18, 48, 78]. Yet, frequent trading escalates transaction costs, bid-ask spreads and commission, posing a persistent challenge to intraday profitability [80].

Core intraday strategies center on short-term price deviations, exemplified by breakouts, momentum, and mean-reversion. Breakout strategies predict sustained moves when prices pierce support or resistance levels, such as prior session highs, driven by accumulated pressure and often accelerated by stop orders [61, 79]. Momentum strategies assume strong price shifts persist as they attract participation, measured by tools like moving averages or the Relative Strength Index [18]. Mean-reversion strategies, conversely, expect extreme deviations, spurred by overreactions or liquidity gaps, to revert, using oscillators or standard deviation bands to pinpoint reversals [61]. Volume bolsters these approaches, as Dow Theory posits: rising volume validates breakouts, sustains momentum, and signals exhaustion in mean-reversion setups, with models like the Probability of Informed Trading linking surges to informed pressure [48, 79]. Strategies can be enhanced with Monte Carlo simulations, generating synthetic price paths to test performance under uncertainty, incorporating factors like volatility shocks and transaction costs [27]. For instance, simulating breakout trades across thousands of scenarios should theoretically reveal their robustness to slippage or false signals, refining execution tactics in fast-paced markets [41, 61].

Technical indicators distill intricate market dynamics into actionable signals, equipping traders to interpret volatility, momentum, and short-term price trends in real time. These tools, grounded in statistical and behavioral insights, convert raw inputs, such as price fluctuations and volume shifts, into structured cues for timing trades in fast-moving markets [61, 79]. Unlike fundamental analysis, they prioritize immediate conditions, spotlighting inefficiencies and directional shifts. For instance, moving averages might track trend persistence, while oscillators signal potential reversals [17, 60]. Their utility, however, relies on alignment with volume patterns and market context, ensuring signals reflect substantive activity rather than fleeting noise [48].

High-frequency data has transformed this landscape, shifting analysis from daily aggregates to tick-level granularity, exposing order book dynamics and amplifying transaction cost scrutiny [18, 78]. This granularity reveals high-frequency trading's dual impact, enhancing liquidity yet stoking volatility and underscores technology's drive toward efficiency [76, 80]. The Adaptive Market Hypothesis suggests this efficiency fluctuates, preserving technical strategy niches amid low-latency execution demands [64, 75]. Modern intraday trading thus blends data-driven design with rapid adaptation, exploiting fleeting edges in an evolving arena [16, 48].

1.1.2 Auction Theory and Market Profile Concepts

Auction theory provides a powerful lens for understanding how prices emerge and trading activity unfolds in financial markets, a process transformed by the advent of electronic platforms. Rooted in economic studies of resource allocation and competitive bidding [65], this model treats markets as ongoing two-way auctions, where buyers and sellers interact to determine fair value through the interplay of price, time, and trade intensity [24]. This perspective gained practical traction in the 1980s when J. Peter Steidlmayer introduced Market Profile at the Chicago Board of Trade, offering traders a way to visualize these interactions as they happened [93]. Since the 2000s, Market Profile has become a vital tool for discretionary traders, especially in electronic markets, by converting raw market data into clear, actionable insights [63].

At its essence, auction theory suggests markets operate fluidly: price signals opportunity, time governs participation, and trade activity reflects negotiation outcomes [24]. In electronic venues, this plays out via an order book, where prices shift along a scale to match trades, aided by transparency that allows graphing of price, time, and volume in real time [63]. Core ideas include pinpointing fair value, the price range drawing the most trades, and recognizing balance (consolidation) versus imbalance (trend), shaped by order flow and participant actions. For instance, climbing prices paired with growing activity suggest bullish momentum, while falling prices with strong volume point to bearish pressure [24, 63].

Market Profile builds on this model by organizing market behavior into a visual structure, spotlighting the Value Area and Point of Control. The VA captures about 70% of Time Price Opportunities, moments when price is hit at a given time, outlining the range of heaviest trading, framed by the Value Area High and Value Area Low . Prices inside this zone reflect fair value, while those beyond it may signal shifts or trends [25]. The POC, the price with the most activity, marks the session's central value, often close to a statistical mean. A prominent POC, backed by many TPOs, shows broad acceptance, and its movement reveals changes in market control [93]. Trade intensity analysis sharpens this picture: the Volume Point of Control highlights the price with peak volume, high-activity zones serve as support or resistance, and low-activity areas indicate swift rejection, based on the idea that volume confirms price stability [63].

This structure aids intraday trading by identifying critical levels and using trade activity as a guide. In breakout plays, moves past VAH or VAL with rising volume suggest new participants and value shifts, especially if the market starts outside the prior day's range [24]. By contrast, breakouts with fading volume often pull back to the center, a pattern key to reversion strategies. Stable markets, shown as bell-shaped profiles with a steady POC, lean toward reversion when activity fails to hold at extremes [63]. Trade intensity thus gauges auction success: robust volume backs trends, while thin activity at edges supports fading moves [25].

Historically, Market Profile ties to auction theory's focus on structuring market data [65], with its utility boosted by electronic transparency [24]. Yet limitations persist, past data can oversimplify complex patterns, like tails or extensions, and volume benchmarks may vary by interpretation [93]. Investor behavior, often irrational, also dents its precision, framing Market Profile as a visualization tool rather than a definitive predictor [63]. Its value shines in flexibility across assets and conditions, trending or range-bound, though effectiveness rests on trader expertise and pairing with other technical analysis tools [25].

1.2 Market modelling and Prediction Techniques

1.2.1 Traditional Statistical and Financial Models

Traditional statistical and financial models form the foundation of quantitative finance, providing structured tools to analyze and predict market behavior using historical price and volume data. Emerging from early economic ideas and statistical advances, these frameworks gained prominence in the 20th century as markets grew more intricate and data more plentiful. Their growth reflects an ongoing quest to capture market dynamics in mathematical terms, blending real-world evidence with theoretical insights.

The journey started with Louis Bachelier's 1900 insight, which viewed stock prices as a random walk, laying the groundwork for stochastic processes in finance [23]. Initially theoretical, this idea took root as data and computing tools emerged. By the mid-20th century, econometrics advanced with Box and Jenkins' [10] ARIMA models for time-series forecasting, while financial theory progressed through Sharpe's [92] Capital Asset Pricing Model and Black and Scholes' [6] option pricing formula [21, 56]. These leaps, supported by innovations like electronic trading, enabled practical testing of market theories [61]. Fama's [35] Efficient Market Hypothesis suggested past prices couldn't predict future ones, yet recurring anomalies fueled further model refinements [21].

Moving Averages were vital in technical analysis, smoothing price data to highlight trends. Simple Moving Averages average prices evenly over a set period, while Exponential Moving Averages give more weight to recent data for faster responses [79]. Traders often use MAs in strategies, such as buying when a 5-day MA crosses above a 20-day MA or incorporate them with tools like MACD [61]. A prominent extension of moving averages is Bollinger Bands, introduced by John Bollinger, which encase a 20-period SMA with bands reflecting two standard deviations of price volatility [9]. These bands dynamically adjust to market conditions, enabling traders to pinpoint volatility shifts—narrow bands signal consolidation, while widening bands herald potential trends. Bollinger [9] highlights their utility in generating trade signals, such as buying when prices break above the upper band with confirmation, a method frequently backtested in short-term strategies [61]. Their straightforward nature and adaptability keep MAs and their derivatives relevant in market analysis.

Volume-Integrated Indicators enhance price tools by factoring in trading activity. On-Balance Volume tracks cumulative volume based on price direction to measure trend strength, while the Money Flow Index merges price and volume to gauge momentum [61]. Volume spikes or divergences, such as rising prices on falling volume, have long served to confirm or challenge trends, with volume reflecting market conviction [79].

Approaches like Chaikin's volume-weighted method deepen this analysis by tying volume to price ranges [47].

Autoregressive Integrated Moving Average models offer a solid statistical approach for time-series prediction. Combining autoregression, differencing to ensure stability (I), and moving averages of errors, ARIMA captures trends like momentum or mean reversion [10]. For instance, an AR(1) model forecasts tomorrow's price as $p_{t+1} = a_0 + a_1 p_t + \epsilon_t$, where ϵ_t is random noise [97]. This methodical framework, shaped by Box and Jenkins, supports precise parameter tuning [62].

Generalized Autoregressive Conditional Heteroskedasticity models, built by Bollerslev [8] from Engle's [32] ARCH, target volatility clustering, where large price swings spark further unrest. A GARCH(1,1) model sets variance as $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, blending past returns and variances [56]. Variants like EGARCH handle uneven volatility effects, proving valuable in risk management and option pricing [47].

Beyond volatility-focused models, cointegration offers a statistical approach to identify long-term equilibrium relationships between asset pairs, enabling pairs trading strategies where prices diverge then converge [33]. This complements single-asset models like ARIMA by exploiting correlated market movements. Markov Models, such as Hidden Markov Models, further extend this toolkit by inferring latent market states, bull or bear regimes, from observable price and volume data, capturing regime shifts probabilistically [84]. Bayesian Methods, including Bayesian Structural Time Series, enhance adaptability by updating predictions with incoming data, providing uncertainty estimates that refine traditional deterministic forecasts [90].

Regression-Based Models rely on statistical links to explain returns. Sharpe's [92] CAPM connects expected returns to market risk with $E[R_i] = R_f + \beta_i(E[R_m] - R_f)$ [21]. Fama and French [36] added size and value factors, while Fama and MacBeth [37] introduced shifting betas [21]. Autoregressive regressions predict prices from past values, though their linear basis curbs adaptability [58].

Option Pricing Models, such as Black-Scholes [6], assume prices follow a lognormal path with constant volatility, expressed as $C = S_0 N(d_1) - Ke^{-rT} N(d_2)$ [56]. Later frameworks, like Heston's stochastic volatility or jump-diffusion models, address issues like volatility smiles, widening their scope [56].

Backtesting employs these tools to craft signals, evaluating performance with metrics like the Sharpe Ratio ($SR = \frac{E[R_p - R_f]}{\sigma_p}$) and Maximum Drawdown across diverse conditions [61]. Weaknesses still persist. MAs lag in swift markets and falter in erratic ones, yielding false signals [26]. ARIMA and regression models depend on linear, stable assumptions, struggling with abrupt changes [97]. GARCH often skips volume or order flow, overlooking structural shifts [58]. Volume data, clouded by off-exchange trades, can distort outcomes [47]. Black-Scholes' steady volatility assumption clashes with real patterns, needing advanced adjustments [56]. Adaptations counter these flaws. Tick-data analysis refines volume estimates, and GARCH-X incorporates factors like realized volatility [47]. Backtesting with out-of-sample trials and scenarios tests durability across calm, trending, or volatile phases, though historical results only suggest future potential [62]. These updates show the models' capacity to evolve with richer data, despite their constraints. Traditional statistical and financial models offer a robust toolkit for analyzing price, volume, and risk. Their development balances simplicity with realism, providing valuable insights into market behavior. However, their struggles with non-linearities and regime shifts highlight the need for more flexible approaches, paving the way for subsequent innovations in financial modeling.

1.2.2 Machine Learning and AI in Financial Markets

The integration of machine learning and artificial intelligence into financial markets has revolutionized the analysis, prediction, and execution of trading strategies. These computational approaches leverage algorithms to discern patterns within vast datasets, facilitating tasks such as price forecasting, algorithmic trading, and risk management [14, 51]. Departing from traditional statistical methods, ML and AI offer a data-driven paradigm that adapts to the complexities of modern financial environments. This section outlines the historical development of these technologies in finance, elucidates their core principles, and examines advanced applications that capitalize on high-frequency data and multifaceted market signals, culminating in a forward-looking perspective on their future role. The adoption of ML and AI in financial markets has unfolded over decades, propelled by technological advancements. Prior to the 2000s, financial modeling leaned on statistical techniques, with early ML efforts introducing linear models and support vector machines as "intelligent computing" tools [57]. The subsequent decade witnessed enhanced computational capabilities, enabling the exploration of neural networks and hybrid architectures, such as genetic-neural systems, for market forecasting [57]. By the 2010s, breakthroughs in deep learning—initially from fields like image recognition—catalyzed its financial applications, supported by the rise of big data, including high-frequency and alternative sources [53]. Today, an "AI-first finance" approach is gaining traction, where flexible ML algorithms increasingly overshadow traditional financial theories, backed by substantial investments in research and infrastructure [14, 51]. This trajectory reflects a shift from rigid, hypothesis-driven models to adaptive, data-centric methodologies. ML and AI in finance are fundamentally data-driven, harnessing extensive historical and real-time data to identify patterns without reliance on pre-set rules [51]. In contrast to traditional statistical models like ARIMA or GARCH, which presuppose linearity, stationarity, and normality [10, 32], ML imposes fewer constraints, adeptly capturing non-linear and non-stationary dynamics prevalent in financial markets [76]. This adaptability is vital given the markets' inherent noise and unpredictability [77]. Central to these techniques are several methodologies. Supervised learning, encompassing logistic regression, SVMs, and random forests, trains models on labeled data to predict outcomes like price movements or returns [50]. Unsupervised learning, including clustering (e.g., k-means) and dimensionality reduction (e.g., PCA), reveals hidden structures in unlabeled data, useful for portfolio diversification or anomaly detection [58]. Reinforcement learning trains agents to optimize trading decisions through environmental interactions, offering dynamic adaptability [94]. Ensemble methods, such as Random Forests and gradient boosting (e.g., XGBoost), enhance predictive power by aggregating multiple models [11, 20]. Feature engineering transforms raw data into meaningful predictors, a critical step distinguishing ML from traditional approaches reliant on simpler indicators [76]. These methods demonstrate prowess in handling high-frequency data, such as 1-minute or faster OHLCV streams from electronic exchanges, where traditional models falter due to data volume and complexity [53, 62]. Deep neural networks, including recurrent neural networks and long short-term memory units, excel at extracting features from raw time-series data, identifying precursors to events like volume surges or price breakouts [54, 58]. Volume fluctuations, for instance, often signal trend conviction, while order book imbalances hint at imminent price shifts—dynamics ML integrates more comprehensively than lagging traditional indicators like moving averages [76]. Research indicates Random Forests often outperform single classifiers in stock direction forecasting, leveraging feature importance to prioritize key signals [3]. This contrasts with traditional volume-based indicators like On-Balance Volume or Volume-Weighted Average Price, which offer static benchmarks that ML can augment through adaptive learning [46, 57]. Available literature and studies also explore volume-weighted SVMs and microstructural features like the volume-synchronized probability of informed trading, highlighting ML's capacity to synthesize complex volume-related predictors [76, 109].

Genetic Algorithms optimize trading rules or model parameters by simulating natural selection, iteratively evolving solutions to maximize performance metrics like the Sharpe Ratio [43, 55]. Graph Neural Networks model relationships between assets as graphs, capturing spatial dependencies—e.g., sector correlations—missed by sequential models like RNNs, making them ideal for portfolio optimization or multi-asset forecasting [88, 104]. Gaussian Process Regression , a non-parametric Bayesian approach, excels in small datasets, offering probabilistic predictions with uncertainty quantification, useful for volatility modeling or less liquid markets [85].

Anyways, challenges temper these advancements. Financial data's low signal-to-noise ratio, combined with risks of overfitting and alpha decay, poses significant hurdles [76]. In high-frequency contexts, transaction costs and execution speed demands exacerbate these issues, while dynamic parameter adjustments may falter amid market regime shifts if not rigorously validated [62, 76]. R

The future of AI in financial markets promises deeper integration, driven by its ability to process heterogeneous data in real time [14]. For fast-paced trading environments, AI offers real-time analytics, optimal execution, and dynamic strategy adjustments, with RL and deep learning poised to enhance adaptability [51, 58]. While traditional statistical models like ARIMA retain relevance in academic settings [10], the rise of data-driven ML suggests a reduced dependence on their linear frameworks, with hybrid approaches potentially bridging this evolution [29]. Recently, researchers have begun experimenting with transformer-based models, originally developed for natural language processing, to capture long-range dependencies in financial time-series data, offering a promising avenue for enhancing predictive accuracy in dynamic market conditions [69, 99]. As financial institutions invest heavily in AI, its role as a competitive differentiator will likely reshape market practices profoundly.

1.2.3 Transformer-based Models for Time Series Prediction

The Transformer architecture, first introduced in "Attention is All You Need"[99], has initiated a significant paradigm shift in time series prediction. Its influence has expanded from sequence modeling in natural language processing and computer vision to the realm of predictive analytics. The core of the Transformer is its self-attention mechanism, which has redefined the processing of sequential data by overcoming the limitations of recurrent neural networks, such as their difficulties with long-range dependencies [54]. The model's encoder-decoder structure, supported by positional encodings to preserve temporal order, enabled efficient parallel processing, where encoders construct rich input representations and decoders generate predictions [99, 112].

Financial time series, characterized by intricate patterns spanning extended periods, are an ideal domain for Transformer-based models [62]. These models excel at capturing complex, long-range temporal dependencies where traditional statistical methods like ARIMA [10] and exponential smoothing [12] often fall short due to their underlying assumptions. The adoption of Transformers for time series analysis was a natural progression from their success in NLP, where they effectively addressed challenges like the vanishing or exploding gradients inherent in RNNs [99].

The initial application of the vanilla Transformer model between 2018 and 2019 confirmed the potential of its parallel processing and attention mechanisms for time series [112]. However, the quadratic computational and memory complexity of self-attention ($O(L^2)$ with respect to sequence length L) presented a significant bottleneck for long-sequence time-series forecasting. This challenge spurred a wave of innovation. The Informer model reduced complexity to $O(L \log L)$ by introducing a *ProbSparse* self-attention mechanism [112], while the Autoformer incorporated a decomposition architecture and an Auto-Correlation mechanism to leverage the periodicity inherent in many time series [103]. Further advancements between 2022 and 2023 included the development of patching techniques, where a time series is segmented into subseries-level patches. Models like PatchTST demonstrated that this approach could improve performance and efficiency [81]. Concurrently, research by [110] challenged the increasing complexity of models by demonstrating that a simple linear model, DLinear, could outperform more complex Transformers on certain benchmarks. This finding invigorated research into task-specific optimizations and fueled interest in developing versatile foundation models, such as TimeGPT, designed for zero-shot forecasting across diverse datasets [40].

A wide array of specialized Transformer-based models has since been developed to address specific challenges in time series analysis. These models can be broadly categorized by their primary innovation:

- **Efficiency and Locality.** To improve computational efficiency, the Informer uses ProbSparse self-attention and a generative decoder for LSTF [112], while LogTrans integrates convolutional self-attention to enhance focus on local patterns [68]. FEDformer combines frequency-enhanced structures with decomposition to achieve linear complexity [113], and Pyraformer employs a pyramidal attention module to efficiently capture long-range dependencies [108].

- **Decomposition and Architecture.** Autoformer fuses decomposition with an Auto-Correlation mechanism to improve long-term forecasting accuracy [103]. TimesNet offers a versatile framework by modeling temporal variations in a 2D space, capturing multi-scale patterns [102]. PatchTST uses a channel-independent design with patching for distinct pattern recognition and has excelled in multivariate benchmarks [81].

- **Handling Distribution Shifts and Dependencies.** The Non-stationary Transformer was explicitly designed to address the common issue of non-stationarity in time series [73]. For multivariate forecasting, iTransformer inverts the standard architecture to better capture cross-variable insights [72], while Crossformer uses a Two-Stage Attention layer to preserve cross-dimensional dependencies [111].

- **Interpretability and Foundation Models.** The Temporal Fusion Transformer blends recurrent layers with attention mechanisms and includes variable selection to enhance interpretability [69]. Pushing the boundaries of generalizability, foundation models have emerged, including TimeGPT for zero-shot forecasting [40], TIME-LLM, which reprograms large language models for few-shot learning [59], and MOIRAI-MOE, which utilizes a sparse mixture-of-experts to handle diverse datasets [71]. Other notable foundation models include Chronos [2], Lag-LLaMA [86], Moment [42], and Timer [74]. TimeMixer focuses on disentangling short- and long-term dynamics, showing strong performance in volatility forecasting [67].

Key applications include:

- **Stock Price and Volatility Forecasting.** Early work demonstrated the effectiveness of Transformers in decoding complex stock market trends [70]. Models like TFT [69] and PatchTST [81] produce robust forecasts by leveraging diverse inputs and patching techniques. For volatility forecasting, TimeMixer has shown excellence in short-term prediction by identifying fine-scale dynamics [67], while TFT is effective for multi-horizon tasks [69].

- **Trading Volume Prediction.** Researchers have successfully used dual-process meta-learning with Transformer encoders for trading volume prediction [19]. PatchTST has also achieved strong benchmark performance by integrating multi-view data [81].

- **Multi-modal Integration.** The fusion of different data types further enhances predictive power. TFT can incorporate static covariates alongside time-varying data [69]. More recent models like TIME-LLM [59] and Text2TimeSeries [66] integrate textual data using LLMs like BERT [28], allowing the models to reflect market sentiment and economic shifts that influence financial markets [61]. Foundation models like TimeGPT enable comprehensive market modeling across a wide range of financial instruments and datasets [40].

The future of Transformer-based models in financial forecasting appears promising and is evolving rapidly. A key trend is the development of universal foundation models like TimeGPT [40], MOIRAI-MOE [71], and Chronos [2], which aim to provide powerful zero-shot forecasting capabilities, thereby reducing the need for task-specific model engineering [101]. Progress in multi-modal fusion is also critical; research into how LLMs understand time series features [39] complements applied models like TIME-LLM [59]. Continuous improvements in efficiency, such as the Adaptive Sharpe Ratio Optimization in TFT-ASRO [106], enhance scalability for large-scale applications. Furthermore, enhancing model interpretability [69] and uncertainty quantification is crucial for building trust and managing risk in financial contexts [14]. Hybrid techniques that combine Transformers with methods like PCA [105] or Retrieval-Augmented Generation [107] promise even greater accuracy and adaptability, which is particularly valuable for high-frequency trading applications [52]. As of late 2024, new methodologies, frameworks, and models continue to be archived and published at a remarkable pace, signaling a rapid advancing area of research.

2. Proposed Methods

2.1 Framework and Data

2.1.1 Framework

Python serves as the primary programming language for this analysis. During development, raw scripts automated specific processes, while Jupyter notebooks facilitated data collection, exploratory analysis, preprocessing, prototyping, experimentation, and modelling. Development occured both locally, using integrated development environments and on cloud to leverage additional computational resources. Data acquisition and algorithmic workflows utilized REST APIs, including the Polygon.io API for market data and others.

2.1.2 Data

Data Selection

NASDAQ was selected as the exchange for this study due to several factors such as hosting many stocks across various sectors, while being fully electronic, which ensures higher data consistency, lower latency, and more reliable execution tracking. Using data from a single exchange simplifies the collection and avoids the need for cross-exchange aggregation, which can introduce discrepancies due to differing market structures and trading hours.
For this thesis, 4798 securities were selected for analysis and trading. These securities were chosen based on their presence in the official listings as of 17/02/2025, and their availability at the start of the backtesting period. The full dataset spanned five years, beginning on 17/02/2020, and was sourced from the Polygon.io API [83]. Each observation included timestamps, ticker symbols, open, high, low, close, and volume prices, previous session high and low, estimated order book data. Later, new features are computed and collected. In the end, the results were obtained only from NASDAQ-100 securities within 9 months of timeseries data.

Volume Data Approximation

While price data is standardized and directly observable, volume data is approximated due to variations in reporting across trading venues. The Polygon.io API provides volume based on trades reported by NASDAQ and participating venues, but it may exclude off-exchange transactions, iceberg orders, or hidden liquidity. Volume is approximated by integrating trade feeds and order book activity, with interpolation techniques addressing delayed reporting or missing data. Consequently, the volume data in this thesis relies on NASDAQ's reporting and Polygon.io's aggregation methodology, potentially not reflecting the absolute true traded volume. That means historical data has almost identical distribution of volume shares traded, while the actual values and range is shifted.

Data Storage

Data was stored locally and on cloud depositories in CSV, JSON, and Parquet and other formats for development, analysis and experimentation. For automated execution, cloud storage would have been employed to ensure scalability and accessibility, supporting efficient data retrieval and processing.

Data Processing

Data processing transformed raw data into a suitable format for analysis, addressing missing values, deriving features, and mitigating biases. Additional features, such as technical indicators and news impact flags, were computed to enhance the dataset. A key objective was to flag trading on days with high-impact news events, as the strategies do not incorporate news or sentiment analysis. A binary indicator flags high-impact news days, defined as scheduled macro-level events (e.g., FOMC Interest Rate Decisions, Non-Farm Payrolls, CPI/PPI releases, U.S. Election days, GDP Reports, Fed Meeting Minutes, ISM PMI, and ECB/BOE/BOJ Policy Decisions) and micro-level events (e.g., Earnings Reports, SEC Filings, Stock Splits, Product Launches, and S&P 500 inclusions). These events are known to drive significant price movements [31].
Unpredictable news events (e.g., M&A announcements, executive changes, or unscheduled regulatory decisions) were not systematically tracked due to their spontaneity. However, a news calendar assigned impacts based on event timing: pre-market news affecting the same day's session, scheduled intra-day news affecting current session, and after-hours news impacting the next trading day. Unscheduled intra-day news was not accounted for, as they require real-time sentiment modelling beyond this thesis's scope. This approach aligns with industry practices for event-driven analysis [15].

Historical News Backtesting Calendar

The historical news backtesting calendar aggregated scheduled macro and micro-events with implied volatility of 1% to 5% (occasionally up to 10%), based on their documented market impact. Macro-events included:

- FOMC meetings and interest rate decisions (Federal Reserve website) [95].
- Federal Reserve Chairman speeches (Federal Reserve website) [96].
- U.S. Election days (Wikipedia) [100].
- Non-Farm Payrolls, CPI/PPI, GDP Reports, and ISM PMI (FRED API) [38].
- ECB, BOE, and BOJ policy decisions (respective central bank websites) [4, 5, 34].

Micro-events included:

- Earnings Reports (Dolthub database) [30].
- SEC 8-K filings, including Earnings Guidance and Regulation FD Disclosures (sec-api.io) [91, 98].
- S&P 500 inclusions (Wikipedia, with S&P Global as an alternative) [87, 100].
- Product launches, partnerships, and milestones, compiled using the Gemini 2.5 Pro model (March 2025 preview, Google Cloud Console) [45], with an estimated accuracy of 0.95 (Wilson score interval [0.88, 0.97]). A.

The calendar maps events to a binary news indicator by ticker and date, flagging the high-impact days during backtesting. See Table 6.1.

Backtesting Dataset Creation

The backtesting dataset was constructed from minute-level data, initially comprising Date, Time, Ticker, Open, High, Low, Close, and Volume, recorded in Eastern Standard Time (EST, UTC-5) per NASDAQ's standard. The raw data exhibited gaps (e.g., missing rows, out-of-bounds timestamps) due to API retrieval issues at 1-minute granularity. To ensure robustness, additional features were derived, and a preprocessing pipeline addressed gaps, scalability, and market dynamics.

The dataset, denoted (D), included observations per ticker at specific date-times, with core features:

- $(High_i)$: Highest price for observation (i).
- (Low_i): Lowest price for observation (i).
- $(Close_i)$: Closing price for observation (i).
- $(Volume_i)$: Trading volume for observation (i).

Derived features included previous session high and low, estimated bid-ask spread, order-book depth and 50-day simple moving average. See Appendix A.

Preprocessing pipeline with Pandas and Dask ensured scalability, focusing on NASDAQ-100 tickers to manage computational demands (subset of ~800,000 rows). Key steps included:

- Timestamp Creation: Combined Date and Time into a unified timestamp.
- Standardization of column names: (e.g., 'Ticker' to 'ticker').
- Deduplication: Handled duplicates in between operations.
- Trading Hours Filter: Retained data from 4:00 AM to 8:00 PM ET, Monday to Friday, covering pre-market, regular, and post-market sessions.

For the handling of missing values, a complete minute-level time index was generated within daily trading bounds, merged with the dataset, and nulls were filled session-specifically:

Regular Trading Hours Volume Volume in regular trading hours was filled using linear interpolation, selected as the most practical method for this study. Spline interpolation and more advanced techniques were considered but deemed overly complex. In liquid stocks, such as those in this subset, volume exhibits clustered volatility and stable patterns over short intervals, with elevated activity at market open and close. Small gaps of 1-3 minutes, likely due to API retrieval issues, were filled by assuming gradual volume changes, consistent with the continuous trading environment of liquid markets where such gaps are typically data artifacts rather than periods of no trading [49].

Pre-Market and Post-Market Volume Volume in pre-market (4:00 AM to 9:30 AM) and post-market (4:00 PM to 8:00 PM) periods was filled with a small non-zero value. Sparse trading in these extended hours, driven by low liquidity, results in frequent gaps of 5-10 minutes or more. Assigning a small non-zero value reflects negligible trading activity while preventing numerical instability in metrics like order book depth, a standard practice in high-frequency data processing that ensures dataset continuity for robust backtesting [49].

Price Data Forward-filling within ticker-timestamp groups for price columns was used, reflecting price stabililty in low-activity periods in a live-trading scenario. Backward-filling addressed a small, negligible remaining percentage of remaining nulls, for the purposes of full dataset, in case of missing data incompatibility with models and simulation scenarios.
The dataset was saved as a ticker-partitioned Parquet file, leveraging Dask for scalability and PyArrow for consistent data types. Float32 for numeric columns and categorical tickers reduced memory usage, enhancing query performance. Full schema of preprocessed dataset in: Table 6.3.

2.2 Proposed strategies

Strategies are designed for intraday trading, meaning all positions are opened and closed within a single market session (9:30 AM to 4:00 PM ET), with no overnight holdings. This thesis introduces three strategies, each with three levels of increasing complexity: baseline, volume-enhanced, and deep learning-enhanced. The strategies are variations of breakout, mean reversion, and momentum paradigms. The baseline level consists of simple, price-action-based implementations. The second level enhances these strategies with volume-based indicators, drawing concepts from "Market Profile" trading systems where volume serves as a critical confirmation signal. The third and most advanced level incorporates a transformer-based deep learning model to forecast future volume, aiming to anticipate market movements and improve performance. This model uses a 60-minute lookback window of market data to predict the trading volume over the subsequent 15-minute horizon. The specific strategy formulas are detailed in Appendix B, with optimizable parameters in Tables in section 6.1.2.

2.2.1 Baseline

Basic Breakout

This strategy identifies and trades price movements that extend beyond the previous session's high or low, using pure price action as the primary signal.
Entry Signals: A buy signal is generated when the price breaks above the previous session's high. Conversely, a sell signal is generated when the price breaks below the previous session's low.
Exit Signals: Positions are managed with a static stop-loss set at a fixed percentage below the entry price for long positions and a take-profit target set at a fixed percentage above the entry price.
Preparation: The strategy requires the calculation of the highest and lowest price levels from the preceding trading session. These levels serve as the primary thresholds for trade entry.
Execution: Trades are executed immediately when the price crosses either the previous session's high or low. All entries are restricted to the time before 15:00 EST to avoid late-day volatility.

Simple Momentum

This strategy initiates trades based on sustained price momentum, identified by a consecutive series of bullish or bearish price movements over a defined lookback period.
Entry Signals: A buy signal is triggered after a specified number of consecutive bullish candles (Close > Open). A sell signal is triggered by an equivalent number of consecutive bearish candles (Close < Open).
Exit Signals: Instead of a fixed take-profit, this strategy employs a trailing stop-loss, which dynamically adjusts to lock in profits as the trade moves favorably.

Preparation: Key parameters include the lookback period for identifying consecutive price moves and the percentage distance for the trailing stop-loss.
Execution: The momentum is confirmed when the price has moved in the same direction for a set number of periods. The trailing stop-loss continuously updates its level based on the highest price reached during a long trade or the lowest price during a short trade.

Mean Reversion using Bollinger Bands

This strategy utilizes Bollinger Bands to identify statistically overbought or oversold conditions, operating on the principle that price will revert to its mean. The bands are calculated based on a simple moving average and its standard deviation.
Entry Signals: A buy signal occurs when the price closes below the lower Bollinger Band, indicating an oversold condition. A sell signal occurs when the price closes above the upper Bollinger Band, indicating an overbought condition.
Exit Signals: The primary exit signal for a long position is when the price reverts and crosses above the middle Bollinger Band (the simple moving average). For a short position, the exit is triggered when the price crosses below the middle band.
Preparation: The strategy requires defining the lookback period for the moving average and the standard deviation multiplier, which determines the width of the bands.
Execution: Trades are initiated upon a clear breach of the upper or lower band, with the expectation of a price reversal. Exits are timed to capture the profit from this reversion.

2.2.2 Volume-enhanced

Breakout using Value Area

This strategy enhances the breakout concept by using the developing intraday Value Area , defined by volume distribution. Trades are triggered when the price breaks out of the Value Area High or Value Area Low , with confirmation from volume and trend strength indicators.
Entry Signals: A buy signal is generated when the price breaks above the developing VAH. A sell signal is generated on a break below the developing VAL.
Exit Signals: Exit levels for stop-loss and take-profit are dynamically set using a multiple of the Average True Range (ATR), adapting to current market volatility.
Preparation: The strategy calculates the developing Point of Control (dPOC), VAH, and VAL using the cumulative volume profile of the current day. It also requires a trend strength filter using the Average Directional Index and a volume surge indicator.
Execution: A breakout is only considered valid if it is accompanied by a significant volume surge and the ADX confirms a strong underlying trend. Entries are typically avoided during the initial market opening period (e.g., first 60 minutes) to allow the volume profile to develop.

Mean Reversion using VWAP and OBV

This strategy identifies mean-reversion opportunities by analyzing price deviations from the Volume-Weighted Average Price, confirmed by On-Balance Volume to gauge buying or selling pressure.
Entry Signals: A buy signal is triggered when the price is below VWAP, OBV is below a lower quantile threshold, the VWAP slope is positive or flat, and the medium-term VWAP trend is down. A sell signal is the inverse: price above VWAP, OBV above an upper quantile, negative or flat VWAP slope, and an upward VWAP trend.
Exit Signals: Stop-loss and take-profit levels are set using ATR multiples to manage risk based on current volatility.
Preparation: The strategy requires calculating the intraday VWAP, its slope, and its medium-term trend. It also uses rolling historical quantiles of OBV to set dynamic overbought/oversold thresholds.
Execution: The combination of price deviation from VWAP and extreme OBV readings identifies a high-probability setup. The VWAP slope and trend filters ensure the trade aligns with the expected short-term reversal.

Volume-driven Momentum

This strategy identifies trend continuation setups by combining price momentum with a surge in relative volume, confirming that significant market participation is driving the move.
Entry Signals: A buy signal requires the price to exceed the prior session's high, relative volume to surpass a key threshold, and ADX to confirm strong trend momentum.

A sell signal requires the price to fall below the prior session's low with similar volume and ADX confirmation.

Exit Signals: The strategy exclusively uses an ATR-based trailing stop-loss to lock in profits during a sustained trend, with no fixed take-profit target.

Preparation: This involves calculating relative volume (current volume divided by its moving average) , the ADX for trend strength, and the prior session's high and low.

Execution: Trades are entered only when a price breakout is validated by both high relative volume and a strong ADX reading, indicating a high conviction move. The trailing stop protects profits as the trend extends.

2.2.3 Deep learning-enhanced

Breakout using Value Area with Predicted Volume

This strategy enhances the Volume-Enhanced Breakout by using a deep learning model's 15-minute-ahead volume predictions to construct the Value Area, providing a leading indication of where institutional interest may lie.

Entry Signals: A buy signal is triggered if the price breaks above the VAH calculated from predicted volume, confirmed by a high predicted volume forecast and a strong ADX reading. The sell signal is triggered on a break below the predicted-volume VAL under similar conditions.

Exit Signals: Risk is managed with dynamic stop-loss and take-profit levels based on ATR multipliers.

Preparation: The core of this strategy is the real-time calculation of the developing VAH and VAL using predicted volume instead of actual volume. It also requires a confirmation that the predicted volume represents a significant surge.

Execution: By using a forecast of future volume, the strategy aims to enter breakouts just as participation is expected to increase, potentially leading to stronger follow-through. The ADX filter remains crucial for confirming directional bias.

Mean Reversion using VWAP with Predicted Volume

This strategy integrates future volume predictions into the VWAP reversion framework. It aims to anticipate reversals by identifying setups where a price deviation from VWAP coincides with a predicted surge in trading activity.

Entry Signals: A buy is signaled when the price is significantly below VWAP, the 15-minute ahead predicted volume is a multiple of current volume, and the VWAP is in a short-term downtrend (primed for reversal). A sell signal is triggered under the opposite conditions.

Exit Signals: The primary exit is a reversion to the mean, triggered when the price crosses the VWAP line. ATR-based stop-loss and take-profit levels serve as secondary, volatility-based exits.

Preparation: A trained deep learning model provides the 15-minute ahead volume forecast. The strategy requires setting a deviation threshold from VWAP and a lookback period to determine the VWAP trend.

Execution: Entry is conditioned on a significant deviation from VWAP, confirmed by a predicted increase in volume, which suggests an imminent reversion. The trade is ideally closed once the reversion thesis is completed.

Volume-driven Momentum with Predicted Volume

This strategy refines the Volume-driven Momentum approach by using a forecast of sustained volume increase to anticipate and confirm trend continuation, rather than relying on reactive volume surges.

Entry Signals: A buy signal is triggered when the price exceeds the prior session's high, the moving average of predicted future volume shows a rising trend, and ADX confirms momentum. A sell signal is triggered on a break of the prior session's low with similar predicted volume and ADX confirmation.
Exit Signals: Profit is captured via an ATR-based trailing stop-loss, which dynamically follows the extending trend without a predefined target.
Preparation: The strategy relies on a forecast of 15-minute ahead volume, which is then smoothed with a moving average to identify a sustained trend of increasing volume. It also requires the prior session's levels and an ADX indicator.
Execution: A trade is entered when a price breakout aligns with a confirmed trend in predicted volume, suggesting that institutional activity will support the momentum. The trailing stop is crucial for maximizing gains from trend continuations.

2.3 Proposed setup

2.3.1 Regime-based categorization

Securities are classified into regime buckets based on their market conditions over a 30-day lookback period, enabling robust experimentation and hypothesis testing. Initially, a multi-filter approach was considered to select securities tailored to specific trading strategies. However, a regime-based classification was adopted to provide a more flexible and adaptive framework for strategy evaluation. This approach uses dynamic indicators to define volatility, trend, and liquidity regimes, with thresholds that adapt to each security's historical distribution over a 252-day rolling window (approximately one trading year), ensuring responsiveness to changing market conditions. The complete formulas for these regimes are detailed in Appendix C.
To avoid statistical neglect of any regime, thresholds are set using quantiles of the 252-day historical data, promoting balanced representation across regimes. If insufficient data (<126 days) is available, static fallback thresholds are applied, or the regimes are marked as 'undefined' if data is inadequate. The resulting data structure is outlined in Table 6.2.

Volatility-Based Regimes Volatility is classified using the Stochastic Oscillator (%K).

- **Low:** The slow Stochastic Oscillator (%K) is below the 15th percentile of its values over the past 252 trading days, indicating low volatility. A fallback to %K $<$ 20 is used if historical data is insufficient.
- **Medium:** The slow Stochastic Oscillator (%K) is between the 15th and 75th percentiles of its 252-day historical distribution, reflecting moderate volatility. A fallback to %K between 20 and 80 is used if necessary.
- **High:** The slow Stochastic Oscillator (%K) is above the 75th percentile of its 252-day history, signaling high volatility. A fallback to %K $>$ 80 is used if necessary.

Trend-Based Regimes Trend classification uses the Average Directional Index (ADX) and Bollinger Band Width.

- **Uptrend:** The ADX is above the 75th percentile of its 252-day history with the positive directional indicator (+DI) greater than the negative indicator (-DI), confirming a strong upward trend. This is supplemented by observing higher highs and higher lows. The fallback is an ADX $>$ 25.
- **Downtrend:** The ADX is above its 75th historical percentile with -DI $>$ +DI, indicating a strong downward trend, accompanied by lower highs and lower lows. The fallback is an ADX $>$ 25.
- **Range:** The ADX is below its 25th historical percentile, and the Bollinger Band Width is below its 5th historical percentile, signaling no clear trend. The fallback is an ADX $<$ 20.

Liquidity-Based Regimes Liquidity is categorized based on the estimated bid-ask spread and order book depth (OBD).

- **High liquidity:** The bid-ask spread is below its 10th historical percentile, and OBD is above its 70th historical percentile, indicating tight spreads and a deep market. A fallback to a spread $< 0.02\%$ is used if needed.
- **Medium liquidity:** The bid-ask spread and OBD are between their respective 10th and 95th historical percentiles, reflecting moderate liquidity. The fallback is a spread between 0.02% and 0.1%.
- **Low liquidity:** The bid-ask spread is above its 95th historical percentile, and OBD is below its 30th historical percentile, indicating wide spreads and shallow liquidity. The fallback is a spread $> 0.1\%$.

Regime-Strategy Performance Hypothesis

- **Breakout strategies** are hypothesized to perform well in medium and high volatility conditions and in both uptrends and downtrends. Performance in ranging markets is uncertain, and low volatility is expected to hinder performance.
- **Reversal strategies** are expected to be effective in low and medium volatility, particularly in ranging and downtrending markets. They are considered less reliable in high volatility, with the exception of the predictive volume models, and not applicable in strong uptrends.
- **Momentum strategies** are hypothesized to perform well in medium and high volatility and in trending markets (up or down). They are expected to underperform in low volatility environments and ranging markets.

2.3.2 Execution phase

The proposed setup intended to run all strategies on the entire NASDAQ-100 subset across multiple trading sessions to enable rigorous evaluation and hypothesis testing. This comprehensive approach, while computationally intensive, was designed to assess strategy performance across diverse market conditions. However, due to time and computational resource constraints, the execution phase was limited to a three-month period from November 17, 2024, to February 17, 2025, focusing on a subset of NASDAQ-100 stocks selected based on the regime categorization. After the preparation and regime-based categorization, strategies are executed on this selected subset.

2.3.3 Incubation phase

The proposed setup envisioned deploying trading strategies live for a 3- to 6-month period to validate backtest results in real market conditions. Strategies would be implemented only in market regimes where they demonstrated statistical success, with only one strategy type per stock per session to enhance execution efficiency. Market data would be collected throughout the period, with performance logs generated for assessment. A broker for the demo trading account and a cloud provider for deployment would be selected. However, due to time and computational resource constraints, the selected incubation period in April 2025 experienced significant market volatility driven by executive changes in the U.S. administration, which led to frequent news-driven disruptions [7]. Since the models did not include sentiment analysis, this rendered the news-impact feature ineffective. Consequently, the incubation phase was halted, and the project was suspended from inclusion in the thesis.

3. Experiments

3.1 Volume modelling

3.1.1 Trade-offs for modelling volume

Choosing to model trading volume instead of price offers a significant trade-off between the predictability of market activity and the direct pursuit of profit. Volume is considerably more predictable than price. It often displays robust statistical patterns, such as the well-documented daily "U-shaped" curve, making it less prone to the erratic noise that can obscure trends in price returns [49]. This superior predictability and the stability of its models make volume forecasting an excellent tool for optimizing trade execution and minimizing market impact, particularly for large institutional orders where even small efficiencies can lead to substantial savings.

The principal drawback of focusing on volume is that its forecasts do not predict the direction of price movements. Consequently, volume modeling is not suited for direct alpha generation, the core objective of speculative trading strategies. While a surge in trading volume might confirm the strength of an existing price trend, it does not, by itself, generate the initial buy or sell signal. In essence, the trade-off is between prioritizing cost minimization and execution efficiency versus direct profit generation. For strategies centered on alpha, price prediction, despite its greater difficulty, remains the primary goal [22].

It is crucial to understand that trading volume is not the same as volatility, although the two are related. Volume measures the quantity of shares traded, reflecting the level of market activity and liquidity. Volatility, on the other hand, measures the magnitude of price fluctuations, indicating the degree of uncertainty and risk. A high trading volume can occur with very little price change if there is a balance between buyers and sellers, signifying high liquidity but low volatility. Conversely, a significant price swing can occur on low volume, especially in illiquid assets where a few trades can disproportionately affect the price. Therefore, while a spike in volume can sometimes trigger a spike in volatility, they are distinct market metrics.

The intraday patterns of volume and volatility are often similar, not divergent. It is a common misconception that maximum volumes are traded in the middle of the session. Both trading volume and volatility tend to follow a similar pattern. They are typically at their highest during the opening and closing hours of the trading day and experience a decrease during the midday period. The heightened activity at the open is driven by reactions to overnight news and the establishment of new positions, while the close is characterized by position adjustments and portfolio rebalancing into the end of the day.

3.1.2 Modelling features

The experimentation dataset comprised of 3 months of AAPL stock data at a 1-minute timeframe, spanning the main trading hours (9:30 AM to 4 PM ET) with a 1-hour pre-market buffer to capture early market dynamics going into the session. The data was normalized, and temporal indicators, regime-based features, and lagged variables were added, ensuring compatibility with transformer-based models. The standardized AAPL data was loaded from its directory and merged based on timestamps. Regime features were encoded to ensure numerical compatibility. To capture autocorrelation in volume and price, 15 lagged features (t-1 to t-15 minutes) for both volume and closing price were computed. A configuration of 60 lags was tested but deemed computationally intensive with marginal performance gain. Missing values in the first lagged values were not present due to the use of pre-market hours to introduce a smooth transition to trading hours. Continuous features including price, volume, estimated order book depth, bid-ask spread, previous session highs and lows, and the 50-day MA were normalized using z-scores. The target variable, trading volume, was log-transformed to handle skewness and scaled to the range of the other variables for stability. 6.4.
To prevent data leakage, modelling avoided using future information. The final dataset excluded the pre-market buffer used for lagged feature computation.

3.1.3 Model selection and training

A suite of transformer architecture models was employed, focusing on their ability to capture temporal dependency and handle high-frequency financial data. The model experimentation included:

- **TFT**: Temporal Fusion Transformer with attention mechanisms, supporting exogenous variables.
- **PatchTST**: Patch-based Transformer optimized for long sequence processing, without exogenous variable support.
- **VanillaTransformer**: Standard Transformer architecture.
- **Informer**: Informer model with efficient attention for long sequences.
- **Autoformer**: Autoformer with decomposition-based attention.
- **FEDformer**: Fourier-enhanced Transformer.
- **iTransformer**: Improved Transformer variant.
- **TimeXer**: Time-series specific Transformer.

The library provided "Auto" model variants that automatically tuned hyperparameters to optimize performance. The models forecasted log-transformed and scaled trading volume 15 minutes ahead, using a 60-minute lookback window. To prevent data leakage, purged cross-validation with a 60-minute embargo ensured no overlap between training and validation sets within the lookback period. Trained with a MAE loss function, the models achieved a validation MAE of approximately 0.3 in the best cases across most trials for the 15-minute horizon, with a test R^2 of approximately 0.85 in scaled space, indicating strong predictive performance.
Since the target variable was scaled, a MAE of 0.3 corresponds to an error of approxi-

mately 15% of the target range D.4. This error reflects moderate accuracy given the volatility and noise in 1-minute stock volume data D.

Training sometimes showed higher training loss than validation loss, likely due to regularization techniques like dropout, which are active during training but disabled during validation. The noisy data and purged cross-validation further amplified this gap. Epoch-by-epoch loss tracking helped monitor underfitting and overfitting risks.

The TFT model was chosen for its superior performance and robustness to correlated features, such as lagged closing prices and returns. The PatchTST was a strong alternative, excelling in speed and efficiency for long sequences, but it is better suited for long-term predictions or tick-level data with millisecond-granularity order book information.

3.1.4 Meta-Learning Framework

The initial plan was to generalize volume forecasting across an index using a global model and ticker-specific fine-tuned versions of the meta learner. A global model should have been trained on a cross-stock dataset to capture shared patterns and volume dynamics, retrained quarterly (20 times over 5 years) to incorporate the latest market trends, and fine-tuning should run daily for each stock using the most recent 30 trading sessions. This approach should ensure adaptability to both market-wide and stock-specific behaviors. [19].

3.1.5 Corrective models

Following the generation of predictions, corrective regressors such as Mincer-Zarnowitz, Isotonic, and XGBoost were experimented with. The aim of these models was to improve predictive accuracy by modeling the residuals between the predicted volume at t+15 and the actual true volume values over a brief training period. However, this approach did not demonstrate significant promise in enhancing the model's capabilities. The conclusion drawn was that correcting the Temporal Fusion Transformer's predictions would necessitate a different strategy. One scrapped alternative idea was to apply isotonic functions for correction specifically during high-volume periods, while leaving the predictions for low-volume periods unadjusted.

3.1.6 Volume modelling results

In the volume modelling phase, the initial methodology was to use a meta-learning framework to generalize forecasting across the index, involving a global model retrained quarterly and fine-tuned daily for each stock. However, due to computational limitations, this approach was not fully implemented. Instead, individual transformer models were trained for each stock using the 3-month window before the 6-month optimization and simulation period for each stock's dataset with the same features as the experimentation AAPL dataset, resulting in 6 months' worth of volume predictions for the 100 stocks. The Temporal Fusion Transformer model was selected for its superior performance, achieving a median R^2 of 0.74, a median MAE of 0.53, and a median RMSE of 0.63

across the 100 stocks in the evaluated period over the scaled space. Keep in mind the predictions on the scaled space reflect the true prediction power because the features going to and from the model can be easily preprocessed and de-scaled in a pipeline setting.

The unique insight from volume evaluation was that there was higher than expected variation across the index. While the model performed well among all securities, even though trained for a time horizon of 15 minutes, for some tickers it performed exceptionally well even in windows of 5-10 minutes, perhaps even better than for longer horizons. That came as a surprise because the model should perform best in the way it was designed, which was for the longer horizons. If computational resources were not an issue, the model could have been trained for multi-horizon prediction. I applied a log-space noise model to generate a benchmark for the evaluation of the TFT prediction. For each ticker, the true volume 15 minutes ahead ($\text{true\_volume}_{t+15}$) was log-transformed as $\ln(1+\text{true\_volume}_{t+15})$. Gaussian noise was then added to simulate prediction errors:

$$\text{pred\_log\_volume}_{t+15} = \ln(1 + \text{true\_volume}_{t+15}) + \mathcal{N}(0, \sigma_{\text{noise}})$$

The noise standard deviation (σ_{noise}) was calibrated using ticker-specific or median metrics:

$$\sigma_{\text{noise}} = \left(\frac{\text{scaled\_MAE} \cdot \sigma_{\text{log\_}volume}}{2} \right) \cdot \sqrt{\max(0, 1 - \text{scaled\_R}^2)} \cdot 1.5$$

Here, scaled_MAE and scaled_R^2 were derived from the TFT's performance metrics, with $\sigma_{\text{log\_}volume} \approx 1.0$ assumed as the standard deviation of log-volume. For tickers with unreliable metrics (scaled_R^2 < 0.5, negative scaled_R^2, negative scaled_MAE, or scaled_MAE > 0.8) or missing metrics, median values (scaled_MAE $= 0.553$, scaled_R^2 $= 0.752$) were used. The predicted volume was then computed as

$$\text{pred\_volume}_{t+15} = \exp(\text{pred\_log\_volume}_{t+15}) - 1$$

This approach ensured realistic error distributions (15–30% relative errors) for evaluation against the true predictions. The true predictions were generated in batches for the optimization and simulation periods and then analyzed against the log noise benchmark. After discrete analysis, the results showed that the TFT managed to capture the distribution tendencies of the volume during the market hours but failed to capture higher volatility in the early hours and ending hours of the session, most likely due to sentiment and news not being incorporated in it. The model showed bias in underestimating the range of volume spikes. A lot of the non-aggregated performance metrics showed unclear results due to outliers and noise in the unscaled target variable space.

3.2 Backtesting and Evaluation

3.2.1 Backtesting procedure

The proposed setup aimed to backtest trading strategies over a comprehensive five-year period, from March 6, 2020, to February 17, 2025, using historical data from 4798 NASDAQ stocks. The goal was to simulate a realistic trading environment with a rolling buffer for preparation and retraining, ensuring robust evaluation across diverse market conditions.

The process was designed to mimic a real trading environment with continuously updated data. On each simulated trading day, the model was supposed to look 30 days into the past and retrain the stock-specific component of the transformer volume prediction model. This retraining was intended for all stocks meeting specific liquidity criteria or exhibiting significant volume changes, such as a deviation greater than 10% from the 30-day average, thereby maximizing adaptability. Following retraining, the model would compute all necessary indicators for execution and execute trades based on the signal generation logic. The process would then shift forward by one day and repeat.

Due to significant computational resource and time constraints, the scope of the backtesting had to be substantially reduced. The backtesting simulation that was ultimately performed spanned only the last three months of trading days data for optimization and another three months for the final simulations. This was conducted on a limited set of the NASDAQ-100 stocks. To manage the computational demands, several adjustments were made. The strategies were run only during the hours of 9:30 AM to 4:00 PM ET. Furthermore, to reduce computational overhead, only one iteration of the prediction model was used for the entire three-month window for each stock, omitting the global model and forgoing the daily retraining process.

3.2.2 Optimization

The optimization phase was designed to identify high-performing hyperparameter sets for each trading strategy and to analyze the resulting performance characteristics.

Initial explorations revealed that many strategies required architectural adjustments for intraday viability. For instance, breakout strategies, which often capture longer-term swings, were modified with a time-based filter to prevent positions from being held into the market close. Such adaptations were crucial in assessing the inherent suitability of each strategy archetype for short-term trading.

The optimization of strategy hyperparameters was conducted using a Batched Random Search methodology. This approach was chosen for its efficiency in exploring high-dimensional parameter spaces where an exhaustive grid search would be computationally intractable. The process was defined as follows: for a given strategy, the set of all possible parameter combinations constitutes the parameter space P, which is the Cartesian product of the discrete values for each hyperparameter. From this extensive space, a sample subset $S \subset P$ of a fixed size (n trials) was generated by drawing parameter vectors uniformly at random.

To leverage parallel processing, this sample set was partitioned into smaller batches, which were submitted to distributed computing workers for evaluation.

For every parameter vector $s \in S$, an objective function $f(s)$ was computed. For the scope of this work, the objective function was the Calmar Ratio E.8. The goal of the optimization was to identify the parameter vector s^* that maximizes this objective function:

$$s^* = \arg\max_{s \in S} f(s)$$

This was achieved by first finding the local maximum within each processed batch and subsequently identifying the global maximum across all batch results. The final reported s^* represented the best-performing parameter set discovered within the randomly drawn samples for each specific security.

Parameter Generalization

The optimization was conducted on a 3-month in-sample period for each of the 100 securities, with a subsequent 3-month period left as a hold-out set for final testing. While the best practice would be to use ticker-specific parameters (s^*) for each security, a generalized parameter set was created for each strategy for the purpose of evaluating its broad-market effectiveness. This was achieved by taking the mean of the best-performing parameter values found across all 100 securities for a given strategy. This averaging approach created a single, robust parameter set that represents a central tendency of what works across the asset universe, acknowledging a trade-off between personalization and generalizability. Surprisingly, or perhaps not, manually selected "best-fit" parameters were found to be very similar to these calculated mean values. The resulting generalized parameters, presented in Tables 6.5 through 6.13, were used for the final simulations.

The table 6.14 presents the mean performance metrics for each strategy in optimization. It is critical to recognize that these results were inherently biased by the choice of the Calmar Ratio as the optimization metric. The performance profile of each strategy was therefore a direct reflection of this specific objective.

The Calmar, Sharpe, and Sortino Ratios were all exceptionally high. This indicates that the strategies generated outstanding returns not only in proportion to their maximum drawdowns but also relative to their total daily volatility and particularly their downside risk In general, Calmar and Sharpe Ratios above 3.0 are considered excellent, and these results far overfitted that threshold. The consistently negative Kelly Criterion highlighted the inherent trade-off of this approach. A positive Kelly value indicate a profitable edge suitable for capital investment, whereas a negative value suggests that despite the high average returns, the risk of large individual losses makes the strategies unsuitable for optimal long-term capital compounding.

3.3 Risk Management

Effective risk management is fundamental to this study, serving two primary purposes: ensuring the theoretical viability of the trading strategies and enabling a realistic performance evaluation. The framework is designed to move beyond a frictionless simulation by imposing practical constraints, thereby creating a robust environment for comparing the baseline, volume-enhanced, and deep learning models. This approach ensures that the evaluation reflects the challenges of capital preservation and cost management inherent in live trading across the different identified market regimes.

3.3.1 Execution Model and Cost

To ground the backtest in realistic market conditions, the execution framework was defined with distinct parameters for the optimization and the final simulation phases. This separation allows for efficient hyperparameter tuning during optimization while ensuring the final evaluation is conducted under a more stringent and realistic set of trading assumptions.

Optimization Phase Execution Parameters

During the hyperparameter optimization process, the following assumptions were made to balance realism with computational feasibility:

- **Slippage:** Based on analysis of NASDAQ bid-ask spreads at one-minute granularity, an estimated slippage of 1–2 basis points was modeled. To maintain a conservative stance, the upper limit of this range was applied, resulting in a fixed 0.02% (2 basis points) slippage cost per trade.
- **Execution Model:** The simulation assumed that all orders were fully filled at the calculated price without failure.

Final Simulation Execution Parameters

For the final backtesting simulations on the hold-out data, a more comprehensive and conservative execution framework was implemented to better reflect live trading conditions:

- **Position Sizing:** A fixed fractional sizing strategy was employed, dedicating 10% of the total portfolio equity to each new position. This standardized, moderately aggressive allocation ensures consistent risk pressure across all trades and strategies, allowing for a fair and direct comparison of their performance under a uniform capital management rule.
- **Dynamic Trade Management:** All strategies integrate dynamic controls for exits. Both stop-loss and take-profit levels are calculated based on the Average True Range. This method allows exit points to adapt to current market volatility, rather than relying on arbitrary static percentages. For momentum strategies, trailing stops are also used to protect unrealized gains while allowing profitable trends to mature.

- **Transaction Cost Simulation:** To accurately model the erosion of returns from trading activity, two primary costs are deducted from the profit and loss of each transaction:
  - **Commission:** A fee of 0.01% of the trade's notional value is applied to simulate a competitive brokerage rate.
  - **Slippage:** A conservative slippage cost of 0.02% (2 basis points) is charged per trade to account for the bid-ask spread and minor price movements during order fulfillment.
- **Execution Uncertainty:** To account for the unpredictability of a live market, a probabilistic order failure model is introduced. Each order faces a 0.5% chance of rejection, simulating real-world events like fleeting liquidity or cancelled orders that prevent a trade from being executed as intended.

4. Results

In the end, the setup was only made of training and test subsets on the full range of data due to computational power and time restrictions. The data were processed, volume modelled, parameters optimized and strategies implemented and ran.

4.1 Simulation environment

The backtesting and evaluation of the intraday trading strategies was conducted on the last 3-months of available data for NASDAQ-100 stocks, simulating trades. The results provide insights into the performance of static and dynamic strategies, their robustness and effectiveness across market regimes. The simulation ran in a timeframe from November 17, 2024 to February 17, 2025, during market hours (9:30 AM to 4:00 PM ET), which means something of a cumulative timeperiod of 17-18 days of pure trading time.

To enhance the realism of this simulation, several market frictions were modeled using the `vectorbt` framework.

It is imperative to acknowledge the inherent simplifications of the backtesting environment. The model assumes the highly liquid stocks of the NASDAQ-100 have infinite liquidity at any given price point, meaning it does not simulate market impact, where a large order could adversely affect execution prices. Consequently, it cannot model partial fills that arise from liquidity gaps in the exchange's order book. The simulation also operates with zero latency, executing trades based on the data of the subsequent bar without accounting for the real-world time delay between signal generation, order routing, and final execution. While the included slippage parameter serves as a proxy for the bid-ask spread, this cost is modeled as a static percentage rather than a dynamic variable that changes with market volatility. Therefore, the presented results should be interpreted as a robust, yet still an estimation, a benchmark that validates the strategies' logic under these specified conditions.

4.1.1 Evaluation outcomes

On an aggregate level, the average results indicate that nearly all strategies, with the sole exception of the *Baseline Bollinger Bands* (+1.01%), failed to generate a positive total return and subsequently underperformed the 4.19% benchmark return on NASDAQ-100 at the time. This headline figure might suggest a failure of the more complex models. A deeper analysis provides a more insightful conclusion. The primary value of the volume-enhanced and deep learning layers was not in generating alpha on average, but in demonstrably improving the underlying baseline strategies, particularly in terms of risk mitigation and loss reduction. For instance, while the *Baseline Breakout* strategy incurred an average loss of -1.23% with a maximum drawdown of 2.14%, its *Volume-Enhanced* counterpart significantly reduced that loss to -0.51% and slashed the maximum drawdown to just 0.89%. A similar, and even more pronounced, improvement was observed with the *Momentum* strategy, where the baseline's -0.79% return and 2.07% drawdown were improved to -0.58% and 0.97% respectively by the volume-enhanced version, and further to -0.55% and 0.89% by the deep learning model. These metrics prove that the enhanced strategies, while not profitable on average, were successful in preserving capital and minimizing losses far more effectively than their simpler baselines. The narrative is further complicated by the "best fit" results, which underscore the critical importance of instrument-specific application. When applied to an optimal ticker at the right time, every single strategy became profitable. This starkly contrasts with the average results and validates the hypothesis that the subpar aggregate performance is more a reflection of a non-optimized, one-size-fits-all approach rather than an inherent flaw in the strategies themselves. The complex strategies are not robust enough to work universally without tuning, whereas the baseline models are less sensitive to parameter choices. For example, the *Deep Learning VWAP Reversion* strategy, which posted an average loss, achieved a strong 4.04% return on the Axon Enterprise Inc stock, demonstrating its potential when correctly deployed.

It is crucial to frame these outcomes within the context of their significant limitations. The volume forecasting models—the very engine of the enhanced strategies—were themselves highly unoptimized. They operated with a prediction error of 15-20%, used a static 15 minute forecast horizon without the use of multi-horizon capabilities, and lacked any form of meta-learning optimization or adjustment for news-driven market events. The fact that these strategies still managed to significantly curtail drawdowns and reduce losses, even when fed by a flawed forecasting engine, is a testament to potential. The results suggest that with proper, ticker-specific hyperparameter tuning and a more sophisticated, dynamically optimized volume forecasting model, the performance of these enhanced strategies could move from simply mitigating risk to actively generating alpha.

The average performance across all tickers is detailed in Tables 6.15, 6.16, and 6.17. The performance of the best-performing strategy-ticker combination for each strategy type is shown in Tables 6.18 and 6.19.

4.2 Regime bucket performance

To test the hypothesis that intraday strategy performance is regime-dependent, a detailed statistical analysis was performed. The methodology involved categorizing every individual trade executed across all backtests based on the prevailing market regime (e.g., volatility and trend) on its entry date. The core metric used for evaluation was the percentage return of each trade. To compare the performance between different regime buckets, a one-tailed independent samples Welch's t-test was conducted, chosen for its robustness to unequal sample sizes and variances. A significance level (α) of 0.05 was used to determine statistical significance. A.

The results provide nuanced support for the overarching hypothesis, with three of the five sub-hypotheses being validated.

Validated Hypotheses ($p < 0.05$)

H1c: Mean-reversion strategies perform better in low/medium volatility

Comparing 34,442 trades in low/medium volatility against 9,435 trades in high volatility, the test yielded a significant result ($p = 0.0053$). This confirms that, as expected, mean-reversion strategies are more profitable in less volatile conditions.

H1d: Mean-reversion strategies perform worse in uptrends

A comparison between 6,944 trades in uptrends and 32,277 trades in ranging markets was statistically significant ($p = 0.0223$), supporting the hypothesis that the performance of mean-reversion strategies degrades in strong directional trends.

H1e: Momentum strategies perform better in trending/high-volatility markets

The test between 705 trades in the trending/high-volatility bucket and 1,092 trades in the ranging/low-volatility bucket was also significant ($p = 0.0246$). This validates the classic expectation that momentum thrives on high volatility and clear direction.

Unsupported Hypotheses ($p \geq 0.05$)

H1a & H1b: Breakout strategies.

Conversely, hypotheses related to breakout strategies were not supported by the data in this sample. The test for improved performance in medium/high volatility (N=21,267) versus low volatility (N=4,650) was not significant ($p = 0.2448$). Similarly, the comparison between trending (N=6,837) and ranging (N=19,080) markets failed to show a significant advantage ($p = 0.7086$).

The concluded results about breakout strategies in this sample goes against norms. The results may signify frequent false breakouts in the included period.

4.3 Strategy comparison

To evaluate the Hypothesis 2, that incorporating volume-based parameters improves strategy performance, a paired comparative analysis was conducted. The methodology involved comparing the final performance metrics of each baseline strategy against its corresponding volume-enhanced version on a per-ticker basis. This pairing is essential as it controls for the unique characteristics of each instrument. For each of the over 100 tickers with complete data, a pair of metric outcomes was established. The statistical tool used was the one-tailed paired t-test, which is designed to determine if the mean difference between two sets of paired observations is significantly different from zero in a specific direction A). The analysis focused on three key performance indicators from the simulation summaries: *Sharpe Ratio* and *Profit Factor* were tested for a significant increase, while *Maximum Drawdown* was tested for a significant decrease. A significance level (α) of 0.05 was used as the threshold to accept or reject the null hypotheses.
The results for this hypothesis provide a clear and compelling, albeit nuanced, conclusion. The primary proposition, that volume-enhanced strategies would be more profitable was not supported by the data.

Profitability Analysis (H2): For all three strategy archetypes (*Breakout*, *Momentum*, and *Mean-Reversion*), the t-tests for both *Sharpe Ratio* and *Profit Factor* failed to demonstrate a statistically significant improvement ($p > 0.05$ in all six tests). The negative t-statistics observed in most of these cases suggest that adding volume confirmation tended to slightly decrease profitability, though not to a significant degree.

In stark contrast, the analysis of risk tells a different story. The hypothesis that volume-enhanced strategies would lead to better risk management was strongly and consistently supported.

Risk Analysis (H2): For all three strategy pairs, the volume-enhanced versions produced a statistically significant reduction in *Maximum Drawdown*. The comparison for *Baseline Breakout* vs. *Volume-Enhanced Breakout* (N=100 pairs) yielded a p-value of 4.38×10^{-18}. The *Momentum* comparison (N=101 pairs) resulted in a p-value of 7.58×10^{-18}, and the *Mean-Reversion* comparison (N=101 pairs) produced a p-value of 0.0007. These exceptionally low p-values ($p < 0.001$) indicate a high degree of confidence in the result.

This suggests that the primary value of volume confirmation lies not in amplifying gains, but in improving a strategy's defensive characteristics by avoiding trades with weaker conviction.

The final stage of the analysis was to evaluate Hypothesis 3, which posited that augmenting strategies with sophisticated, transformer-based volume forecasts would yield a further performance improvement over the simpler, heuristic-based Volume-Enhanced models. The same methodology was applied, this time comparing the Volume-Enhanced strategies against their Deep Learning counterparts across the same metrics and ticker cohort A).

The results of this analysis were unequivocal, the hypothesis that transformer-based models provide a superior signal generation framework was conclusively rejected. Across all nine statistical tests, none showed a statistically significant improvement ($p > 0.05$). For both profitability and risk, the Deep Learning strategies failed to outperform the simpler volume-enhanced versions. The inconsistent t-statistics and high p-values indicate that any observed differences in performance were likely attributable to random chance rather than a genuine modeling advantage.

This finding suggests that, within this study's framework, the added complexity and computational expense of employing a transformer model did not translate into tangible improvements in trading outcomes. A possible explanation for this outcome is the static nature of the model employed; a model trained on 3-6 months old data was used without periodic retraining. For a time-series model like this to be effective, it likely requires continual weight updates (daily or weekly) to adapt to the current market environment. This points to a important fact, that the retraining frequency should be increased in best conditions. The simpler, heuristic-based volume parameters were sufficient, and in this context, the advanced forecasting model provided no discernible edge.

5. Discussion

5.1 Strengths

Despite its limitations, the developed framework has several key strengths. The primary advantage lies in its successful implementation of a tiered strategy architecture, progressing from simple price-action rules to complex, deep learning-enhanced models across Big Data at 1 minute frequency. This structure provides a clear and methodical way to isolate and evaluate the incremental value of adding volume-based analytics and predictive forecasting. The results confirmed that these enhancements, even when imperfect, contributed positively to risk management.

The choice of the TFT model, while not the most recent architecture available, proved to be a pragmatic and effective decision. The model is well-documented, has a functional implementation, and demonstrated strong predictive performance (median R^2 of 0.74 in scaled space) on the complex task of 15-minute volume forecasting. This allowed the project to proceed with a reliable forecasting engine without the need to develop or debug a more experimental, state-of-the-art model from scratch, which might have been gated or proprietary.

Implementation of a regime classification system, though methodologically simple, provided a crucial analytical lens for interpreting performance. It enabled hypothesis testing that validated established market wisdom, such as the superior performance of mean-reversion strategies in low-volatility environments and momentum strategies in trending markets, grounding the study's findings in established financial theory.

5.2 Weaknesses

The framework's weaknesses, largely stemming from practical constraints, span data processing, model implementation, and optimization.

A foundational challenge originated from data quality. The dataset sourced from Polygon.IO contained missing records, which necessitated the use of linear interpolation for some trading-hours volume. This method, while practical, is a simplification that may not perfectly represent true market activity. Furthermore, the inherent inconsistencies in how exchanges report volume introduce a layer of noise that no amount of cleaning can fully eliminate.

The validation process was also constrained. The project did not include a true 6-month incubation period of demo or live trading. The planned incubation phase was halted due to highly volatile, news-driven market conditions, for which the strategies were not designed. This highlights a significant limitation: the absence of news and sentiment analysis in the models.

The volume forecasting model itself, while functional, was not fully optimized. It was not retrained in a rolling fashion throughout the full backtest period, which would be essential in a live environment to adapt to evolving market dynamics.

Implementing true rolling predictions, even when batched, proved to be computationally prohibitive with the TFT architecture for available compute. Moreover, the model could have been improved through several avenues that were not explored:

- **Multi-horizon forecasting:** The model was trained only for a static 15-minute horizon, while capable for multi horizon predictions to the future or to the past (for example dynamic windows), some tickers showed potential for better accuracy at shorter or longer horizons.
- **Feature Engineering:** Additional relevant features could have been engineered and included in the TFT model.
- **Advanced Techniques:** The framework did not employ meta-learning to generalize across tickers, nor did it use corrective regressors (e.g., Isotonic or XGBoost models on residuals) to systematically scale and improve volume predictions.

The broader strategic and portfolio-level approach was simplified. Strategy parameters were generalized by averaging optimal values across all securities rather than using ticker-specific optimized sets. The regime classification relied on the simplest available methods rather than more advanced clustering or statistical techniques. At the portfolio level, there was no optimization of leverage or capital allocation, which would be a critical component of a real-world trading system.

5.3 Market conditions and regime shifts

A central theme of this thesis is the profound impact of changing market conditions on strategy performance. The financial markets are not static; they cycle through distinct regimes of volatility, trend, and liquidity. A strategy that is profitable in one environment may perform poorly in another.

The regime classification system was implemented specifically to address this challenge by attempting to categorize the market environment and test strategy performance accordingly. The results from this analysis were insightful, confirming that mean-reversion and momentum strategies are indeed highly sensitive to the prevailing regime. However, the analysis failed to find a statistically significant performance advantage for breakout strategies in their hypothesized optimal environments (trending, high-volatility markets), possibly due to a prevalence of false breakouts during the simulation period. The non-robust nature of the classification model itself is a key limitation. Using more advanced methods like clustering or hidden Markov models could provide a more accurate and dynamic segmentation of market states. More importantly, the experience during the aborted incubation phase underscores the model's vulnerability to exogenous shocks. The strategies, blind to news events, were ill-equipped to handle the sudden, sentiment-driven volatility. This demonstrates that for a trading framework to be truly robust, it must either incorporate data sources that capture such events or include risk management systems that can dynamically deactivate strategies during periods of extreme, un-modelled uncertainty, adapting to these shifts remains a primary challenge for any trading system.

Conclusion

The thesis explored the domain of intraday trading, systematically evaluating the incremental benefit of integrating volume-based analytics and predictive deep learning into classical trading strategies. By designing a tiered framework of baseline, volume-enhanced, and deep learning-enhanced strategies, this study sought to quantify the value of increasing analytical complexity. The core investigation was anchored by three hypotheses concerning the influence of market regimes, the role of volume as a confirmation signal, and the predictive power of a transformer forecasting model.

The empirical results yielded a nuanced narrative. The primary finding was that while incorporating volume-based parameters did not lead to a statistically significant increase in profitability, it offered a powerful advantage in risk management. The volume-enhanced strategies consistently and significantly reduced maximum drawdowns across all strategy archetypes, confirming the hypothesis that volume serves as an effective filter for avoiding signals with low conviction. This underscores a crucial trade-off: the added complexity serves a defensive purpose, preserving capital more effectively than simpler price-only models.

Conversely, the hypothesis that a sophisticated transformer-based volume forecast would further enhance performance was conclusively rejected. The predictive strategies failed to outperform their simpler, volume-enhanced counterparts, suggesting that within the constraints of this study, the added computational complexity did not translate into a tangible trading edge. Alongside this, the research validated the long-held market wisdom that strategy performance is regime-dependent, with mean-reversion and momentum strategies showing statistically significant performance shifts between volatile, trending, and ranging conditions.

It is crucial to frame these conclusions within the context of the work's limitations. The study was constrained by data-quality issues, which required interpolation, and lacked computational resources, which limited the backtesting period and precluded the implementation of a dynamic, rolling-retraining scheme for the volume forecasting model. The model's static nature, trained only once on historical data, is a significant departure from what would be required in a live environment. The framework's vulnerability to exogenous, news-driven events was highlighted during the process.

The most immediate path for improvement involves enhancing the forecasting engine by implementing rolling-window retraining, exploring multi-horizon predictions, and integrating meta-learning to generalize across securities. A second crucial direction is the incorporation of multi-modal data, particularly news and sentiment analysis, to build strategies that are more resilient to market shocks.

An optimized framework could explore more sophisticated regime detection models, such as Hidden Markov Models, and implement portfolio-level optimizations to translate strategic signals into effective capital allocation. These enhancements would move the framework closer to a commercially viable and adaptive trading system.

6. Figures & Tables

6.1 Tables

This section outlines the schema for the various dataframes used in the analysis and the parameters for the trading strategies.

6.1.1 Data Schemas

Table 6.1

Schema of Preprocessed Final News Dataframe

Column Name	Data Type	Description
event_type	string	Type of news event (e.g., "Earnings Report", "Acquisition")
date	date32[day]	Date of the news event (e.g., "2020-03-06")
time	string	Time of the news event (e.g., "00:00:00", "08:00:00")
ticker	string	Stock ticker associated with the news event
impact_date	date32[day]	Date when the news is expected to impact the stock

Table 6.2

Schema of Preprocessed Regimes Dataframe

Column Name	Data Type	Description
ticker	string	Stock ticker (e.g., "ABNB")
date	date32[day]	Date of the regime data (day-level granularity)
volatility_regime	string	Volatility regime label (encoded or "undefined")
trend_regime	string	Trend regime label (encoded or "undefined")
liquidity_regime	string	Liquidity regime label (encoded or "undefined")
news_impact	int8	News impact score (0 for no impact, 1 for impact)

Table 6.3

Schema of Preprocessed Backtesting Dataset

Column Name	Data Type	Description
ticker	string	Stock ticker
timestamp	timestamp[ns]	Unified timestamp combining Date and Time
open	float32	Opening price for the 1-minute interval
high	float32	Highest price for the 1-minute interval
low	float32	Lowest price for the 1-minute interval
close	float32	Closing price for the 1-minute interval
volume	float32	Trading volume for the 1-minute interval
prev_session_high	float32	Previous session's highest price
prev_session_low	float32	Previous session's lowest price
estimated_bid_ask_spread	float32	Estimated bid-ask spread
estimated_obd	float32	Estimated order book depth
50_day_sma	float32	50-day simple moving average

Table 6.4

Schema of Preprocessed AAPL Data for NeuralForecast

Column Name	Data Type	Description
ticker	string	Stock ticker ("AAPL")
timestamp	timestamp[ns]	Timestamp of the 1-minute interval
open	float64	Normalized opening price (z-score)
high	float64	Normalized highest price (z-score)
low	float64	Normalized lowest price (z-score)
close	float64	Normalized closing price (z-score)
volume	float64	Normalized trading volume (z-score)
prev_session_high	float32	Normalized previous session high price (z-score)
prev_session_low	float32	Normalized previous session low price (z-score)
estimated_bid_ask_spread	float32	Normalized estimated bid-ask spread (z-score)
estimated_obd	float32	Normalized estimated order book depth (z-score)
50_day_sma	float32	Normalized 50-day simple moving average (z-score)
news_impact	int8	News impact score from standardized data
hour	int64	Hour of the day (9–16)
day_of_week	int64	Day of the week (0=Monday, 6=Sunday)
minute	int64	Minute of the hour (0–59)
time_since_open	float64	Minutes since market open (9:30 AM ET)
is_trading	int64	Binary indicator (1 for trading hours)
volatility_regime	int64	Label-encoded volatility regime
trend_regime	int64	Label-encoded trend regime
liquidity_regime	int64	Label-encoded liquidity regime
news_impact_regime	int8	News impact score from regimes data
volume_lag_1 to volume_lag_15	float64	Normalized lagged volume (t-1 to t-15, z-score)
close_lag_1 to close_lag_15	float64	Normalized lagged closing price (t-1 to t-15, z-score)
log_volume	float64	Log-transformed and scaled volume (target variable)
returns	float64	Percentage change in closing price

6.1.2 Strategy Thresholds

Baseline Strategy Thresholds

Table 6.5

Breakout Strategy Parameters

Parameter	Description	Optimized Value
δ_{sl}	Breakout Stop-Loss	0.011
δ_{tp}	Breakout Take-Profit	0.010

Table 6.6

Momentum Strategy Parameters

Parameter	Description	Optimized Value
τ_{mom}	Momentum Buy/Sell Consecutive Bars	6
$\delta_{\text{mom\_trail}}$	Momentum Trailing Stop-Loss	0.024

Table 6.7

Bollinger Bands Strategy Parameters

Parameter	Description	Optimized Value
τ_{bb}	Bollinger Bands Period	25
γ_{bb}	Bollinger Bands Standard Deviations	2.5

Volume-Enhanced Strategy Thresholds

Table 6.8

Volume Breakout Strategy Parameters

Parameter	Description	Optimized Value
ϕ_{va}	Value Area Coverage for VAH/VAL	0.80
κ_{surge}	Volume Surge Multiplier	3.5
τ_{adx}	ADX Period	20
θ_{adx}	ADX Threshold	20
α_{sl}	Volume Breakout Stop-Loss ATR Multiplier	3.0
α_{tp}	Volume Breakout Take-Profit ATR Multiplier (SL $\times$ $\mathcal{R}$)	6.0
$\mathcal{R}$	Reward-to-Risk Ratio	2.00

Table 6.9

Volume VWAP Reversion Strategy Parameters

Parameter	Description	Optimized Value
τ_{obv}	OBV Threshold Period	40
θ_{obv}	OBV Threshold Percentile	0.9
ϵ_{slope}	VWAP Slope Threshold	0.0001
$\tau_{\text{vwap\_trend}}$	VWAP Trend Window	19
α_{sl}	VWAP Stop-Loss ATR Multiplier	3.2
α_{tp}	VWAP Take-Profit ATR Multiplier (SL $\times$ $\mathcal{R}$)	6.0
$\mathcal{R}$	Reward-to-Risk Ratio	1.88

Table 6.10

Volume Momentum Strategy Parameters

Parameter	Description	Optimized Value
τ_{adx}	ADX Period	17
$\kappa_{\text{vol\_mom}}$	Volume Momentum Relative Volume Threshold	2.8
θ_{adx}	ADX Threshold	33
α_{sl}	Volume Momentum Stop-Loss ATR Multiplier	3.3

Deep Learning-Enhanced Strategy Thresholds

Table 6.11
DL Breakout Strategy Parameters

Parameter	Description	Optimized Value
ϕ_{va}	VAH/VAL Adjustment Factor	0.75
κ_{dl}	Deep Learning Volume Multiplier	1.6
τ_{adx}	ADX Period	16
θ_{adx}	ADX Threshold	20
α_{sl}	DL Stop-Loss ATR Multiplier	3.0
α_{tp}	DL Take-Profit ATR Multiplier (SL $\times$ $\mathcal{R}$)	4.0
$\mathcal{R}$	Reward-to-Risk Ratio	1.33

Table 6.12
DL Volume Momentum Strategy Parameters

Parameter	Description	Optimized Value
τ_{adx}	ADX Period	20
κ_{dl}	Deep Learning Volume Multiplier	1.5
θ_{adx}	ADX Threshold	30
α_{sl}	DL Volume Momentum Stop-Loss ATR Multiplier	4.0
$\tau_{\text{vol\_trend}}$	DL Volume Trend Window	7

Table 6.13
DL VWAP Reversion Strategy Parameters

Parameter	Description	Optimized Value
δ_{vwap}	VWAP Buffer	0.003
$\tau_{\text{vwap\_trend}}$	VWAP Trend Window	18
κ_{dl}	Volume Multiplier	1.5
α_{sl}	DL VWAP Stop-Loss ATR Multiplier	3.0
α_{tp}	DL VWAP Take-Profit ATR Multiplier (SL $\times$ $\mathcal{R}$)	7.0
$\mathcal{R}$	Reward-to-Risk Ratio	2.33

6.1.3 Optimization Results

The table below presents the key performance metrics achieved during the optimization phase for each trading strategy. These results reflect the average performance across all tickers used in the optimization process. The metrics include risk-adjusted return measures such as the Calmar, Sharpe, and Sortino Ratios, as well as the Kelly Criterion and the average number of trades executed. This provides a baseline understanding of each strategy's potential before backtesting on the full dataset.

For brevity, the following abbreviations are used for the trading strategies in the tables below:

- **BBB**: Baseline Bollinger Bands
- **BB**: Baseline Breakout
- **BM**: Baseline Momentum
- **VEB**: Volume-Enhanced Breakout
- **VEM**: Volume-Enhanced Momentum
- **VEV**: Volume-Enhanced VWAP Reversion
- **DLB**: Deep Learning Breakout
- **DLM**: Deep Learning Momentum
- **DLV**: Deep Learning VWAP Reversion

Table 6.14

Optimization Results

Strategy	Calmar	Sharpe	Sortino	Kelly	Trades
BBB	129.65	9.17	18.04	-1.06	154.35
BB	46.64	4.59	7.50	-1.47	81.34
BM	67.38	6.09	9.63	-0.43	18.76
VEB	271.07	7.71	33.77	-0.40	42.74
VEM	94.44	6.71	13.55	-0.35	25.40
VEV	157.94	9.95	20.73	-0.45	35.83
DLB	135.06	7.59	17.01	-0.37	60.66
DLM	105.64	6.09	16.12	-0.34	21.68
DLV	66.45	6.37	11.18	-0.98	164.75

6.1.4 Simulation Results

The simulation results are presented in two parts. First, we examine the average performance of each strategy across all tickers. Second, we highlight the key performance and risk metrics for the single best-performing strategy-ticker combination.

Average Strategy Performance

The following tables summarize the performance, trading characteristics, and risk-adjusted returns averaged across all simulations for each strategy.

Table 6.15

Key Performance Metrics (Average Results)

Strategy	Total Return [%]	Benchmark Return [%]	Max DD [%]	Profit Factor
BBB	1.01	4.19	1.01	1.34
BB	−1.23	4.19	2.14	0.79
BM	−0.79	4.19	2.07	0.80
VEB	−0.51	4.19	0.89	0.00
VEM	−0.58	4.19	0.97	0.65
VEV	−0.12	4.19	0.81	1.06
DLB	−1.10	4.19	1.57	0.78
DLM	−0.55	4.19	0.89	0.65
DLV	−0.09	4.19	1.40	0.99

Table 6.16

Trading Behavior (Average Results)

Strategy	Win Rate [%]	Total Trades	Avg Win [%]	Avg Loss [%]
BBB	61.13	257.51	0.37	−0.47
BB	46.88	71.55	1.25	−1.47
BM	33.15	23.14	2.74	−1.94
VEB	35.87	47.59	0.79	−0.60
VEM	28.87	37.19	0.77	−0.55
VEV	37.66	21.97	1.45	−0.94
DLB	41.99	137.46	0.53	−0.52
DLM	25.84	29.16	0.80	−0.56
DLV	49.40	154.94	0.69	−0.70

Table 6.17

Risk-Adjusted Ratios (Average Results)

Strategy	**Sharpe**	**Sortino**	**Calmar**
BBB	5.06	8.92	94.70
BB	−5.60	−7.02	−9.46
BM	−2.89	−3.65	−2.89
VEB	0.00	0.00	−8.35
VEM	−5.88	−7.08	−8.43
VEV	−1.16	0.04	13.38
DLB	−7.60	−9.31	−5.70
DLM	−5.66	−6.76	−8.60
DLV	−0.83	−0.33	15.58

Best-Fit Individual Performance

The following tables deconstruct the performance of the single best-fit strategy for a specific ticker, focusing on the most relevant performance and risk metrics.

Table 6.18
Best Fit Performance Summary

Strategy	Ticker	Total Ret. [%]	Bench. Ret. [%]	Win Rate [%]	Profit Factor
BBB	ON	5.38	−25.25	66.76	1.89
BB	NFLX	2.23	24.71	56.86	1.78
BM	SBUX	2.46	13.62	62.50	4.50
VEB	LULU	0.59	19.89	56.52	1.35
VEM	MELI	1.25	8.57	57.89	4.44
VEV	DASH	1.79	16.22	55.00	3.44
DLB	DXCM	0.94	12.50	52.43	1.33
DLM	INTU	0.43	−14.65	57.14	2.31
DLV	AXON	4.04	9.82	56.34	1.84

Table 6.19
Best Fit Risk & Drawdown

Strategy	Ticker	Max DD [%]	Sharpe	Sortino	Calmar
BBB	ON	0.55	15.92	27.66	488.26
BB	NFLX	1.09	6.03	17.12	66.04
BM	SBUX	0.82	9.33	13.21	94.20
VEB	LULU	0.62	4.12	5.34	33.92
VEM	MELI	0.21	14.38	26.62	543.45
VEV	DASH	0.38	11.08	31.97	205.80
DLB	DXCM	0.71	4.94	7.38	54.34
DLM	INTU	0.27	7.76	11.37	87.77
DLV	AXON	0.60	14.88	28.87	518.66

References

[1] Irene Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Wiley, 2013.

[2] Abdul Fatir Ansari et al. *Chronos: A Foundation Model for Time Series Forecasting*. 2024. arXiv: `2403.07815`.

[3] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels and Ruben Gryp. 'Evaluating multiple classifiers for stock price direction prediction'. In: *Expert Systems with Applications* (2015).

[4] Bank of England. *News and Events*. `https://www.bankofengland.co.uk/news/events`. 2025.

[5] Bank of Japan. *Monetary Policy Meetings*. `https://www.boj.or.jp/en/mopo/mpmsche_minu/index.htm`. 2025.

[6] Fischer Black and Myron Scholes. 'The pricing of options and corporate liabilities'. In: *Journal of Political Economy* (1973).

[7] Bloomberg. *U.S. Market Coverage*. Accessed: 2025-06-24. 2025. URL: `https://www.bloomberg.com/markets`.

[8] Tim Bollerslev. 'Generalized autoregressive conditional heteroskedasticity'. In: *Journal of Econometrics* (1986).

[9] John Bollinger. *Bollinger on Bollinger Bands*. McGraw-Hill, 2001.

[10] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.

[11] Leo Breiman. 'Random forests'. In: *Machine Learning* (2001).

[12] Robert G. Brown. *Statistical Forecasting for Inventory Control*. McGraw-Hill, 1959.

[13] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever and Dario Amodei. 'Language Models are Few-Shot Learners'. In: *arXiv preprint arXiv:2005.14165* (2020). arXiv: `2005.14165 [cs.CL]`.

[14] F. Cecconi, ed. *AI in the Financial Markets: New Algorithms and Solutions*. Computational Social Sciences. Springer, 2023.

[15] CFA Institute. *CFA Program Curriculum, Level II, Volume 4: Equity Valuation, Fixed Income*. CFA Institute, 2024. ISBN: 978-1-953337-58-0.

[16] Ernest P. Chan. *Advances in financial machine learning*. Wiley, 2017.

[17] Ernest P. Chan. *Algorithmic trading: Winning strategies and their rationale*. Wiley, 2013.

[18] Ernest P. Chan. *Quantitative trading: How to build your own algorithmic trading business*. Wiley, 2008.

[19] Ruibo Chen, Wei Li, Zhiyuan Zhang, Ruihan Bao, Keiko Harimoto and Xu Sun. 'Stock Trading Volume Prediction with Dual-Process Meta-Learning'. In: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD)*. 2022. DOI: 10.48550/arXiv.2211.01762. arXiv: 2211.01762 [q-fin.TR].

[20] Tianqi Chen and Carlos Guestrin. 'XGBoost: A scalable tree boosting system'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.

[21] John H. Cochrane. *Asset Pricing*. Revised. Princeton University Press, 2005.

[22] Rama Cont. 'Empirical properties of asset returns: stylized facts and statistical issues'. In: *Quantitative Finance* 1.2 (2001), pp. 223–236.

[23] Paul H. Cootner, ed. *The Random Character of Stock Market Prices*. MIT Press, 1964.

[24] James Dalton. *Mind Over Markets: Power Trading with Market Generated Information*. Wiley, 2012.

[25] James Dalton, Robert Dalton and Eric Jones. *Markets in Profile: Profiting from the Auction Process*. Wiley, 2007.

[26] Kevin J. Davey. *Building Winning Algorithmic Trading Systems*. Wiley Trading Series. Wiley, 2014.

[27] Kevin J. Davey. *Building winning algorithmic trading systems: A trader's journey from data mining to Monte Carlo simulation to live trading*. Wiley, 2014.

[28] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019.

[29] Matthew F. Dixon, Igor Halperin and Paul Bilokon. *Machine Learning in Finance: From Theory to Practice*. Springer, 2020.

[30] DoltHub. *Earnings Calendar Database*. https://www.dolthub.com/repositories/post-no-preference/earnings/data/master/earnings_calendar. 2025.

[31] Louis H. Ederington and Jae Ha Lee. 'How Markets Process Information: News Releases and Volatility'. In: *The Journal of Finance* (1993).

[32] Robert F. Engle. 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation'. In: *Econometrica* (1982).

[33] Robert F. Engle and Clive W. J. Granger. 'Co-integration and error correction: Representation, estimation, and testing'. In: *Econometrica* (1987).

[34] European Central Bank. *Monetary Policy Decisions*. https://www.ecb.europa.eu/press/press_conference/visual-mps/2025/html/mopo_statement_explained_april.en.html. 2025.

[35] Eugene F. Fama. 'Efficient capital markets: A review of theory and empirical work'. In: *Journal of Finance* (1970).

[36] Eugene F. Fama and Kenneth R. French. 'The cross-section of expected stock returns'. In: *Journal of Finance* (1992).

[37] Eugene F. Fama and James D. MacBeth. 'Risk, return, and equilibrium: Empirical tests'. In: *Journal of Political Economy* (1973).

[38] Federal Reserve Bank of St. Louis. *Federal Reserve Economic Data (FRED) API*. https://fred.stlouisfed.org/docs/api/fred/. 2025.

[39] Elizabeth Fons, Rachneet Kaur, Soham Palande, Zhen Zeng, Tucker Balch, Manuela Veloso and Svitlana Vyetrenko. 'Evaluating Large Language Models on Time Series Feature Understanding: A Comprehensive Taxonomy and Benchmark'. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2024. DOI: 10.48550/arXiv.2404.16563. arXiv: 2404.16563 [cs.CL].

[40] A. Garza and M. Mergenthaler-Canela. *TimeGPT-1*. 2023.

[41] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

[42] Anubha Goel, Puneet Pasricha and Juho Kanniainen. *Time-Series Foundation AI Model for Value-at-Risk Forecasting*. 2024. DOI: 10.48550/arXiv.2410.11773. arXiv: 2410.11773 [q-fin.RM].

[43] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[44] Ian Goodfellow, Yoshua Bengio and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[45] Google. *Google Cloud Vertex AI, Gemini Pro Model*. https://cloud.google.com/vertex-ai. 2025.

[46] Joseph E. Granville. *Granville's New Key to Stock Market Profits*. Prentice-Hall, 1963.

[47] Xing Guo, Tze Leung Lai, Howard Shek and Samuel Po-Shing Wong. *Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization*. CRC Press, Taylor & Francis Group, 2017.

[48] Larry Harris. *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press, 2003.

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

[50] Trevor Hastie, Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.

[51] Yves Hilpisch. *Artificial Intelligence in Finance: A Python-Based Guide*. O'Reilly Media, 2020.

[52] Yves Hilpisch. *Python for Algorithmic Trading: From Idea to Cloud*. O'Reilly Media, 2020.

[53] Yves Hilpisch. *Python for Finance: Mastering Data-Driven Finance*. O'Reilly Media, 2018.

[54] Sepp Hochreiter and Jürgen Schmidhuber. 'Long short-term memory'. In: *Neural Computation* (1997).

[55] John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

[56] John C. Hull. *Options, Futures, and Other Derivatives*. 9th ed. Pearson, 2014.

[57] Stefan Jansen. *Hands-On Machine Learning for Algorithmic Trading*. Packt Publishing, 2018.

[58] Stefan Jansen. *Machine Learning for Algorithmic Trading: Predictive Models*. Packt Publishing, 2020.

[59] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan and Qingsong Wen. 'Time-LLM: Time Series Forecasting by Reprogramming Large Language Models'. In: *The Twelfth International Conference on Learning Representations (ICLR)*. 2024. DOI: `10.48550/arXiv.2310.01728`. arXiv: `2310.01728 [cs.LG]`.

[60] Barry Johnson. *Algorithmic trading & DMA: An introduction to direct access trading strategies*. 4Myeloma Press, 2010.

[61] Perry J. Kaufman. *Trading systems and methods*. 5th ed. Wiley, 2013.

[62] Colm Kelliher. *Quantitative Finance with Python: A Practical Guide*. Taylor & Francis Group, 2022.

[63] J. Keppler. *Profit with the Market Profile: Identifying Market Value in Real Time*. Marketplace Books, 2011.

[64] J. H. Kim, A. Shamsuddin and K.-P. Lim. 'Stock return predictability and the Adaptive Markets Hypothesis: Evidence from century-long U.S. data'. In: *Journal of Empirical Finance* (2011).

[65] Paul Klemperer. *Auctions: Theory and Practice*. Princeton University Press, 2004.

[66] Litton Jose Kurisinkel, Pruthwik Mishra and Yue Zhang. *Text2TimeSeries: Enhancing Financial Forecasting through Time Series Prediction Updates with Event-Driven Insights from Large Language Models*. 2024. DOI: `10.48550/arXiv.2407.03689`. arXiv: `2407.03689 [cs.CL]`.

[67] Alex Li. *Volatility Forecasting in Global Financial Markets Using TimeMixer*. 2024. DOI: 10.48550/arXiv.2410.09062. arXiv: 2410.09062 [q-fin.ST].

[68] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang and Xifeng Yan. 'Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019. DOI: 10.48550/arXiv.1907.00235. arXiv: 1907.00235 [cs.LG].

[69] Bryan Lim, S. Ö. Arık, Nicolas Loeff and Tomas Pfister. 'Temporal fusion transformers for interpretable multi-horizon time series forecasting'. In: *International Journal of Forecasting* (2021).

[70] Jialin Liu et al. 'Transformer-based Capsule Network for Stock Movement Prediction'. In: *Proceedings of FinNLP*. 2019.

[71] Xu Liu, Juncheng Liu, Gerald Woo, Taha Aksu, Yuxuan Liang, Roger Zimmermann, Chenghao Liu, Silvio Savarese, Caiming Xiong and Doyen Sahoo. *Moirai-MoE: Empowering Time Series Foundation Models with Sparse Mixture of Experts*. 2024. DOI: 10.48550/arXiv.2410.10469. arXiv: 2410.10469 [cs.LG].

[72] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma and Mingsheng Long. 'iTransformer: Inverted Transformers Are Effective for Time Series Forecasting'. In: *The Twelfth International Conference on Learning Representations (ICLR)*. 2024. DOI: 10.48550/arXiv.2310.06625. arXiv: 2310.06625 [cs.LG].

[73] Yong Liu, Haixu Wu, Jianmin Wang and Mingsheng Long. 'Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting'. In: *Advances in Neural Information Processing Systems*. 2022.

[74] Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang and Mingsheng Long. *Timer: Transformers for Time Series Analysis at Scale*. 2024. DOI: 10.48550/arXiv.2402.02368. arXiv: 2402.02368 [cs.LG].

[75] Andrew W. Lo. 'The Adaptive Markets Hypothesis: Market efficiency from an evolutionary perspective'. In: *Journal of Portfolio Management* (2004).

[76] Marcos Lopez de Prado. *Advances in financial machine learning*. Wiley, 2018.

[77] Burton G. Malkiel. *A Random Walk Down Wall Street*. W.W. Norton & Company, 1973.

[78] Albert J. Menkveld. 'The economics of high-frequency trading: Taking stock'. In: *Annual Review of Financial Economics* (2016). DOI: 10.1146/annurev-financial-121415-033010.

[79] John J. Murphy. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. New York Institute of Finance, 1999.

[80] Rishi K. Narang. *Inside the black box: The simple truth about quantitative trading*. Wiley, 2009.

[81] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong and Jay Kalchbrenner. 'A Time Series is Worth 64 Words: Long-term Forecasting with Transformers'. In: *The Eleventh International Conference on Learning Representations (ICLR)*. 2023.

[82] Maureen O'Hara. 'High frequency market microstructure'. In: *Journal of Financial Economics* (2015). DOI: 10.1016/j.jfineco.2015.01.003.

[83] Polygon.io. *Polygon.io Stock Market Data API*. https://polygon.io. 2025.

[84] Lawrence R. Rabiner. 'A tutorial on hidden Markov models and selected applications in speech recognition'. In: *Proceedings of the IEEE* (1989).

[85] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[86] Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, Marin Biloš, Sahil Garg, Anderson Schneider, Nicolas Chapados, Alexandre Drouin, Valentina Zantedeschi, Yuriy Nevmyvaka and Irina Rish. *Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting*. 2023. DOI: 10.48550/arXiv.2310.08278. arXiv: 2310.08278 [cs.LG].

[87] S&P Global. *S&P Dow Jones Indices | News & Announcements*. https://www.spglobal.com/spdji/en/media-center/news-announcements/. 2025.

[88] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner and Gabriele Monfardini. 'The graph neural network model'. In: *IEEE Transactions on Neural Networks* (2009).

[89] Jack D. Schwager. *The New Market Wizards: Conversations with America's Top Traders*. HarperBusiness, 1992.

[90] Steven L. Scott and Hal R. Varian. 'Predicting the present with Bayesian structural time series'. In: *International Journal of Mathematical Modelling and Numerical Optimisation* (2014).

[91] SEC API IO. *SEC Filings API*. https://sec-api.io/. 2025.

[92] William F. Sharpe. 'Capital asset prices: A theory of market equilibrium under conditions of risk'. In: *Journal of Finance* (1964).

[93] J. Peter Steidlmayer and Kevin Hawkins. *Steidlmayer on Markets: Trading with Market Profile*. 2nd ed. Wiley, 2003.

[94] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[95] The Federal Reserve Board. *FOMC Calendars*. https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm. 2025.

[96] The Federal Reserve Board. *Speeches and Statements*. https://www.federalreserve.gov/newsevents/speeches.htm. 2025.

[97] Ruey S. Tsay. *Analysis of Financial Time Series*. 3rd ed. Wiley, 2010.

[98] U.S. Securities and Exchange Commission. *Search for Company Filings*. https://www.sec.gov/search-filings. Accessed: 2025-06-26. 2025.

[99] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin. 'Attention is all you need'. In: *Advances in Neural Information Processing Systems*. 2017.

[100] Wikipedia. *List of elections in the United States*. https://en.wikipedia.org/wiki/List_of_elections_in_the_United_States. 2025.

[101] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese and Doyen Sahoo. *Unified Training of Universal Time Series Forecasting Transformers*. 2024. DOI: 10.48550/arXiv.2402.02592. arXiv: 2402.02592 [cs.LG].

[102] Haixu Wu, Tengge Chen, Yong Wang and Mingsheng Long. 'TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis'. In: *The Eleventh International Conference on Learning Representations (ICLR)*. 2023.

[103] Haixu Wu, Jiehui Xu, Jianmin Wang and Mingsheng Long. 'Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting'. In: *Advances in Neural Information Processing Systems*. 2021.

[104] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang and Philip S. Yu. 'A comprehensive survey on graph neural networks'. In: *IEEE Transactions on Neural Networks and Learning Systems* (2020).

[105] Jingjing Xu, Caesar Wu, Yuan-Fang Li and Pascal Bouvry. *Transformer Multivariate Forecasting: Less is More?* 2023. DOI: 10.48550/arXiv.2401.00230. arXiv: 2401.00230 [cs.LG].

[106] Jingyun Yang, Pan Li, Yiwen Cui, Xu Han and Mengjie Zhou. 'Multi-Sensor Temporal Fusion Transformer for Stock Performance Prediction: An Adaptive Sharpe Ratio Approach'. In: *Sensors* 25.3 (2025), p. 976. DOI: 10.3390/s25030976.

[107] Silin Yang, Dong Wang, Haoqi Zheng and Ruochun Jin. *TimeRAG: Boosting LLM Time Series Forecasting via Retrieval-Augmented Generation*. 2024. arXiv: 2412.16643 [cs.AI].

[108] Yongbo Yu, Weizhong Yu, Feiping Nie and Xuelong Li. *PRformer: Pyramidal Recurrent Transformer for Multivariate Time Series Forecasting*. 2024. DOI: 10.48550/arXiv.2408.10483. arXiv: 2408.10483 [cs.LG].

[109] Kamil Żbikowski. 'Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy'. In: *Expert Systems with Applications* (2015).

[110] Ailing Zeng et al. *Long-term Forecasting with TiDE: Time-series Dense Encoder*. 2023. arXiv: 2304.08424.

[111] Yunhao Zhang and Junchi Yan. 'Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting'. In: *The Eleventh International Conference on Learning Representations (ICLR)*. 2023. eprint: 2212.09184. URL: https://openreview.net/forum?id=v_iW0s9t5K.

[112] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong and Wancai Zhang. 'Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting'. In: *AAAI*. 2021.

[113] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun and Rong Jin. 'FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting'. In: *Proceedings of the 39th International Conference on Machine Learning (ICML)*. Proceedings of Machine Learning Research. 2022.

Appendices

A. Data processing and Hypothesis testing

1. Wilson Score Interval for Event Accuracy

Calculates the confidence interval for the estimated accuracy of event predictions using the Wilson score method.

$$\text{Lower Bound} = \frac{\hat{p} + \frac{z^2}{2n} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$$

$$\text{Upper Bound} = \frac{\hat{p} + \frac{z^2}{2n} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$$

Where:

- $\hat{p}$: Estimated accuracy (e.g., 0.95).
- n: Sample size (e.g., 100).
- z: Z-score for confidence level (e.g., 1.96 for 95%).

Note: Returns NaN if n = 0 or data missing.

2. Previous Session High and Low

Captures the high and low prices from the previous session for price pattern analysis.

$$D[\text{Prev High}]_i = \text{High}_{i-1}, \quad D[\text{Prev Low}]_i = \text{Low}_{i-1}$$

Where:

- *D[Prev High]*$_i$: Previous session high for observation i.
- *D[Prev Low]*$_i$: Previous session low for observation i.
- $High_{i-1}$: High price of previous observation.
- Low_{i-1}: Low price of previous observation.

3. Estimated Bid-Ask Spread

Approximates the bid-ask spread using the daily trading range with a minimum threshold.

$$\text{BAS}_i = \max(\text{High}_i - \text{Low}_i, 0.0001), \quad D[\text{BAS}]_i = \text{BAS}_i$$

Where:

- BAS_i: Estimated bid-ask spread for observation i.
- $High_i$: High price for observation i.
- Low_i: Low price for observation i.
- *D[BAS]*$_i$: Bid-ask spread added to dataset.

4. Order Book Depth

Estimates liquidity by calculating the order book depth based on trading volume and bid-ask spread.

$$\text{OBD}_i = \frac{\text{Volume}_i}{\text{BAS}_i}, \quad D[\text{OBD}]_i = \text{OBD}_i$$

Where:

- OBD_i: Order book depth for observation i.
- $Volume_i$: Trading volume for observation i.
- BAS_i: Estimated bid-ask spread for observation i.
- $D[OBD]_i$: Order book depth added to dataset.

5. 50-Day Simple Moving Average

Computes the 50-day simple moving average of closing prices as a trend indicator.

$$\text{SMA}_{(50,i)} = \frac{1}{50} \sum_{j=i-49}^{i} \text{Close}_j, \quad D[\text{SMA}_{50}]_i = \text{SMA}_{(50,i)}$$

Where:

- $SMA_{(50,i)}$: 50-day simple moving average for observation i.
- $Close_j$: Closing price for observation j.
- $D[\text{SMA}_{50}]_i$: 50-day SMA added to dataset.

6. Welch's t-test

The Welch's t-test is performed to compare the means of two independent groups when their variances are assumed to be unequal. It is defined by two formulas: one for the t-statistic and another for calculating the degrees of freedom (ν).

1. t-statistic:

The t-statistic is calculated as the difference between the two sample means, divided by the standard error of the difference.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

2. Degrees of Freedom (ν):

The degrees of freedom are calculated using the Welch-Satterthwaite equation, which accounts for the different sample sizes and variances.

$$\nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

Where:

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means of Group 1 and Group 2.
- s_1^2 and s_2^2 are the sample variances of Group 1 and Group 2.

- n_1 and n_2 are the sample sizes of Group 1 and Group 2.
- ν is the effective degrees of freedom for the test.

7. Paired t-test

The paired t-test determines if the mean difference between two sets of paired observations is zero. The test works by first calculating the difference (d_i) for each pair. Then, it calculates the t-statistic for the mean of these differences.

t-statistic:

The t-statistic is calculated by dividing the mean of the differences by the standard error of the mean difference.

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

Where:

- t: The t-statistic.
- $\bar{d}$: The mean of the differences for all pairs.
- s_d: The standard deviation of the differences.
- n: The number of pairs.

B. Strategy Metrics

B.1 Baseline Strategy Indicators

B.1.1 Breakout Strategy

(BB)

Buy Signal

Formula:

$$\text{BB-Buy}_t = \begin{cases} 1 & \text{if } C_t > H_{d-1}^{\text{session}} \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

where $H_{d-1}^{\text{session}} = \max\{H_s \mid s \in \text{session}_{d-1}\}$.
Granularity: 1-minute.
Description: Signals a buy when the close price exceeds the previous session's high. Entries are not taken after 15:00 EST.

Sell Signal

Formula:

$$\text{BB-Sell}_t = \begin{cases} 1 & \text{if } C_t < L_{d-1}^{\text{session}} \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

where $L_{d-1}^{\text{session}} = \min\{L_s \mid s \in \text{session}_{d-1}\}$.
Granularity: 1-minute.
Description: Signals a sell when the close price falls below the previous session's low. Entries are not taken after 15:00 EST.

Stop-Loss

Concept: A fixed percentage below (for longs) or above (for shorts) the entry price.
Formula (Long):

$$\text{StopPrice} = C_{\text{entry}} \cdot (1 - \delta_{sl})$$

Granularity: Per-trade.
Description: Sets a stop-loss at a percentage δ_{sl} away from the entry price.

Take-Profit

Concept: A fixed percentage above (for longs) or below (for shorts) the entry price.
Formula (Long):

$$\text{ProfitTarget} = C_{\text{entry}} \cdot (1 + \delta_{tp})$$

Granularity: Per-trade.
Description: Sets a take-profit target at a percentage δ_{tp} away from the entry price.

B.1.2 Bollinger Bands Strategy (BBB)

Upper, Middle, and Lower Bands

Formulas:

$$\begin{aligned}
\mathrm{BB}_{\mathrm{Upper},t} &= \mathrm{SMA}_t(\tau_{bb}) + \gamma_{bb} \cdot \sigma_t(\tau_{bb}) \\
\mathrm{BB}_{\mathrm{Middle},t} &= \mathrm{SMA}_t(\tau_{bb}) \\
\mathrm{BB}_{\mathrm{Lower},t} &= \mathrm{SMA}_t(\tau_{bb}) - \gamma_{bb} \cdot \sigma_t(\tau_{bb})
\end{aligned}$$

Granularity: 1-minute.
Description: Standard Bollinger Bands calculated over a τ_{bb}-period window with γ_{bb} standard deviations.

Buy Signal (Mean Reversion)

Formula:

$$\text{BBB-Buy}_t = \begin{cases} 1 & \text{if } C_t < \mathrm{BB}_{\mathrm{Lower},t} \text{ and } t_{\mathrm{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy when the close price crosses below the lower Bollinger Band, anticipating a reversion to the mean.

Sell Signal (Mean Reversion)

Formula:

$$\text{BBB-Sell}_t = \begin{cases} 1 & \text{if } C_t > \mathrm{BB}_{\mathrm{Upper},t} \text{ and } t_{\mathrm{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell when the close price crosses above the upper Bollinger Band, anticipating a reversion to the mean.

Long Exit Signal

Formula:

$$\text{BBB-LongExit}_t = \begin{cases} 1 & \text{if } C_t > \mathrm{BB}_{\mathrm{Middle},t} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Exits a long position when the price reverts and crosses above the middle band (SMA).

Short Exit Signal

Formula:

$$\text{BBB-ShortExit}_t = \begin{cases} 1 & \text{if } C_t < \text{BB}_{\text{Middle},t} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Exits a short position when the price reverts and crosses below the middle band (SMA).

B.1.3 Momentum Strategy (BM)

Buy Signal

Formula:

$$\text{BM-Buy}_t = \begin{cases} 1 & \text{if } \sum_{i=0}^{\tau_{\text{mom}}-1} \mathbb{1}\{C_{t-i} > O_{t-i}\} \geq \tau_{\text{mom}} \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy if the last τ_{mom} consecutive bars have been bullish ($C > O$).

Sell Signal

Formula:

$$\text{BM-Sell}_t = \begin{cases} 1 & \text{if } \sum_{i=0}^{\tau_{\text{mom}}-1} \mathbb{1}\{C_{t-i} < O_{t-i}\} \geq \tau_{\text{mom}} \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell if the last τ_{mom} consecutive bars have been bearish ($C < O$).

Trailing Stop-Loss

Concept: A stop-loss that trails the highest price (for longs) or lowest price (for shorts) reached during a trade.
Formula (Long):

$$\text{StopPrice}_t = \max(\text{StopPrice}_{t-1}, \text{HighPriceSinceEntry}_t \cdot (1 - \delta_{\text{mom\_trail}}))$$

Granularity: Per-trade.
Description: Sets a trailing stop-loss at a percentage $\delta_{\text{mom\_trail}}$ below the highest price observed since the position was opened.

B.2 Volume-Enhanced Strategy Indicators

B.2.1 Shared Indicators

Developing Value Area High (dVAH), Low (dVAL), and Point of Control (dPOC)

Concept: These metrics are calculated using an expanding window of the current day's data, not the previous session's. Let $S_d(t)$ be the set of time indices from the start of day d up to the current time t.
Formulas:

$$V_d(p,t) = \sum_{s \in S_d(t): \lfloor C_s/0.01 \rfloor \cdot 0.01 = p} V_s$$

$$\text{dPOC}_t = \arg\max_p V_d(p,t)$$

$P_d(t)$ is the set of prices starting from dPOC_t and adding prices in descending order of volume $V_d(p,t)$ until $\sum_{p \in P_d(t)} V_d(p,t) \geq \phi_{va} \cdot \sum_p V_d(p,t)$.

$$\text{dVAH}_t = \max\{p \in P_d(t)\}$$

$$\text{dVAL}_t = \min\{p \in P_d(t)\}$$

Granularity: 1-minute (recalculated each minute).
Description: The POC, VAH, and VAL calculated based on the cumulative volume profile of the current trading day up to time t.

Average Directional Index (ADX)

Formula: Standard ADX calculation over a τ_{adx} period.
Granularity: 1-minute.
Description: Measures trend strength over a τ_{adx}-period window, scaled 0–100. Used as a filter instead of RSI.

Volume Surge

Formula:

$$\text{VS}_t = \begin{cases} 1 & \text{if } V_t > \kappa_{\text{surge}} \cdot \text{SMA}(V_t, \tau_{\text{vol\_avg}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals when current volume exceeds κ_{surge} times the recent average volume.

ATR-Based Stop-Loss & Take-Profit

Concept: Stop-loss and take-profit levels are set based on a multiple of the daily Average True Range (ATR).
Formulas (Long):

$$\text{StopPrice} = C_{\text{entry}} - \alpha_{sl} \cdot \text{ATR}_d$$

$$\text{ProfitTarget} = C_{\text{entry}} + \alpha_{tp} \cdot \text{ATR}_d$$

Granularity: Daily ATR applied per-trade.
Description: Sets risk management levels by subtracting/adding a multiple (α_{sl}, α_{tp}) of the daily ATR from the entry price. The take-profit multiple is often dependent on the stop-loss multiple via a risk-reward ratio ($\alpha_{tp} = \alpha_{sl} \cdot \text{RR\_Ratio}$).

B.2.2 Volume-Enhanced Breakout Strategy (VEB)

Buy Signal

Formula:

$$\text{VEB-Buy}_t = \begin{cases} 1 & \text{if } (C_t > \text{dVAH}_t) \wedge \text{VS}_t \wedge (\text{ADX}_t > \theta_{\text{adx}}) \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy when price breaks above the developing VAH with a concurrent volume surge and a strong trend confirmed by ADX.

Sell Signal

Formula:

$$\text{VEB-Sell}_t = \begin{cases} 1 & \text{if } (C_t < \text{dVAL}_t) \wedge \text{VS}_t \wedge (\text{ADX}_t > \theta_{\text{adx}}) \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell when price breaks below the developing VAL with a concurrent volume surge and a strong trend confirmed by ADX.

Stop-Loss & Take-Profit

Uses the ATR-Based definitions above.

B.2.3 Volume-Enhanced Momentum Strategy (VEM)

Relative Volume

Formula:

$$\text{RV}_t = V_t / \text{SMA}(V_t, \tau_{\text{vol\_avg}})$$

Granularity: 1-minute.
Description: Current bar's volume as a multiple of its recent average.

Buy Signal

Formula:

$$\text{VEM-Buy}_t = \begin{cases} 1 & \text{if } (\text{RV}_t > \kappa_{\text{vol\_mom}}) \wedge (C_t > H_{d-1}^{\text{session}}) \wedge (\text{ADX}_t > \theta_{\text{adx}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy when relative volume is high, the price breaks the previous day's high, and trend strength is confirmed by ADX.

Sell Signal

Formula:

$$\text{VEM-Sell}_t = \begin{cases} 1 & \text{if } (\text{RV}_t > \kappa_{\text{vol\_mom}}) \wedge (C_t < L_{d-1}^{\text{session}}) \wedge (\text{ADX}_t > \theta_{\text{adx}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell when relative volume is high, the price breaks the previous day's low, and trend strength is confirmed by ADX.

Trailing Stop-Loss

An ATR-based stop-loss that trails the price, as defined by `sl_stop` and `sl_trail=True` in the code.

B.2.4 Volume-Enhanced VWAP Reversion Strategy (VEV)

On-Balance Volume Threshold (OBVT)

Formula:

$$\text{OBVT}_t(\theta_{\text{obv}}, \tau_{\text{obv}}) = Q_{\theta_{\text{obv}}}(\{\text{OBV}_{t-i} \mid i = 0, \dots, \tau_{\text{obv}} - 1\})$$

Granularity: 1-minute.
Description: The lower (θ_{obv}) and upper ($1 - \theta_{\text{obv}}$) percentile of OBV over a recent τ_{obv}-minute window.

Buy Signal

Formula:

$$\text{VEV-Buy}_t = \begin{cases} 1 & \text{if } (C_t < \text{VWAP}_t) \wedge (\text{OBV}_t < \text{OBVT}_t(\theta_{\text{obv}}, \tau_{\text{obv}})) \wedge (\text{VWAPS}_t \geq \epsilon_{\text{slope}}) \wedge \\ & (\text{VWAP}_t < \text{VWAP}_{t-\tau_{\text{vwap\_trend}}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy on price weakness ($C < \text{VWAP}$), low OBV, positive VWAP slope, and a medium-term downward VWAP trend.

Sell Signal

Formula:

$$\text{VEV-Sell}_t = \begin{cases} 1 & \text{if } (C_t > \text{VWAP}_t) \wedge (\text{OBV}_t > \text{OBVT}_t(1 - \theta_{\text{obv}}, \tau_{\text{obv}})) \wedge (\text{VWAPSe}_t \leq \\ & -\epsilon_{\text{slope}}) \wedge (\text{VWAP}_t > \text{VWAP}_{t-\tau_{\text{vwap\_trend}}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell on price strength ($C >$ VWAP), high OBV, negative VWAP slope, and a medium-term upward VWAP trend.

Stop-Loss & Take-Profit

Uses the ATR-Based definitions.

B.3 Deep Learning-Enhanced Strategy Indicators

B.3.1 Shared Indicator

Predicted Volume Confirmation

Concept: A condition confirming that the predicted volume 15 minutes ahead ($\hat{V}_{t+15}$) is significantly larger than the recent average volume.
Formula:

$$\text{PVC}_t = (\hat{V}_{t+15} > \kappa_{dl} \cdot \text{SMA}(V_t, \tau_{\text{vol\_avg}}))$$

Description: Used across DL strategies to ensure trades are taken in anticipated high-volume environments.

B.3.2 Deep Learning Breakout Strategy (DLB)

Developing VAH/VAL (Predicted Volume)

Concept: Same as the dVAH/dVAL, but the volume profile is built using predicted 15-minute ahead volume ($\hat{V}_{t+15}$) instead of actual current volume.
Description: Creates VAH and VAL based on where volume is expected to occur.

Buy Signal

Formula:

$$\text{DLB-Buy}_t = \begin{cases} 1 & \text{if } (C_t > \text{dVAH}_t^{\text{pred}}) \wedge \text{PVC}_t \wedge (\text{ADX}_t > \theta_{\text{adx}}) \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy on a breakout above the predicted volume VAH, confirmed by high future volume and strong ADX trend.

Sell Signal

Formula:

$$\text{DLB-Sell}_t = \begin{cases} 1 & \text{if } (C_t < \text{dVAL}_t^{\text{pred}}) \wedge \text{PVC}_t \wedge (\text{ADX}_t > \theta_{\text{adx}}) \text{ and } t_{\text{hour}} < 15:00 \text{ EST} \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell on a breakdown below the predicted volume VAL, confirmed by high future volume and strong ADX trend.

Stop-Loss & Take-Profit

Uses the ATR-Based definitions.

B.3.3 Deep Learning Volume Momentum Strategy (DLM)

Predicted Volume Trend Confirmation (PVTC)

Formula:

$$\mathrm{PVTC}_t = (\mathrm{SMA}(\hat{V}_{t+15}, \tau_{\mathrm{vol\_trend}}) > \kappa_{dl} \cdot \mathrm{SMA}(V_t, \tau_{\mathrm{vol\_avg}}))$$

Description: A modified confirmation where a moving average of the predicted volume must exceed a multiple of the average of actual volume, ensuring a sustained expectation of high volume.

Buy Signal

Formula:

$$\text{DLM-Buy}_t = \begin{cases} 1 & \text{if } (C_t > H_{d-1}^{\mathrm{session}}) \land \mathrm{PVTC}_t \land (\mathrm{ADX}_t > \theta_{\mathrm{adx}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy on a breakout of yesterday's high, confirmed by a rising trend in predicted volume and strong ADX.

Sell Signal

Formula:

$$\text{DLM-Sell}_t = \begin{cases} 1 & \text{if } (C_t < L_{d-1}^{\mathrm{session}}) \land \mathrm{PVTC}_t \land (\mathrm{ADX}_t > \theta_{\mathrm{adx}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell on a breakdown of yesterday's low, confirmed by a rising trend in predicted volume and strong ADX.

Trailing Stop-Loss

An ATR-based stop-loss that trails the price.

B.3.4 Deep Learning VWAP Reversion Strategy (DLV)

Predicted Volume Confirmation (vs. Current Volume)

Formula:

$$\mathrm{PVC}_t^{\mathrm{current}} = (\hat{V}_{t+15} > \kappa_{dl} \cdot V_t)$$

Description: A simpler confirmation where predicted future volume must simply be κ_{dl} times larger than the single most recent bar's volume.

Buy Signal

Formula:

$$\text{DLV-Buy}_t = \begin{cases} 1 & \text{if } (C_t < (1 - \delta_{\text{vwap\_buy}}) \cdot \text{VWAP}_t) \wedge \text{PVC}_t^{\text{current}} \wedge (\text{VWAP}_t < \text{VWAP}_{t-\tau_{\text{vwap\_trend}}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a buy when price is significantly below VWAP, future volume is predicted to be high, and VWAP is in a short-term downtrend (expecting a reversal).

Sell Signal

Formula:

$$\text{DLV-Sell}_t = \begin{cases} 1 & \text{if } (C_t > (1 + \delta_{\text{vwap\_sell}}) \cdot \text{VWAP}_t) \wedge \text{PVC}_t^{\text{current}} \wedge (\text{VWAP}_t > \text{VWAP}_{t-\tau_{\text{vwap\_trend}}}) \\ 0 & \text{otherwise} \end{cases}$$

Granularity: 1-minute.
Description: Signals a sell when price is significantly above VWAP, future volume is predicted to be high, and VWAP is in a short-term uptrend (expecting a reversal).

Exit Signal

Concept: In addition to stop-loss/take-profit, positions are exited if the price reverts to the VWAP line.
Formulas:
Long Exit: if C_t crosses above VWAP_t
Short Exit: if C_t crosses below VWAP_t
Granularity: Per-trade.
Description: A primary exit mechanism based on the mean reversion thesis being completed.

Stop-Loss & Take-Profit

Uses the ATR-Based definitions, serving as a secondary exit.

C. Market Regimes

(a) **Stochastic Oscillator**

Calculates the percentage of the close price relative to the 14-day high-low range.

- $$\text{stoch\_k}_t = \frac{C_t - \min_{i=0,\ldots,13}\{L_{t-i}\}}{\max_{i=0,\ldots,13}\{H_{t-i}\} - \min_{i=0,\ldots,13}\{L_{t-i}\}} \cdot 100 \tag{C.1}$$
- **Where:**
  - stoch_k$_t$: Stochastic oscillator %K at time t.
  - C_t, H_t, L_t: Close, high, and low prices at time t.
  - $t - i$: The last 14 days (where $i = 0, \ldots, 13$).

(b) **Volatility Regimes**

Classifies volatility based on the slow stochastic oscillator relative to its 252-day percentiles.

- $$V_t = \begin{cases} \text{Low} & \text{if slow\_k}_t < Q_{0.15}(\text{slow\_k}_{252}) \\ \text{Medium} & \text{if } Q_{0.15}(\text{slow\_k}_{252}) \leq \text{slow\_k}_t \leq Q_{0.75}(\text{slow\_k}_{252}) \\ \text{High} & \text{if slow\_k}_t > Q_{0.75}(\text{slow\_k}_{252}) \\ \text{Undefined} & \text{if slow\_k}_t \text{ or } Q_{0.15}, Q_{0.75} \text{ is NaN} \end{cases} \tag{C.2}$$
- **Where:**
  - V_t: Volatility regime (Low, Medium, High, Undefined).
  - slow_k$_t$: Slow stochastic %K at time t.
  - $\text{slow\_k}_{252} = \{\text{slow\_k}_{t-i} \mid i = 0, \ldots, 251\}$: Slow %K over the last 252 days.
  - Q_θ: The θ-th percentile over the 252-day period.

(c) **Average Directional Index (ADX)**

Measures trend strength over a 14-day period.

- $$\text{adx}_t = 100 \cdot \frac{\frac{1}{14}\sum_{i=0}^{13} \frac{|\text{p\_dm}_{t-i} - \text{m\_dm}_{t-i}|}{\text{tr}_{t-i}}}{\frac{1}{14}\sum_{i=0}^{13} \text{tr}_{t-i}} \tag{C.3}$$
- **Where:**
  - adx$_t$: Average Directional Index at time t.
  - $\text{p\_dm}_t = \max(H_t - H_{t-1}, 0)$ if $H_t - H_{t-1} > L_{t-1} - L_t$, else 0.
  - $\text{m\_dm}_t = \max(L_{t-1} - L_t, 0)$ if $L_{t-1} - L_t > H_t - H_{t-1}$, else 0.
  - $\text{tr}_t = \max(H_t - L_t, |H_t - C_{t-1}|, |L_t - C_{t-1}|)$: True range.

(d) **Directional Indicators**

Quantifies directional movement strength over 14 days.

- $$\text{p\_di}_t = 100 \cdot \frac{\frac{1}{14}\sum_{i=0}^{13} \text{p\_dm}_{t-i}}{\text{atr}_t}, \quad \text{m\_di}_t = 100 \cdot \frac{\frac{1}{14}\sum_{i=0}^{13} \text{m\_dm}_{t-i}}{\text{atr}_t} \tag{C.4}$$
- **Where:**
  - $\text{p\_di}_t, \text{m\_di}_t$: Positive and negative directional indicators.

- $\text{atr}_t = \frac{1}{14}\sum_{i=0}^{13} \text{tr}_{t-i}$: Average true range.
- $\text{p\_dm}_t, \text{m\_dm}_t, \text{tr}_t$: As defined previously.

(e) **Bollinger Band Width**

Measures price volatility relative to a 14-day moving average.

- $$\text{bb\_width}_t = \frac{\text{ub}_t - \text{lb}_t}{\text{mb}_t} \tag{C.5}$$

- **Where:**
  - bb_width_t: Bollinger Band Width at time t.
  - $\text{ub}_t = \text{SMA}_t(14) + 2 \cdot \sqrt{\frac{1}{14}\sum_{i=0}^{13}(C_{t-i} - \text{SMA}_t(14))^2}$: Upper band.
  - $\text{lb}_t = \text{SMA}_t(14) - 2 \cdot \sqrt{\frac{1}{14}\sum_{i=0}^{13}(C_{t-i} - \text{SMA}_t(14))^2}$: Lower band.
  - $\text{mb}_t = \text{SMA}_t(14) = \frac{1}{14}\sum_{i=0}^{13} C_{t-i}$: Middle band.

(f) **Trend Regimes**

Classifies trend direction and strength using ADX, directional indicators, and price patterns.

- $$T_t = \begin{cases} \text{Uptrend} & \text{if } \text{adx}_t > Q_{0.75}(\text{adx}_{252}) \wedge \text{p\_di}_t > \text{m\_di}_t \wedge (H_t > H_{t-1} \vee L_t > L_{t-1}) \\ \text{Downtrend} & \text{if } \text{adx}_t > Q_{0.75}(\text{adx}_{252}) \wedge \text{m\_di}_t > \text{p\_di}_t \wedge (H_t < H_{t-1} \vee L_t < L_{t-1}) \\ \text{Range} & \text{if } \text{adx}_t < Q_{0.25}(\text{adx}_{252}) \wedge \text{bb\_width}_t < Q_{0.05}(\text{bb\_width}_{252}) \\ \text{Range} & \text{otherwise} \\ \text{Undefined} & \text{if any indicator or percentile is NaN, or } \text{bb\_width}_t = 0 \end{cases} \tag{C.6}$$

- **Where:**
  - T_t: Trend regime (Uptrend, Downtrend, Range, Undefined).
  - $\text{adx}_{252}, \text{bb\_width}_{252} = \{\text{adx}_{t-i}, \text{bb\_width}_{t-i} \mid i = 0, \ldots, 251\}$: ADX and bb_width over 252 days.
  - Q_θ: The θ-th percentile over the 252-day period.
  - $\text{p\_di}_t, \text{m\_di}_t, H_t, L_t$: As defined previously.

(g) **Liquidity Regimes**

Classifies liquidity based on bid-ask spread, order book depth, and volume.

- $$L_t = \begin{cases} \text{High} & \text{if } \text{BAS}_t < Q_{0.10}(\text{bas}_{252}) \wedge (\text{obd}_t > Q_{0.70}(\text{obd}_{252}) \vee \text{vol}_t > Q_{0.70}(\text{vol}_{252})) \\ \text{Low} & \text{if } \text{BAS}_t > Q_{0.95}(\text{bas}_{252}) \wedge \text{obd}_t < Q_{0.30}(\text{obd}_{252}) \\ \text{Medium} & \text{otherwise} \end{cases} \tag{C.7}$$

- **Where:**
  - L_t: Liquidity regime (High, Low, Medium).
  - $\text{BAS}_t = \frac{\text{est\_BAS}_t}{100}$: Bid-ask spread percentage.
  - $\text{bas}_{252}, \text{obd}_{252}, \text{vol}_{252} = \{\text{BAS}_{t-i}, \text{obd}_{t-i}, \text{volume}_{t-i} \mid i = 0, \ldots, 251\}$: Over 252 days.
  - obd_t: Order book depth, clipped at a minimum of 1000.
  - Q_θ: The θ-th percentile over the 252-day period.

D. Volume prediction modelling

(a) **Z-Score Normalization**
Standardizes continuous features to have zero mean and unit standard deviation.

- $$z = \frac{x - \mu}{\sigma} \tag{D.1}$$
- **Where:**
  - z: Normalized z-score.
  - x: Original feature value.
  - μ: Mean of the feature across the dataset.
  - σ: Standard deviation of the feature.

(b) **Log-Transformation and Scaling of Volume**
Log-transforms trading volume to reduce skewness and scales it to the range [-1, 1].

- $$y = \log(v + 1), \quad y_{\text{scaled}} = 2 \cdot \frac{y - \mu_y}{\sigma_y} - 1 \tag{D.2}$$
- **Where:**
  - v: Original trading volume.
  - y: Log-transformed volume.
  - y_{scaled}: Scaled log-transformed volume.
  - μ_y: Mean of log-transformed volume.
  - σ_y: Standard deviation of log-transformed volume.

(c) **Mean Absolute Error (MAE)**
Measures the average absolute difference between predicted and actual log-transformed volumes.

- $$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{D.3}$$
- **Where:**
  - MAE: Mean Absolute Error.
  - n: Number of predictions.
  - y_i: Actual log-transformed volume for observation i.
  - $\hat{y}_i$: Predicted log-transformed volume for observation i.

(d) **Percentage Error from MAE**
Calculates the percentage error based on MAE relative to the scaled volume range.

- $$\text{Percentage Error} = \frac{\text{MAE}}{\text{Range}} \cdot 100 \tag{D.4}$$

(e) **Coefficient of Determination (R^2)**
Measures the proportion of variance in the actual volume that is explained by the model.

- $$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{D.5}$$

- **Where:**
  - R^2: Coefficient of Determination.
  - y_i: Actual log-transformed volume for observation i.
  - $\hat{y}_i$: Predicted log-transformed volume for observation i.
  - $\bar{y}$: Mean of actual log-transformed volumes.

(f) **TFT Input Representation**

Structures the input tensor for the Temporal Fusion Transformer with a 60-minute lookback window.

- $$X_t = [y_{t-60}, \ldots, y_{t-1}, x^h_{t-60}, \ldots, x^h_{t-1}, x^f_t, \ldots, x^f_{t+15}] \tag{D.6}$$

- **Where:**
  - X_t: Input tensor at time t.
  - y_{t-i}: Log-transformed volume at time $t - i$ ($i = 1, \ldots, 60$).
  - x^h_{t-i}: Historical exogenous variables (e.g., price, volume lags, regimes).
  - x^f_{t+j}: Future exogenous variables (e.g., time, regimes, $j = 0, \ldots, 15$).

(g) **TFT Gated Linear Unit (GLU)**

Processes input features with a gated linear unit to control the flow of information.

- $$\text{GLU}(x) = (W_1 x + b_1) \odot \sigma(W_2 x + b_2) \tag{D.7}$$

- **Where:**
  - $\text{GLU}(x)$: Gated linear unit output.
  - x: Input feature vector.
  - W_1, W_2: Weight matrices.
  - b_1, b_2: Bias vectors.
  - σ: The sigmoid activation function.
  - $\odot$: Element-wise (Hadamard) product.

(h) **TFT Self-Attention Mechanism**

Captures temporal dependencies using a multi-head self-attention mechanism.

- $$\begin{aligned} \text{Attention}(Q, K, V) &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \\ \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \\ \text{where head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{D.8}$$

- **Where:**
  - Q, K, V: Query, Key, and Value matrices from the input sequence.
  - W_i^Q, W_i^K, W_i^V: Weight matrices for the i-th attention head.
  - W^O: Output projection matrix.
  - d_k: Dimension of each head's key/query (e.g., 128 / 2 = 64).
  - h: Number of attention heads (e.g., 2).

(i) **TFT Output Prediction**

Predicts the log-transformed volume 15 minutes ahead using a linear output layer.

- $$\hat{y}_{t+15} = W_{\text{out}} h_t + b_{\text{out}} \tag{D.9}$$

- **Where:**
  - $\hat{y}_{t+15}$: Predicted log-transformed volume at time $t+15$.
  - h_t: Hidden state after attention and gating layers at time t.
  - W_{out}: Output weight matrix.
  - b_{out}: Output bias vector.

(j) **TFT Loss Function**

Optimizes the Temporal Fusion Transformer using the MAE loss function.

- $$L = \frac{1}{n} \sum_{i=1}^{n} |y_{t+15,i} - \hat{y}_{t+15,i}| \tag{D.10}$$

- **Where:**
  - L: The loss value.
  - n: Number of predictions in the batch.
  - $y_{t+15,i}$: Actual log-transformed volume at $t+15$ for observation i.
  - $\hat{y}_{t+15,i}$: Predicted log-transformed volume at $t+15$ for observation i.

E. Evaluation and Backtesting Metrics

(a) **Sharpe Ratio**
Measures the annualized risk-adjusted return of a strategy relative to the risk-free rate.

- $$\mathrm{SR} = \frac{\bar{R} - R_f}{\sigma_R} \cdot \sqrt{252} \tag{E.1}$$
- **Where:**
  - SR: Annualized Sharpe ratio.
  - $\bar{R}$: Mean daily return of the strategy.
  - R_f: Mean daily risk-free rate (e.g., US Treasury bill yield).
  - σ_R: Standard deviation of daily strategy returns.
  - 252: Number of trading days per year for annualization.

(b) **Maximum Median Drawdown**
Quantifies the maximum median loss from a peak to a trough in portfolio value over a period.

- $$\mathrm{MMD} = \mathrm{median}\left(\max_t \left(\frac{P_{\mathrm{peak},t} - P_t}{P_{\mathrm{peak},t}}\right)\right) \tag{E.2}$$
- **Where:**
  - MMD: Maximum median drawdown.
  - P_t: Portfolio value at time t.
  - $P_{\mathrm{peak},t}$: Peak portfolio value before time t.
  - $\max_t$: Maximum over all time periods.
  - median: Median across multiple simulations or periods.

(c) **Market Capture Ratio**
Compares the strategy's return to the market's return to assess performance.

- $$\mathrm{MCR} = \frac{R_{\mathrm{strategy}}}{R_{\mathrm{market}}} \tag{E.3}$$
- **Where:**
  - MCR: Market capture ratio.
  - R_{strategy}: Cumulative return of the strategy over the period.
  - R_{market}: Cumulative return of the benchmark market index (e.g., NASDAQ-100).

(d) **Profit Factor**
Evaluates the ratio of gross profits to gross losses from trades.

- $$\mathrm{PF} = \frac{\sum_{i \in \mathrm{wins}} |R_i|}{\sum_{j \in \mathrm{losses}} |R_j|} \tag{E.4}$$
- **Where:**

- PF: Profit factor.
- R_i: Return of winning trade i.
- R_j: Return of losing trade j.
- $\sum_{i \in \text{wins}}$: Sum over all winning trades.
- $\sum_{j \in \text{losses}}$: Sum over all losing trades.

(e) **Kelly Criterion for Optimal Leverage**

Determines the optimal fraction of capital to allocate per trade to maximize growth.

- $$f = \frac{p \cdot b - (1 - p)}{b} \tag{E.5}$$
- **Where:**
  - f: Optimal fraction of capital to bet.
  - p: Probability of a winning trade.
  - b: Net odds received on winning trades (profit/loss ratio per trade).

(f) **Maximum Favorable Excursion**

Measures the maximum profit achieved during a trade before closing.

- $$\text{MFE}_i = \max_{t \in T_i} \left(\frac{P_{t,i} - P_{\text{entry},i}}{P_{\text{entry},i}} \right) \tag{E.6}$$
- **Where:**
  - MFE_i: Maximum favorable excursion for trade i.
  - $P_{t,i}$: Price at time t during trade i.
  - $P_{\text{entry},i}$: Entry price of trade i.
  - T_i: Time period of trade i.

(g) **Maximum Adverse Excursion**

Measures the maximum loss incurred during a trade before closing.

- $$\text{MAE}_i = \max_{t \in T_i} \left(\frac{P_{\text{entry},i} - P_{t,i}}{P_{\text{entry},i}} \right) \tag{E.7}$$
- **Where:**
  - MAE_i: Maximum adverse excursion for trade i.
  - $P_{t,i}$: Price at time t during trade i.
  - $P_{\text{entry},i}$: Entry price of trade i.
  - T_i: Time period of trade i.

(h) **Calmar Ratio**

Measures the risk-adjusted return relative to the maximum drawdown. A higher Calmar ratio is generally better.

- $$\text{Calmar Ratio} = \frac{\text{Annualized Return}}{\text{Maximum Drawdown}} \tag{E.8}$$
- **Where:**
  - Annualized Return: The geometric mean of the portfolio's returns, scaled to a yearly figure.
  - Maximum Drawdown: The largest peak-to-trough decline in portfolio value.

Of course. Here are the additions of Calmar Ratio, Sortino Ratio, and Omega Ratio, formatted to match the style of your document.

(i) **Sortino Ratio**

Similar to the Sharpe ratio, but it only penalizes for downside volatility. It differentiates harmful volatility from total overall volatility.

-
$$\text{Sortino Ratio} = \frac{\bar{R} - R_f}{\sigma_d} \cdot \sqrt{252} \tag{E.9}$$

- **Where:**
  - $\bar{R}$: Mean daily return of the strategy.
  - R_f: Mean daily risk-free rate.
  - σ_d: Standard deviation of the downside, considering only returns that fall below a minimum acceptable return.
  - 252: Number of trading days per year for annualization.

(j) **Omega Ratio**

A risk-return performance measure that considers all moments of the return distribution. It is the ratio of the probability of winning to the probability of losing.

-
$$\Omega(\theta) = \frac{\int_\theta^\infty (1 - F(r))\, dr}{\int_{-\infty}^\theta F(r)\, dr} \tag{E.10}$$

- **Where:**
  - $\Omega(\theta)$: The Omega ratio at a given threshold θ.
  - θ: The target return threshold that defines gains and losses.
  - $F(r)$: The cumulative probability distribution function of the returns.
  - $\int_\theta^\infty (1 - F(r))\, dr$: The probability-weighted returns above the threshold θ.
  - $\int_{-\infty}^\theta F(r)\, dr$: The probability-weighted returns below the threshold θ.

F. Risk Management Formulas

- **Tharp Expectancy**
  The Tharp Expectancy formula calculates the expected profit per trade based on possible trade outcomes.
  -
  $$E = \sum_{i=1}^{n} (P_i \cdot R_i) \tag{F.1}$$
    Where:
    * E: Expectancy (expected profit per trade)
    * n: Number of possible trade outcomes
    * P_i: Probability of outcome i (where $\sum P_i = 1$)
    * R_i: Return for outcome i (profit or loss in dollars)
  - **Binary Win/Loss Scenario:**
  $$E = (P_{\text{win}} \cdot R_{\text{avg win}}) - (P_{\text{loss}} \cdot R_{\text{avg loss}}) \tag{F.2}$$
    Where:
    * P_{win}: Probability of a winning trade
    * $R_{\text{avg win}}$: Average profit per winning trade
    * P_{loss}: Probability of a losing trade ($P_{\text{loss}} = 1 - P_{\text{win}}$)
    * $R_{\text{avg loss}}$: Average loss per losing trade
- **Kelly Criterion**
  The Kelly Criterion determines the optimal fraction of capital to allocate to a trading strategy.
  -
  $$F_i^* = \frac{m_i}{s_i^2} \tag{F.3}$$
    Where:
    * F_i^*: Optimal fraction of capital to allocate to strategy i
    * m_i: Mean return of strategy i per trade (net of costs)
    * s_i^2: Variance of returns for strategy i
  - **Binary Outcome:**
  $$F^* = \frac{bp - q}{b} \tag{F.4}$$
    Where:
    * F^*: Optimal fraction of capital
    * b: Net odds (profit per unit wagered)
    * p: Probability of winning
    * q: Probability of losing ($q = 1 - p$)