HEC MONTRÉAL

Intraday Stock Return Prediction :
A Comparative Study of Econometrics and Machine Learning Models

par

Zahrasadat Hashemi

Vincent Grégoire, Professeur, Département de Finance
HEC Montréal
Directeur de recherche

Sciences de la gestion
(Spécialisation Applied Financial Economics)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

August 2025

Résumé

Les traders font face à une décision cruciale à la mi-journée : déterminer avec quelle agressivité exécuter les ordres restants sans savoir comment évolueront les prix et la liquidité. Lorsque l'exécution est en avance ou en retard par rapport au calendrier, les écarts par rapport au VWAP rendent ce choix particulièrement risqué. Des prévisions fiables des rendements de l'après-midi et des pressions latentes de la demande sont donc essentielles pour améliorer la qualité d'exécution au-delà du simple suivi du VWAP. Cette recherche évalue la valeur prédictive des indicateurs de marché disponibles à la mi-journée pour anticiper les rendements et les déséquilibres de flux d'ordres entre 13 h et la clôture. Dix années de données intrajournalières issues des principales bourses nord-américaines sont exploitées, avec trois cibles : le rendement, le déséquilibre du nombre de transactions (TCI) et le déséquilibre du volume de transactions (TVI). Les prédicteurs couvrent plusieurs catégories, notamment les mesures de prix, les indicateurs transactionnels, la liquidité, les métriques d'informativité et de volatilité, les attributs liés aux symboles et au temps, ainsi que les flux d'ordres institutionnels et de détail. Un cadre à fenêtre roulante est appliqué à la régression Lasso et au modèle Random Forest pour comparer structures linéaires et non linéaires.

Les résultats montrent que les mesures de flux d'ordres sont nettement plus prévisibles que les rendements, le TCI surpassant systématiquement le TVI. La précision s'améliore avec une stratification par industrie, ce qui met en évidence des schémas microstructurels sectoriels. Le VWAP a un rôle faible et incohérent en Lasso mais améliore nettement la précision dans Random Forest, surtout pour les rendements, révélant des effets d'interaction non linéaires. L'analyse de regroupement souligne en outre une hétérogénéité persistante entre titres. Un backtest sur 2024 montre qu'une stratégie long–short quotidienne génère des ratios de Sharpe élevés et des gains asymétriques. Cette étude propose un cadre robuste pour la prévision mi-journée–clôture et offre des pistes pour des stratégies VWAP adaptatives. Ses limites concernent l'absence de données de carnet d'ordres complet et de sentiment, ouvrant la voie à des travaux futurs avec des modèles sensibles aux régimes et des simulations prolongées.

Abstract

Traders face a critical decision at midday: how aggressively to complete outstanding orders without knowing how prices and liquidity will evolve. When execution is ahead or behind schedule, deviations from VWAP make this choice especially risky. Reliable forecasts of afternoon returns and latent demand pressures are therefore essential for improving execution quality beyond static VWAP tracking. This research examines the predictive value of midday market indicators for forecasting stock returns and order-flow imbalances between 1:00 p.m. and the close. Using ten years of intraday data from major North American exchanges, three targets are considered: return, trade count imbalance (TCI), and trade volume imbalance (TVI). Predictors span broad categories, including price and return measures, transaction-based indicators, liquidity measures, informativeness and volatility metrics, symbol- and time-related attributes, and retail versus institutional order-flow indicators. A rolling-window framework is applied to Lasso regression and Random Forest models to compare linear and nonlinear predictive structures.

Results show that order-flow measures are substantially more predictable than returns, with TCI consistently outperforming TVI. Forecast accuracy improves when models are stratified by industry, underscoring the value of sector-specific microstructure patterns. VWAP plays a weak and inconsistent role in the linear framework but significantly improves predictive accuracy in Random Forest, particularly for returns, suggesting nonlinear interaction effects. Clustering analysis further reveals structural heterogeneity across equities, with midday total size, turnover, and shares outstanding emerging as key differentiators. A 2024 backtest demonstrates that a daily rebalanced long–short strategy built from the forecasts achieves high Sharpe ratios and asymmetric gains, as long legs deliver positive returns while short legs generate larger-magnitude losses. The study contributes a practical framework for noon-to-close forecasting, clarifies when predictive content is strongest, and offers insights for adaptive VWAP execution. Limitations include the ab-

sence of full-depth order book features and sentiment data, which future research should address through regime-aware and live-simulation approaches.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Vincent Grégoire, for his continuous support, guidance, and encouragement at every stage of this research. I am thankful to the Laboratoire de calcul et d'exploitation des données (LACED) and Mohamed Jabir for providing the computational resources and technical assistance that made large-scale data analysis possible. I also wish to thank Prof. Manuel Morales, the Fin-ML program, and IVADO for creating valuable resources on artificial intelligence and data science, which greatly enriched the methodological foundation of this work. I also thank the WRDS support team for their fast and comprehensive assistance. Finally, I am deeply grateful to my family and my spouse for their constant encouragement and belief in me, which supported me throughout this journey.

1 Introduction

In the fast-paced and highly competitive environment of modern financial markets, traders face the dual challenge of executing orders efficiently and achieving performance benchmarks that measure execution quality. Among these, the Volume Weighted Average Price (VWAP) has become a widely adopted yardstick, particularly for institutional investors executing large trade volumes. Achieving a fill price better than VWAP is considered a signal of skillful execution, but doing so consistently requires anticipating short-term market dynamics with precision. Institutional execution strategies must balance market impact and timing risk, especially during the midday decision point. By 1:00 p.m., a trader must decide how to complete the remaining portion of the order before market close, without knowing how prices, liquidity, and order flow will evolve during the afternoon. This decision is complicated by the heterogeneous objectives of market participants, macroeconomic announcements, and algorithmic flows, all of which can alter the market state in ways that affect execution costs. Even small improvements in predicting the direction and magnitude of price changes between midday and the close can have material financial consequences for large orders (L. Gao et al. 2018).

VWAP-based execution faces inherent risks due to the uneven intraday distribution of liquidity. Liquidity is typically concentrated near the market open and close, leaving a thinner and less predictable environment at midday. This intraday liquidity fragmentation, driven by inventory management by dealers, portfolio rebalancing, and event-driven trading, creates conditions in which prices can drift away from VWAP and execution risk rises (Fallahi 2023). When prices deviate materially from VWAP, the execution challenge intensifies. If a trader is ahead of schedule, meaning they have executed more of the order than planned, and the price has moved favorably relative to VWAP, slowing down execution may risk losing the advantageous price if the market reverts. Conversely, if a trader is behind schedule and the price has moved unfavorably, the need to "catch up" can require more aggressive trading, increasing market impact and signaling urgency to other participants. These scenarios imply that the gap between price and VWAP at midday is not just a descriptive benchmark but a potential indicator of latent demand or supply pressure in the market. Large deviations, especially when combined with information on order flow imbalances, can reveal the presence of large hidden orders or the urgency of execution by other traders. If such latent pressures can be reliably inferred, they could form the basis for adaptive execution strategies that improve upon static VWAP tracking.

This research builds on that insight by hypothesizing that midday deviations from VWAP, combined with liquidity and order-flow measures, contain predictive information about the direction and magnitude of afternoon price movements and trade imbalances. In other words, the midday state of the market may encode signals about both where prices are likely to move and how liquidity will evolve in the remaining hours of the trading day. The primary objective of this thesis is to develop and compare econometric and machine learning models for forecasting noon-to-close market dynamics using only information available by midday. The study defines and constructs predictive variables that capture midday market conditions, including morning trade count imbalance (TCI) and trade volume imbalance (TVI), relative VWAP, turnover, and intraday momentum. The prediction targets are the afternoon return from 1 p.m. to close, as well as afternoon TCI and TVI, which are relevant both for execution quality and for understanding market microstructure.

Forecasting these measures poses significant methodological challenges. Intraday financial data is noisy, high-frequency, and often nonstationary. The interaction of heterogeneous market participants, each with different objectives and constraints, further complicates modeling efforts (Matías and Reboredo 2012). Traditional econometric approaches, though well-established, may fail to capture nonlinear relationships and adaptive dynamics in modern markets. At the same time, machine learning techniques offer greater flexibility but raise concerns about interpretability, overfitting, and stability in real-time environments (Huddleston, F. Liu, and Stentoft 2023) This research therefore aims to provide a thorough comparison between econometric and machine learning approaches in effectively capturing latent market pressures under real-world trading conditions.

To address these objectives, the analysis estimates and compares two distinct modeling approaches: LASSO regression, selected for its embedded feature selection and interpretability, and Random Forests, chosen for their ability to capture nonlinear interactions among predictors. The role of VWAP is evaluated through controlled exclusion tests to quantify its incremental predictive contribution and to determine whether its effect is linear, as captured by LASSO, or nonlinear, as potentially better captured by Random Forest. In addition, structural patterns are examined through equity clustering, grouping stocks with similar predictive profiles to identify whether certain market segments share common drivers of predictability. Finally, the study evaluates the practical relevance of the forecasts through backtesting, simulating execution strategies that incorporate the model outputs to measure potential performance gains over standard VWAP execution.

From a theoretical perspective, this research contributes to the literature on intraday predictability by isolating the noon-to-close window, a decision point that is operationally critical but underexplored in predictive modeling. While many existing studies assume access to full-session data, this thesis deliberately restricts inputs to midday information, aligning the predictive framework with the constraints faced by real-world traders. From a practical standpoint, the proposed framework offers the potential to improve VWAP-based execution strategies. By forecasting not just price direction but also liquidity imbalances, the approach addresses both components of execution quality: price performance and market impact. The explicit focus on the trader's relative progress versus execution schedule, interpreted through the VWAP gap, provides a microstructure-based rationale for adapting execution speed and aggressiveness in real time.

Forecasting intraday dynamics in this setting involves several methodological challenges. The data are noisy, high-frequency, and prone to nonstationary patterns and regime shifts. Stocks differ widely in their liquidity profiles, volatility characteristics, and sensitivity to order flow, introducing substantial cross-sectional heterogeneity. Furthermore, modeling involves trade-offs between interpretability and flexibility: econometric models offer transparency and theoretical grounding, while machine learning models can exploit nonlinearities and interactions at the expense of greater opacity. The methodology addresses these challenges through rigorous data cleaning, filtering, and feature engineering to ensure robustness and avoid look-ahead bias, a rolling-window estimation design to simulate real-time forecasting and capture evolving market structure, and cross-sectional stratification by security type, size, and industry to measure heterogeneity in predictability. Parallel evaluation of LASSO and Random Forest enables a balanced assessment of interpretability versus predictive power.

In sum, this introduction positions the thesis at the intersection of market microstructure theory, predictive modeling, and execution strategy design. By linking statistical predictability to economically meaningful signals, such as the midday VWAP gap and order-flow imbalances, it aims to provide both an empirical contribution to the study of intraday market dynamics and a practical tool for institutional execution. The subsequent chapters develop this framework in detail, beginning with a review of the literature on econometric and machine learning approaches to intraday forecasting, followed by data construction, model implementation, and empirical results. The overarching goal is to bridge the gap between predictive modeling advances and actionable decision support for traders operating under real-time constraints.

2 Literature Review

2.1 Traditional Econometric Approaches to Intraday Forecasting

Intraday return forecasting has long been a subject of intense research, primarily due to its implications for trade execution, market efficiency, and liquidity management. Its simplicity lies in modeling the relationship between inputs and the target as a straight line, which provides clear insights into variable influence. Building on this foundation, penalized extensions such as Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, and Elastic Net introduce regularization to prevent overfitting and handle high-dimensional data. LASSO, in particular, is valued for its ability to perform automatic feature selection by shrinking less relevant coefficients to zero (Rastogi et al. 2021).

Several studies have highlighted the effectiveness of LASSO regression in stock price forecasting, demonstrating its superior performance over traditional models across diverse contexts. Roy et al. (2015) introduced a LASSO-based linear regression model that outperformed both Ridge regression and a Bayesian regularized neural network, achieving a test Root Mean Square Error (RMSE) of 2.5401 and a Mean Absolute Percentage Error (MAPE) of 1.4726% for predicting Goldman Sachs stock. Similarly, Rastogi et al. (2021) applied LASSO to predict the NIFTY 50 index[1] using Principal Component Analysis (PCA)-reduced technical indicators and time-lagged features, attaining an RMSE of 14.78 and a MAPE of 2.98% after optimal data preprocessing. Expanding the data scope, Xu et al. (2023) integrated LASSO with external financial variables and demonstrated its predictive advantage across three airline stocks, reaching MAPEs as low as 1.94%. In a more complex architecture, Sheng et al. (2025) combined LASSO-based feature selec-

[1]The NIFTY 50 is an Indian stock market index that represents the float-weighted average of 50 of the largest Indian companies listed on the National Stock Exchange.

tion with the Non-stationary Autoformer model and financial sentiment inputs, improving MAE by 8.75% over the base model and achieving high stability in 10-step forecasts. Collectively, these studies underscore LASSO's robust forecasting capability and adaptability to high-dimensional, and multi-source financial data environments.

Given Elastic Net regression in stock price forecasting, some efforts have been conducted to compare it to other machine learning or penalized regression models. Ding (2024) evaluated the performance of Elastic Net in predicting Apple Inc.'s stock prices over a ten-year period, reporting remarkably high accuracy with an R^2 of 0.998 and a low MSE of 3.244. A distinctive aspect of this study is its coefficient analysis, which identified MACD[2] and EMA50[3] as the most influential predictors, while indicators such as RSI[4], volume, and True Range had negligible influence. In contrast, Sai et al. (2023) compared Elastic Net with an Long Short-Term Memory (LSTM) model for forecasting Nifty stock prices and found Elastic Net to be less effective, with LSTM achieving an R^2 of 0.990 and a lower Mean Absolute Error (MAE) of 10.72. Together, these studies highlight the Elastic Net model's potential for high accuracy and interpretable feature selection in linear settings, while also revealing its limitations in capturing complex nonlinear dynamics when benchmarked against deep learning approaches.

Comparative studies offer mixed but insightful evidence on the relative strengths and limitations of regression-based models in financial forecasting. Schorno (2022) examined probit models augmented with regularization and found that while Ridge, LASSO, and Elastic Net improved in-sample predictive performance for the S&P 500, these gains did not generalize to out-of-sample forecasts, except for the LASSO probit model applied to large-cap firms, which outperformed standard probit benchmarks. In contrast, X. Wang, W. Wang, and S. Zhang (2023) conducted a direct performance comparison across four public companies and concluded that Ridge regression consistently outperformed OLS, LASSO, and Elastic Net in terms of R^2 and MSE, attributing Ridge's superiority to its stable L2 penalty that avoids the over-shrinkage effect observed in LASSO. Meanwhile,

[2]Moving Average Convergence Divergence
[3]50-period Exponential Moving Average
[4]Relative Strength Index

Neba et al. (2023) identified LASSO as the most effective model for forecasting Netflix's adjusted closing prices, outperforming both Ridge and Elastic Net, which ranked second and third, respectively. Collectively, these studies highlight that LASSO's capacity for effective variable selection and model simplification often provides a distinct advantage, particularly in contexts where overfitting or the inclusion of irrelevant predictors could otherwise impair predictive performance.

Early approaches to intraday return forecasting predominantly relied on traditional econometric techniques such as Autoregressive Moving Average (ARMA), Autoregressive Conditional Heteroskedasticity (ARCH), and its extensions like GARCH, EGARCH, and GJR-GARCH (Engle 1982; Bollerslev 1986; Nelson 1991; Glosten, Jagannathan, and Runkle 1993). These models effectively captured volatility clustering and time-varying risk, offering foundational tools for high-frequency return analysis. However, their reliance on linear assumptions and inability to adapt to nonlinear dynamics and structural breaks limited their performance in real-time intraday settings.

While ex-post evaluations based on daily squared returns have often suggested poor volatility forecast performance, it has been shown that this conclusion largely reflects noisy measurement rather than model failure. Using realized-volatility measures constructed from high-frequency intraday returns (i.e., summing squared intraday returns), standard ARCH/GARCH models were found to deliver strikingly accurate day-ahead (daily) volatility forecasts, explaining a large share of latent volatility variation; this also established the value of high-frequency data for improved ex-post measurement and forecast assessment (Andersen and Bollerslev 1998).

Vector Autoregressive (VAR) models extended this literature by modeling interactions between multiple time series such as prices, order flow, and liquidity (Hasbrouck 1991; Engle and Russell 1998). While VAR and regime-switching models (Hamilton 1989; Tsay 1998) provide greater modelling flexibility by allowing relationships among variables to evolve over time, they still fall short in capturing self-reinforcing feedback effects, such as price changes influencing order flow, which in turn further moves prices, along with liquidity-driven regime shifts and the pronounced nonstationarity that characterizes intraday markets.

The rise of high-frequency data offered finer resolution for modeling intraday behaviors. Studies conducted by Heston, Korajczyk, and Sadka (2010) and L. Gao et al. (2018) documented predictable return patterns across time intervals, highlighting intraday momentum effects and institutional trading influence. Yet, these insights were constrained by evolving liquidity conditions and market noise, motivating a shift toward data-driven models better suited to dynamic trading environments.

Out-of-sample concerns were partially mitigated by imposing economically motivated restrictions on predictive regressions (e.g., sign constraints on coefficients and ruling out negative equity premia), which stabilized forecasts and yielded economically meaningful gains despite small R^2, underscoring the value of theory-guided constraints in real-time settings (Campbell and Thompson 2008). Subsequent survey evidence highlights that out-of-sample performance improves when model uncertainty and parameter instability are addressed via theory-based restrictions, forecast combination, diffusion-index factors, and regime-shift models, and practical designs directly relevant for short-horizon execution contexts (Rapach and Zhou 2013). Consistent with these findings, portfolio tests based on stabilized or combined forecasts better translate modest predictability into utility gains once frictions are acknowledged.

2.2 Machine Learning and Hybrid Approaches

Unlike statistical models, Machine Learning (ML) algorithms can capture complex nonlinear dependencies, adapt to dynamic market conditions, and process large volumes of high-frequency data in real time. Notable ML applications in finance focused on neural networks and Support Vector Machines (SVMs) to model stock price movements. G. Zhang, Patuwo, and Hu (1998) demonstrated that Artificial Neural Networks (ANNs) could outperform ARIMA models in short-term forecasting by capturing nonlinear interactions between market variables. Similarly, Kim (2003) applied SVMs to stock return prediction, showing improved classification accuracy over traditional logistic regression models. Illa, Parvathala, and Sharma (2022) found that the Random Forest (RF) model achieved 81.6% and 83.3% accuracy across two setups, outperforming the SVM, for predicting whether the price of a stock will be higher than its price on a given day. These models predicted the short-term trend of the market. Over a long-term period, applying RF to stock market data yielded consistently high accuracy (85–95%) in predicting stock price trends—rising, sideways, or falling—for Apple, Samsung, and GE, outperforming alternatives such as SVM and logistic regression (Zheng et al. 2024).

While these early ML models showed promise, they lacked interpretability and required extensive parameter tuning, making them difficult to implement in live trading environments. A major breakthrough came with the introduction of LSTM networks, a specialized type of Recurrent Neural Network (RNN) designed to handle sequential dependencies in time series data. Huddleston, F. Liu, and Stentoft (2023) demonstrated that LSTM models significantly outperformed traditional time-series methods in predicting intraday returns, as they effectively captured long-term dependencies and nonlinear relationships in stock price movements. Similarly, Fischer and Krauss (2018) showed that deep LSTM architectures, when trained on large financial datasets, could generate consistent predictive signals, making them attractive for algorithmic trading applications.

In terms of clustering, Bini and Mathew (2016) evaluated hierarchical clustering against other methods and found it to be a less effective approach for identifying profitable companies, with its performance ranking below partitioning methods such as K-Means. To address this, Renugadevi et al. (2016) employed Hierarchical Agglomerative Clustering (HAC) and K-Means clustering to create a portfolio of recommended stocks on a short-term basis. The hierarchical agglomerative clustering component was specifically used to create an informative structure from unstructured data. The K-Means algorithm then refined these initial clusters by reducing the sample size, which allowed for a better choice of centroids. This combined method was used to determine stock prices and provide a final list of recommended stocks to investors.

Despite these advancements, ML models faced several practical challenges in intraday forecasting. One of the most pressing issues was data overfitting, where models performed well on historical data but failed to generalize to unseen market conditions. This issue was particularly problematic in high-frequency trading, where market regimes shift rapidly due to liquidity shocks, news events, and institutional trading flows. Additionally, the black-box nature of deep learning models made them difficult to interpret, posing a challenge for institutional traders who require transparency in execution strategies.

To address these concerns, researchers began integrating hybrid models that combined ML algorithms with econometric techniques. Avellaneda and Lee (2010) proposed a framework where traditional stochastic models were augmented with ML-based feature selection, allowing for greater adaptability without sacrificing interpretability. Similarly, X. Gao et al. (2021) implemented ensemble learning methods, blending tree-based models such as RFs and gradient boosting with regression-based approaches to enhance predictive stability across different market conditions. Chen et al. (2024) proposed a hybrid stock price prediction model that combines multi-feature calculation, LASSO feature selection, and a novel cascaded LSTM (Ca-LSTM) network to enhance forecasting accuracy and training efficiency. Their contributions include a focus on data processing by introducing 57 technical indicators for a richer feature set, from which LASSO selects the optimal combination. The Ca-LSTM model is shown to be superior to other time-series prediction models and conventional LSTM approaches, and its integration with an accumulation-based VMD-LSTM model further enhances forecasting accuracy. These hybrid approaches provide a balance between accuracy and interpretability, making them more applicable to institutional trading environments.

A particularly relevant area where these predictive models have been applied is Volume Weighted Average Price (VWAP) execution strategies, widely used by institutional traders to minimize slippage and market impact. Given that intraday return predictability is crucial for optimizing VWAP-based execution, ML models have been explored as tools for improving trade placement and order execution timing. However, existing ML-based VWAP models often rely on full-session data, limit their usefulness for traders making execution decisions at midday without access to end-of-day price movements. This gap has led to an increased focus on adaptive execution algorithms, where machine learning is used not just for forecasting but also for real-time trade execution optimization. Reinforcement learning (RL) approaches, enable models to dynamically adjust execution strategies based on evolving market conditions. These RL-based strategies show promise in minimizing VWAP slippage, yet their reliance on large training datasets and simulation-based learning makes them difficult to implement in highly frequency trading environments. While ML has revolutionized intraday forecasting, it has not fully addressed the practical challenges faced by institutional traders. Existing models often fail to provide reliable noon-to-close predictions, as they are either trained on static historical patterns or assume access to full intraday data. This limitation underscores the need for a real-world applicable model that can predict afternoon price movements using only market information available at noon, a gap that remains largely unaddressed in the literature.

2.3 Intraday Liquidity, Market Microstructure, and VWAP Execution Strategies

Against this backdrop of uneven intraday liquidity, Biais, Hillion, and Spatt (1995) provided foundational evidence on how limit-order book (LOB) states shape short-horizon price dynamics and liquidity supply in a centralized, computerized market, reinforcing why intraday execution decisions must condition on contemporaneous book conditions. They documented that order flow concentrates near the quotes, thin books elicit new orders while thick books trigger trades, and traders rapidly place orders inside the quotes when spreads or quote depth are large, behaviors consistent with priority incentives and transient liquidity events. They also showed that bid/ask quotes adjust asymmetrically after large trades, linking informational shocks to immediate microstructure responses. Taken together, these LOB regularities motivate our use of midday state variables, specifically TCI/TVI and the VWAP gap, to summarize expected drift and impact over the noon-to-close horizon and to inform adaptive execution.

Intraday forecasting is inherently linked to liquidity dynamics and market microstructure features. VWAP-based execution strategies depend heavily on the timing and intensity of liquidity throughout the trading day. Empirical evidence suggests that liquidity is not evenly distributed, with concentration at the open and close, and thinner conditions around midday (Hallam and Olmo 2014). Providing the theoretical foundation for this U-shaped pattern, Admati and Pfleiderer (2015) showed that clustering of discretionary liquidity trading endogenously attracts informed traders, concentrating activity at the open and close and thinning it at midday, which formalizes the intraday liquidity profile referenced above. These variations introduce execution risk, particularly when traders must commit to orders without full-day visibility.

A theoretical foundation is provided by Kyle (1985), whose sequential-auction model shows how private information and noise trading jointly determine market depth, resiliency, and the gradual incorporation of information, implying Brownian price paths with constant volatility, and thereby motivates the use of order-flow–based proxies (depth, spread, OFI) for short-horizon return prediction and execution risk.

Research conducted by Almgren and Chriss (2000) and Cont, Kukanov, and Stoikov (2013) demonstrated how order flow and liquidity indicators influence price movements and execution costs. Subsequent extensions to this framework integrated additional market microstructure features, such as spread dynamics, depth imbalance, and trading pressure, which enhanced VWAP execution strategies but often relied on the unrealistic assumption of continuous, high-quality order book data (Hasbrouck 1991; Engle and Russell 1998). More recent findings by Fallahi (2023) revealed the limitations of such indicators in guiding midday execution under volatile liquidity conditions.

A particularly relevant concept in this domain is Order Flow Imbalance (OFI), also known as trade imbalance, which measures the net pressure of buyer- and seller-initiated trades. Typically computed as a normalized difference between buy and sell market order volumes; OFI serves as a proxy for short-term demand-supply asymmetries and is closely tied to price movements (Chordia, Roll, and Subrahmanyam 2002; Q. Wang et al. 2021). Techniques such as Lee and Ready algorithm (1991) are widely used for classifying trade direction, though recent innovations such as bivariate Hawkes processes, offering more nuanced modeling of trade arrivals and self-excitation patterns in high-frequency settings (Anantha and Jain, 2024).

These microstructure-based indicators offer valuable insights for short-term prediction, especially in the absence of sentiment or macro signals. However, their effectiveness depends on the quality and frequency of trade and quote data, which may not be available in all institutional contexts.

2.4 Forecast-Based Portfolio Construction and Backtesting

Forecast–driven portfolio construction is widely explored in both academic finance and quantitative investment practice, where predictive signals are integrated into allocation rules and evaluated through rigorous backtesting. Ślusarczyk and Ślepaczuk (2025) shows how ARIMA–GARCH and XGBoost return forecasts can serve as expected–return inputs to a Markowitz program on DJIA constituents (2007–2022), spanning 152 strategies with varied estimation windows, rebalancing frequencies, and transaction–cost assumptions. Under certain parameterizations, especially when targeting the Global Maximum Information Ratio, forecast–informed portfolios surpass equal–weight and benchmark portfolios, albeit with sensitivity to tuning choices. Corberán-Vallet et al. (2023) departs from asset-level modeling by forecasting the portfolio value path directly with damped-trend models; the ensuing mean/variance forecasts feed a bi-objective genetic algorithm, thereby bypassing explicit covariance estimation and offering an alternative route from forecasts to actions. Ma and Pohlman (2008) shift attention from point forecasts to conditional return distributions via quantile regression, exploiting distributional asymmetries to design allocations that better account for tail risks—an idea that is especially pertinent intraday, where skewness and kurtosis meaningfully shape realized outcomes. At the interface of forecasting and factors, Chauhan, Alberg, and Lipton (2020) predict forward fundamentals (e.g., next-year EBIT) with deep networks, transform them into forward-looking factor scores adjusted for forecast uncertainty, and rank stocks accordingly; an industrial-grade backtester with costs, slippage, and capacity constraints yields a 17.7% compound annual return and 0.84 Sharpe, materially above a baseline factor strategy. Ta, C.-M. Liu, and Tadesse (2020) applies LSTM forecasts within equal-weight, Monte Carlo, and mean–variance allocation schemes and documents sizable active returns over an S&P 500 benchmark when forecasts are paired with disciplined rebalancing.

Taken together, this literature establishes three points that guide our design: (i) forecasts can be operationalized through multiple allocation paradigms—mean–variance optimization, distribution-aware rules, and ranking-based selection; (ii) credible evaluation requires out-of-sample testing and explicit treatment of frictions; and (iii) realized performance is highly contingent on parameter choices and uncertainty management. In contrast to predominantly daily-horizon studies, the present thesis operates at the *intraday* level, using information observable by 1:00 p.m. to forecast the return to the close. We leverage microstructure-rich predictors, order-flow imbalances, liquidity/turnover, and a relative VWAP measure, and translate model scores into cross-sectional rankings implemented as equal-weight decile long–short portfolios. Relative VWAP is treated not only as an execution benchmark but also as a state variable: when a trader is ahead (behind) schedule and the price drifts above (below) VWAP, the required catch-up trading can reveal latent buying (selling) pressure that is informative about near-term price dynamics. The ranking approach is deliberately simple and robust for intraday horizons, where full mean–variance optimization is fragile. Throughout, we emphasize strict time ordering, rolling estimation, and interpretation of raw backtest metrics as an *upper bound* pending explicit cost and feasibility analyses.

Consistent with credible out-of-sample evaluation, data-snooping risk is addressed by noting that many in-sample 'significant' predictors are no longer significant once multiple testing is controlled, with suitable adjustments for predictive regressions provided by Harvey, Y. Liu, and Zhu (2016).

2.5 Gaps and Research Contribution

The core gap addressed in this thesis is the measurement of latent demand and supply pressure at midday and its translation into actionable noon-to-close forecasts. As discussed in the Introduction, execution risk can pivot around noon as liquidity conditions and order-flow pressure evolve, yet traders must commit to execution trajectories without end-of-day visibility. We posit that the price–VWAP gap observed by 1:00 p.m., together with pre-1:00 p.m. order-flow imbalance, contains information about these hidden forces; jointly modeling these observables provides a practical route to infer the direction and intensity of afternoon pressure.

This thesis contributes a deployable, out-of-sample framework that operationalizes those constructs at 1:00 p.m. and evaluates their informativeness for the remainder of the session. First, we engineer noon-available predictors centered on relative VWAP and trade-imbalance metrics (TCI/TVI), alongside liquidity and momentum controls. Second, we test their predictive value for the return from 1:00 p.m. to close and for afternoon trade imbalances; a controlled exclusion design isolates VWAP's incremental contribution. Third, we compare sparse linear (LASSO) and nonlinear ensemble (Random Forest) models under strictly rolling, out-of-sample estimation, examine cross-sectional heterogeneity via industry and firm-type stratifications, and use clustering to diagnose structural commonalities that may underpin predictive differences. Finally, we translate model scores into equal-weight decile long–short portfolios formed at 1:00 p.m., assess symmetry across long and short legs using mean, dispersion, Sharpe (total and excess), and annualized return, and interpret raw results as pre-cost upper bounds given transaction costs, slippage, short-selling frictions, and operational challenges inherent to intraday equal-weighting. By centering measurement on midday latent pressure and validating noon-observable proxies (price–VWAP gap and order-flow imbalance) in out-of-sample tests, the study directly fills the gap identified in the Introduction and advances an execution-relevant approach to intraday forecasting.

3 Data

The analysis leverages four datasets sourced from the Wharton Research Data Services (WRDS) including: (i) Millisecond Intraday Indicators[1], (ii) CRSP daily stocks[2], (iii) TAQ - Millisecond Consolidated Trades[3], and (iv) WRDS-processed NYSE Matched Trades and Quotes (WCT)[4]. The sample spans from January 2015 to December 2024, covering 10 years of stocks listed across North American Security Exchanges NYSE[5], NASDAQ[6], NYSE American (formerly AMEX)[7], NYSE ARCA[8], and mutual funds (as quoted by NASDAQ). The final dataset comprises 15,175 stocks across more than 21 million stock-day observations (i.e. datapoints).

3.1 Data Sources

3.1.1 Millisecond Intraday Indicators

: This dataset offers more than 190 daily indicators, including liquidity, spread, and transaction-based metrics calculated at millisecond resolution, capturing real-time market microstructure characteristics. Due to the extremely large size of the raw files, all data were stored in a hierarchical `Parquet` format. This columnar, compressed structure—functionally similar in storage efficiency to a `.zip` archive—substantially reduces

[1]`https://wrds-www.wharton.upenn.edu/pages/get-data/nyse-trade-and-quote/millisecond-trade-and-quote-daily-product-2003-present-updated-daily/taq-millisecond-tools/millisecond-intraday-indicators-by-wrds/`

[2]`https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/annual-update/stock-security-files/daily-stock-file/`

[3]`https://wrds-www.wharton.upenn.edu/pages/get-data/nyse-trade-and-quote/millisecond-trade-and-quote-daily-product-2003-present-updated-daily/consolidated-trades/`

[4]`https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/nyse-trade-and-quote-taq/`

[5]New York Stock Exchange

[6]National Association of Securities Dealers Automated Quotations

[7]American Stock Exchange

[8]NYSE Archipelago Exchange

file size while preserving full data fidelity and enabling efficient querying. For context, the `Parquet` archive containing the full ten-year sample of millisecond intraday indicators occupies approximately 40 GB, a fraction of the space that would be required by more common formats such as CSV or uncompressed tables.

3.1.2 CRSP Intraday Daily File (IDF)

: Provides daily open, high, low, close prices, volumes and share classification codes. This data has been used to calculate intraday variables such as intraday return and turnover.

3.1.3 TAQ Consolidated Trades

: Contains nanosecond-level trade data, used to compute precise Volume Weighted Average Price (VWAP) measures up to 1 PM with nanosecond precision for each stock. The VWAP calculation follows:

$$\mathrm{VWAP}_{\mathrm{AM}} = \frac{\sum_{i=1}^{n} P_i Q_i}{\sum_{i=1}^{n} Q_i} \tag{3.1}$$

where P_i and Q_i denote the price and size of trade i, respectively. This dataset was also used to calculate morning (before 1 PM) and afternoon (after 1 pm) values for variables such as size and volume. Due to the exceptionally large size of the TAQ dataset, directly loading all trade-level observations into a local environment for VWAP computation was computationally infeasible. To address this, we implemented an efficient in-database computation strategy WRDS Jupyterhub: VWAP values were calculated directly on the WRDS data servers using SQL aggregate queries, ensuring that only the aggregated results—rather than the full underlying trade data—were extracted and stored. This approach significantly reduced memory and processing requirements while preserving exact VWAP values, and it enabled timely processing of the full sample without compromising accuracy.

3.1.4 WRDS-processed NYSE Matched Trades and Quotes (WCT)

: This dataset aligns each trade with the most recent National Best Bid and Offer (NBBO) quote at the time of execution. It includes trade attributes such as price and size, as well as corresponding quote fields (National Best Bid, National Best Offer, and quote time), enabling detailed market microstructure analysis. In this study, it is specifically used to compute trade imbalance, a directional measure of buyer- vs. seller-initiated trades.

Given the extremely large size of the WCT dataset, direct access and manipulation through conventional local environments were not feasible. Furthermore, due to the computational complexity of applying the Lee–Ready classification to nanosecond-level matched trade–quote pairs, even in-database SQL processing on WRDS was insufficient to handle the required operations. To overcome these limitations, all trade classification and imbalance calculations were executed using `SAS` within `SAS Studio` on the WRDS servers. This approach leveraged the platform's optimized processing capabilities and allowed the computations to be performed without downloading the raw dataset. The full classification and aggregation process required more than 80 hours of continuous computation, underscoring both the scale of the data and the computational intensity of the task.

Trade imbalance is computed by classifying each trade using the Lee–Ready algorithm, which combines a quote test and, when needed, a tick test. In the first step, the quote test, each trade price is compared to the midpoint of the prevailing NBBO. A trade is classified as buyer-initiated if its execution price is above the midpoint, and seller-initiated if below. When the trade price falls exactly at, or within a small tolerance of, the midpoint, making the quote test inconclusive, the algorithm applies the tick test. This test compares the current trade price to the most recent different price: if the price has increased, the trade is classified as buyer-initiated; if it has decreased, as seller-initiated; and if it is unchanged from the prior trade, the classification is inherited from the last price change direction. In this study, two distinct versions of trade imbalance are computed to capture directional market pressure:

- **Trade Count Imbalance (TCI)** is defined in this research as the difference between the number of buyer- and seller-initiated trades, normalized by the total number of trades:

$$\text{TCI} = \frac{\text{Number of Buy Trades} - \text{Number of Sell Trades}}{\text{Total Number of Trades}} \tag{3.2}$$

- **Trade Volume Imbalance (TVI)** also follows a similar structure but uses the aggregated trade volume instead of trade counts:

$$\text{TVI} = \frac{\text{Volume of Buy Trades} - \text{Volume of Sell Trades}}{\text{Total Trade Volume}} \tag{3.3}$$

While the count-based version reflects the directional frequency of trades, the volume-based version captures the directional intensity of trading activity. Both versions are calculated separately for the morning session (9:00 a.m. to 1:00 p.m.) and afternoon session (1:00 p.m. to 4:00 p.m.), based on trade classification using the Lee–Ready algorithm. It is notable that the processing time for this procedure was approximately eighty hours.

3.2 Data Preparation

3.2.1 Data Cleaning

To construct the final dataset, the four data sources described above were merged on the basis of common ticker–date pairs. Given the exceptional scale of the dataset and the heavy processing requirements, all subsequent preprocessing and modelling steps from this stage onward were executed on `Narval`, a high-performance computing (HPC) cluster operated by the Digital Research Alliance of Canada. Narval's large-memory compute nodes and long-runtime capabilities allowed for processing datasets of this size without exceeding memory limits, while its job scheduling system ensured stable execution of tasks requiring many hours to complete. Jobs were submitted and managed using the `SLURM` workload manager, enabling controlled resource allocation, automatic job resumption in case of interruptions, and efficient queue management.

Data merges were performed in memory-efficient chunks to avoid exceeding node memory capacity. Redundant columns arising from multiple data sources were identified and removed, duplicate entries based on ticker and date were eliminated to ensure temporal and cross-sectional uniqueness, and datapoints with missing values in any critical feature were discarded. This workflow ensured computational feasibility and methodological rigour while preserving the integrity of the dataset.

Several preprocessing filters have been applied to ensure data reliability. First, Columns with more than 5% missing values were removed to avoid sparsity-driven noise. Categorical encoding was applied to structural variables such as exchange codes, share codes, and industry identifiers. To address extreme values (outliers), two steps were applied: (i) all numerical variables were winsorized by capping values at the 1st and 99th percentiles, and (ii) stocks trading below \$5 at the market open were excluded to mitigate distortions caused by penny stocks. It should be noted that the outlier handling steps substantially improved model performance, increasing the average R^2 from near zero to above 40%.

To avoid look-ahead bias, we restrict inputs to information observable by 1:00 p.m. on day (t). Any feature that partly or fully depends on afternoon information—e.g., close-based indicators or any statistic aggregating the full trading day—is replaced by its value from the previous trading day (t-1). This lagging scheme is illustrated in Figure 3.1: features are constructed using data up to 1:00 p.m. on day (t) (blue), while variables that would otherwise incorporate post-1:00 p.m. observations are shifted back one day; the model then forecasts the 1:00–4:00 p.m. interval on day (t) (red). As part of feature engineering, we constructed moving-average features at 1-month (21-day), 3-month (63-day), and 1-year (252-day) horizons for key variables (returns, closing prices), capturing short-, medium-, and long-term dynamics. Furthermore, Variables exhibiting high skewness ($|\text{skewness}| > 2$) and positive support were log-transformed to stabilize variance and enhance normality. Finally, we examined the correlation matrix and removed variables with pairwise absolute correlations exceeding 0.90 to avoid multicollinearity in modelling.

A suite of engineered features was developed to extract economically meaningful information from the raw variables. These include return from 1 PM to market close,

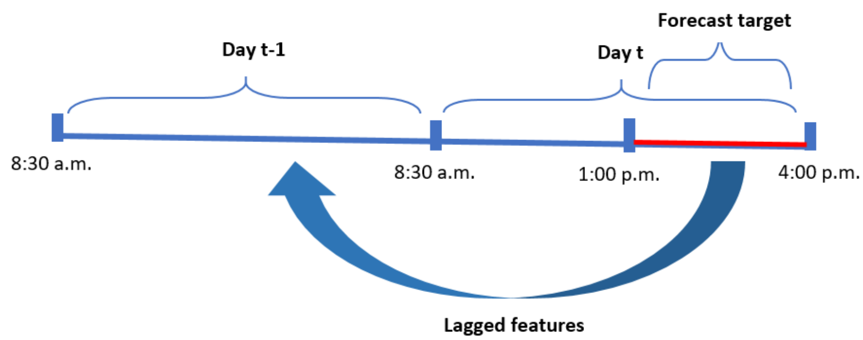


Figure 3.1: Lagging scheme to prevent look-ahead bias. Inputs on day t use only information available by 1:00 p.m. (blue). Any feature requiring afternoon or end-of-day data is replaced by its value from day $t-1$. The model then forecasts the 1:00–4:00 p.m. interval on day t (red).

VWAP-to-Price Ratio, Intraday Momentum and Turnover at 1 PM (defined as the ratio of dollar trading volume to market capitalization).

3.2.2 Subsampling for Model Comparison

To evaluate model robustness and interpret heterogeneity across firm types and market structures, the final dataset has been partitioned into several subsamples based on the following criteria:

1. **Equity Type (Common Stocks vs. ETFs)**
   Equities were divided based on CRSP's share code: entries with a share code of 10 and 11 were classified as common stocks, while those with a share code of 73 were labeled as Exchange-Traded Funds (ETFs). This distinction enabled comparative modeling between traditional equities and passive index-tracking instruments.

2. **Market Capitalization (Size-Based Stratification)**
   At the start of each calendar year, NYSE-listed equities were sorted by market capitalization as of the first trading day. The 30th and 70th percentiles of this cross-sectional distribution were recorded as size thresholds for that year. These NYSE-based thresholds were then applied to all equities across the five markets in the sample to ensure a consistent size classification framework. Stocks with market

capitalization below the 30th percentile threshold were labeled small-cap, between the 30th and 70th percentiles mid-cap, and above the 70th percentile large-cap. Once assigned, a stock's size category remained fixed for the remainder of that year.

3. **Industry Classification (Fama-French 12 Sectors)**
   Using SIC codes[9] provided in CRSP, we mapped each firm to one of the 12 major industry sectors as defined by the Fama-French classification.[10] This mapping allows for the control of industry-specific return patterns and microstructure behavior in downstream modeling.

A set of target and predictor variables was constructed to capture the intraday informational environment relevant for forecasting stock price dynamics. The target variables comprise the evening return (1 p.m. to market close), the evening trade count imbalance (TCI) (Eq. 3.2), and the evening trade volume imbalance (TVI) (Eq. 3.3). The predictor set includes the morning TCI, morning TVI, morning relative VWAP (Eq. 5.1), morning momentum (see appendix B, Eq. 1), and turnover at 1 p.m. (see appendix B, Eq. 2). These engineered predictors proved highly informative; all of them were identified among the top-ranked variables in the clustering analysis of stock price behaviour (see Table 6.1), underscoring their effectiveness in differentiating trading patterns across equities.

3.3 Final Dataset Characteristics

Table 3.1 presents the descriptive statistics for key continuous variables in the final dataset after preprocessing and integration of all four WRDS sources. It includes central tendency measures (mean, median), dispersion metrics (standard deviation, interquartile range), and extremes (minimum and maximum). Variables such as VWAP, total size, total value, volume, and order imbalance reflect the primary market microstructure measures used

[9] Standard Industrial Classification Code (siccd)

[10] Industry definitions are sourced from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html`.

in the study, summarizing their scale and variability across the sample period. This provides context for interpreting model inputs and assessing the magnitude of market activity captured in the data.

Table 3.2 summarizes the feature composition of the final dataset by category, listing the number of variables in each group and example features. Categories span price/return measures, transaction-based measures, liquidity, informativeness, volatility, temporal attributes, and retail/institutional order flow. This table outlines the breadth of information available for modelling, clarifies the data's multidimensional structure, and demonstrates how different market dimensions are captured for predictive analysis.

Table 3.1: Table 3.1: Descriptive statistics of key variables in the final dataset.

Variable	Mean	Std Dev	Min	25%	Median	75%	Max
VWAP	40.05	129.09	0.00	8.15	20.85	44.31	28,767.30
Total Size	61,290.33	411,106.10	1.00	100.00	1311.00	26,087.00	314,462,300.00
Total Value	50,680,680.76	522,084,400.00	0.00	224,578.57	1,656,404.02	14,624,598.96	168,363,700,000.00
Return	0.00	0.04	−0.98	−0.01	0.00	0.01	39.73
Volume	958,439.87	5,566,376.00	0.00	18,079.00	106,794.00	512,608.00	2,655,406,000.00
Quoted Spread	0.12	0.28	0.00	0.02	0.05	0.11	5.00
Order Imbalance	0.25	0.26	0.00	0.06	0.14	0.34	1.00

Table 3.2: Overview of feature categories distribution in the final dataset

Categories	**Number of Variables**	**Example Features**
Price and Return Measures (Indicators + CRSP)	∼50	Closing Price (CRP), Opening Price (QPR), CRSP Close, CRSP Return
Transaction: Trade, Volume, Value (Indicators + TAQ-derived)	∼65	Total Number of Trades, Total Volume, VWAP at 1pm, Total Size by 1pm, Total Value by 1pm
Liquidity Measures (Indicators)	24	Quoted Spread (Dollar/Percent), Effective Spread
Informativeness Measures (Indicators)	5	Order Imbalance, Price Impact Lambda
Volatility Measures (Indicators)	8	Trade-based Volatility, Quote-based Volatility
Symbol, Time, and Other Attributes (Indicators)	24	Date, Symbol, Closing Time, Trade Sequence Number
Retail and Institutional Order Flow (Indicators)	39	Retail Buys/Sells (Volume and Value)
Total	∼221–228	—

4 Forecasting Models

4.1 Lasso Modeling

4.1.1 Modeling Framework

To evaluate the predictive content of midday indicators for afternoon stock price movement, a Lasso regression framework is adopted within a rolling-window estimation strategy. This choice is motivated by the methodological requirements of the problem. A simple OLS regression is unsuitable, as the number of predictors is large relative to the effective sample size and multicollinearity is substantial. Elastic Net is also avoided because, in preliminary tests, its combined L_1–L_2 penalty tended to shrink nearly all coefficients toward zero, obscuring the identification of truly persistent predictors. Time-series models designed for step-ahead forecasting are inappropriate here, as the dataset does not form a consistent, evenly spaced time series, our focus is solely on forecasting the "evening" point using midday cross-sectional information. Lasso's L_1 regularization directly addresses these challenges by performing simultaneous coefficient shrinkage and variable selection, mitigating overfitting, and enhancing interpretability in high-dimensional, noisy financial settings. These properties are essential for isolating which liquidity and order-flow measures retain stable explanatory power across time and market segments.

Formally, the Lasso estimator solves the following optimization problem:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - X_i^\top \beta \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{4.1}$$

where $\lambda \geq 0$ controls the amount of ℓ_1 regularization.

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - X_i^\top \beta \right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t, \tag{4.2}$$

where $t > 0$ determines the ℓ_1-norm bound.

where y_i denotes the response variable, X_i is the predictor vector for observation i, β is the coefficient vector, and $\lambda \geq 0$ is the regularization parameter controlling the strength of the L_1 penalty. Larger values of λ increase the shrinkage applied to the coefficients, driving more of them to exactly zero and thereby performing implicit variable selection.

4.1.2 Rolling Forecast Procedure

A rigorous rolling forecast has been implemented to reflect the sequential nature of investment decision-making. The model is trained on a fixed three-year historical window and tested on the subsequent year, advancing the window in annual increments from 2018 through 2024. This setup ensures that each forecast relies solely on information available at the time of prediction, thereby eliminating look-ahead bias.

To account for structural heterogeneity, the sample is stratified by equity type, market capitalization group, and industry classification. For each subset, the regularization hyperparameter (λ) is re-tuned at the start of each calendar year using five-fold cross-validation on the preceding three-year training window (rolling-origin), and the selected λ is then held fixed for that subset throughout the year to ensure comparability across daily test windows. In principle, a fully optimal protocol would re-select λ at *every* rolling step via CV on the contemporaneous training fold (i.e., nested hyperparameter search within walk-forward evaluation). However, with daily re-estimation over 2018–2024, a three-year rolling training span, and $k=5$ folds, proved computationally infeasible on the Narval cluster (faced out-of-memory errors). The adopted *annual* re-tuning thus represents a tractable, variance-reducing compromise that limits hyperparameter staleness while avoiding leakage, and preserves year-over-year comparability of R^2 and error profiles across rolling windows.

Each forecast iteration applies a standardized preprocessing pipeline: (i) mean imputation for missing values to preserve sample size without distorting distributional properties, (ii) z-score scaling to place predictors on a comparable scale, and (iii) Lasso regression using the previously tuned λ. Model performance is evaluated using MSE, MAE, and out-of-sample R^2, allowing for consistent assessment across models and periods.

The pipeline is implemented in Python using the `scikit-learn` library, ensuring reproducibility and transparency of the modeling process. Data preprocessing and modeling steps are encapsulated in modular scripts, enabling consistent execution across all subsets and windows. Identifier columns and categorical labels are systematically excluded from the predictor matrix to prevent data leakage.

Beyond point forecasts, the analysis examines the stability of variable selection across rolling windows, highlighting indicators whose coefficients remain persistently non-zero. This longitudinal perspective is critical for addressing the research objective of identifying robust drivers of intraday price behavior. As such, the Lasso results not only serve as a benchmark for more flexible machine learning methods presented later in this chapter but also provide economically interpretable insights into market microstructure dynamics.

The adoption of the Lasso regression with a rolling forecast framework directly addresses the core research questions by rigorously testing whether midday liquidity, order flow, and VWAP-related indicators possess stable predictive power for subsequent intraday returns. The method's embedded feature selection aligns with the objective of isolating the most informative predictors, while the rolling design ensures temporal validity and realistic investment applicability. Stratifying by equity type, market capitalization, and industry enables the analysis to capture cross-sectional heterogeneity, thereby meeting the methodological relevance criterion. The preprocessing, parameter tuning, and consistent evaluation across multiple subsets and years demonstrate the utmost rigour in data handling and model estimation. Collectively, these design choices ensure that the analysis not only delivers accurate forecasts but also produces interpretable results that fully satisfy the stated objectives and contribute meaningful insights into the drivers of intraday market movements.

4.1.3 Lasso Forecasting Results

The out-of-sample R^2 patterns from the rolling Lasso estimation reveal clear differences in predictability across the three targets, afternoon return, trade volume imbalance (TVI), and trade count imbalance (TCI), and across market capitalisation, security type, and industry classifications. The most striking and consistent finding is the marked disparity in forecastability between returns and order-flow measures. While TVI and TCI typically exhibit R^2 values in the range of 0.50–0.62, indicating substantial and persistent predictability, the corresponding figures for return are an order of magnitude lower, generally between 0.06 and 0.18 (see table 1). This difference holds across all subgroup definitions and throughout the 2018–2024 evaluation period, underlining that midday information provides far greater insight into the evolution of afternoon order flow than into directional price changes. It is notable that the processing time for this procedure was approximately ten days.

Across time, order-flow predictability appears relatively stable, though not entirely invariant. Many series exhibit a modest trough around 2020–2021, followed by partial recovery in 2022–2023. This pattern is observed across size, security type, and industry groups, and may reflect the influence of broad market-wide factors, potentially including shifts in liquidity provision and changes in retail participation during the pandemic period, on the persistence of intraday flow patterns. By 2024, R^2 values for TVI and TCI remain high, though slightly below their early-sample peaks, particularly in ETFs. The return forecasts display a noisier version of the same shape, with a pronounced drop in 2019, a mild rebound into 2022, and softening thereafter.

Size-based patterns are modest for returns, where all three capitalisation groups cluster at similarly low R^2 levels, but are more pronounced for the order-flow measures. Small-cap stocks tend to post the highest TVI and TCI R^2 values, especially early in the sample, consistent with greater persistence in their liquidity imbalances, possibly due to a higher share of less-informed trading and more concentrated market-making activity. Large and mid-cap stocks track closely together, with slightly lower but still robust predictability in

Figure 4.1: Out-of-sample R^2 Comparison for Lasso model forecasts of afternoon Return, Trade Count Imbalance (TCI), and Trade Volume Imbalance (TVI) across stock types, market capitalization groups and industries (rolling test years from 2018 to 2024).

TVI and TCI.

Comparing common stocks and ETFs reveals a distinct trajectory. For returns, both series remain weakly predictable, with minor fluctuations over time. For order-flow measures, ETFs start the sample with the highest predictability, TCI R^2 exceeding 0.63 in 2018, before undergoing a gradual, monotonic decline toward the mid-0.50s by 2024. Common stocks, by contrast, maintain a narrower range with smaller drifts, suggesting a more stable microstructural environment. The ETF decline may reflect the erosion of mechanical flow patterns from creation/redemption activity as competition among arbitrageurs increased and as execution strategies adapted to post-2020 conditions.

Industry disaggregation confirms the overall ranking of target variables while revealing some heterogeneity in level. For returns, a handful of sectors, notably Chemicals, Durables, Business Equipment, and Manufacturing, occasionally post R^2 near 0.18–0.19, especially around 2022, whereas sectors like Financials and Telecom generally sit at the lower end. In contrast, TVI and TCI are strong across the board, with sectors such as Business Equipment, Chemicals, Health, Energy, and Other frequently in the upper half of the range and maintaining values near or above 0.58 for extended periods. Utilities and Telecom tend toward the lower end, but even here, order-flow predictability remains materially high.

Taken together, these results demonstrate that the Lasso model consistently extracts robust predictive signals for intraday order-flow variables, with high R^2 across time and market segments. The persistence and stability of TVI and TCI forecasts indicate that midday liquidity and trade activity patterns carry substantial information about the state of the market later in the day, even if that information does not translate directly into predictable returns. The modest time variation, common across segments, suggests that changes in market-wide conditions can affect the strength of these relationships, but without eliminating them. These findings directly inform the execution problem introduced in this thesis. The relatively weak predictability of returns, contrasted with the strong and persistent forecastability of TCI and TVI, implies that the information available by midday is more reflective of latent trading pressure than of directional price expectations. In other words, while price changes between noon and close remain largely driven by stochastic shocks, liquidity and order-flow variables exhibit structured and repeatable behavior that traders can exploit. This distinction is central to the hypothesis that midday VWAP deviations embody execution-relevant information: they do not deterministically forecast prices, but they proxy for temporary supply–demand imbalances that shape short-horizon price impact and fill quality. The Lasso framework, by isolating these stable linear relationships, thus establishes a baseline understanding of how the midday market state encodes the "urgency" and asymmetry of liquidity that VWAP-based strategies must respond to in real time.

4.2 Random Forest Modeling

4.2.1 Modeling Framework

As a nonparametric alternative to Lasso, we estimate Random Forest (RF) regressions to capture nonlinear relationships and interactions among predictors. Random Forests are ensemble learners that average the predictions of many decorrelated decision trees built on bootstrap samples and random feature subsets, thereby reducing variance and improving out-of-sample stability. Formally, for B trees $\{T_b\}_{b=1}^{B}$ grown on bootstrap samples, the RF prediction for a new observation x is

$$\hat{y}_{\mathrm{RF}}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{4.3}$$

The randomisation in sampling and feature selection mitigates overfitting and makes the estimator robust in high-dimensional, noisy settings typical of intraday data.

4.2.2 Rolling Forecast Procedure and Implementation

To ensure full comparability with the Lasso benchmarks, we employ the same rolling-window design: for each test year from 2018 to 2024, the model is trained on the preceding three calendar years and evaluated on the following year. All preprocessing and estimation steps are executed strictly within each fold to avoid look-ahead bias.

The modeling pipeline is implemented in `scikit-learn` as a two-step estimator comprising mean imputation (calibrated on the training slice only) followed by a Random Forest regressor. Because decision-tree splits are invariant to monotonic transformations, no feature scaling is applied. Predictors are restricted to numerical variables; identifier fields and administrative columns (`date`, `Year`, `ticker`, `hsiccd`, `shrcd`, `market_cap`) are excluded from the feature matrix to prevent leakage.

Consistent with the code, hyperparameters are fixed ex-ante across all folds and subsets: we use $B = 100$ trees with parallel fitting (`n_jobs=-1`) and a fixed random seed for reproducibility. Although Random Forest studies often use 500–1000 trees to further reduce variance, our dataset's per-fold memory footprint repeatedly triggered out-of-memory (OOM) failures on the Narval cluster; we therefore cap the ensemble at $B = 100$ trees to ensure stable execution. Other tree-growth parameters are left at their library defaults. This choice prioritises stability and transparency over fold-by-fold re-tuning and avoids contaminating the temporal comparison with changing model capacity.

For each subset–year combination we record mean squared error (MSE), mean absolute error (MAE), and out-of-sample R^2. In addition, we extract the model's embedded impurity-based importances (mean decrease in impurity, MDI) for all predictors in that fold. Storing these importance vectors across folds enables us to study which variables repeatedly contribute to predictive accuracy and whether their relevance shifts across time and market segments. All metrics and importances are written to structured CSV files, and a predicted-versus-actual scatter plot is produced per subset and target for visual diagnostics.

The RF specification complements the Lasso benchmark by relaxing linearity and sparsity assumptions and allowing for rich nonlinearities and interaction effects among liquidity, order-flow, and intraday price variables. Whereas Lasso yields a parsimonious, interpretable linear signal, the Random Forest probes whether additional structure can be harnessed for incremental out-of-sample gains. The joint evidence from these two model classes therefore provides a comprehensive assessment of the predictive content of midday indicators for afternoon market dynamics.

4.2.3 Random Forest Forecasting Results

The Random Forest (RF) estimates uncover substantial and stable predictability for order-flow measures and more moderate, yet economically meaningful, predictability for returns. Averaged across all subgroups and years (2018–2024), the out-of-sample R^2 is 0.771 for trade count imbalance (TCI; range 0.641 to 0.850), 0.659 for trade volume imbalance (TVI; range 0.541 to 0.758), and 0.168 for the afternoon return (range -0.037 to

0.246) (see table 1). These magnitudes indicate that nonlinear structure and interactions captured by the RF translate into very strong forecasts for the state of order flow and into nontrivial, albeit smaller, gains for directional returns.

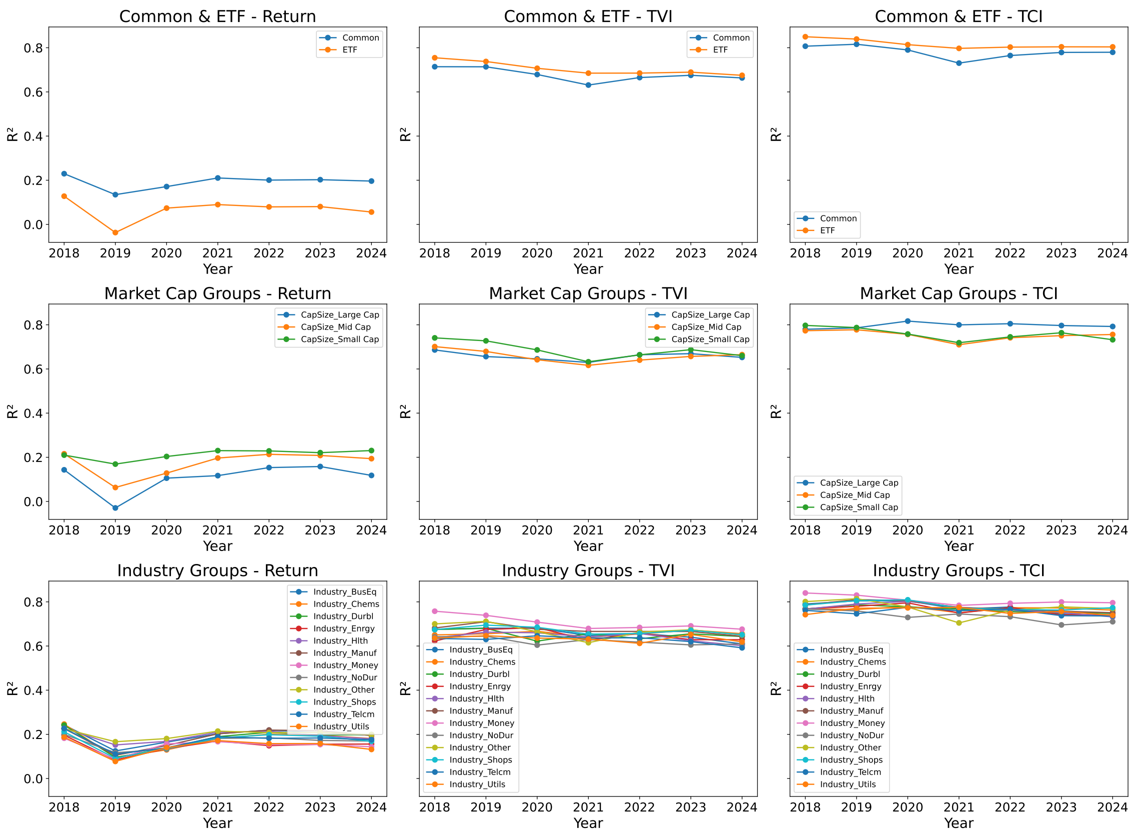


Figure 4.2: Out-of-sample R^2 Comparison for Random Forest forecasts of afternoon Return, Trade Count Imbalance (TCI), and Trade Volume Imbalance (TVI) across stock types, market capitalization groups and industries (rolling test years from 2018 to 2024).

Time-series profiles are broadly consistent across targets. For TCI, average yearly R^2 peaks in 2019 at 0.794 and trends gently lower thereafter, settling near 0.758 in 2024. TVI follows a similar path, from 0.686 in 2019 to 0.641 in 2024, with a trough around 2021. Return predictability displays a noisier pattern: a dip in 2019 (0.099) is followed by a steady improvement through 2021–2022 (about 0.185) and a mild softening in 2024 (0.171). The common movement in order-flow targets suggests regime-level changes in intraday participation and inventory management, which compress persistence during 2020–2021 without eliminating it.

Cross-sectional heterogeneity aligns with microstructure intuition. By security type, ETFs exhibit very high order-flow predictability (mean TCI $R^2 = 0.816$; mean TVI $R^2 = 0.705$) but much weaker return forecasts (mean $R^2 = 0.067$ with a negative reading in 2019), whereas common stocks sustain comparably strong order-flow predictability (TCI 0.781, TVI 0.677) and materially higher R^2 (mean $R^2 = 0.192$) that is economically significant. This gap suggests that mechanical and arbitrage-related ETF flows are persistent and thus forecastable as flows, but these patterns translate only weakly into subsequent price changes at the horizon considered. By market capitalisation, small caps lead in return forecasting (mean $R^2 = 0.213$ versus 0.174 for mid caps and 0.109 for large caps), consistent with higher frictions and more persistent intraday momentum in the tail of the size distribution (these R^2 differences are economically significant). For order flow, leadership splits by measure: TVI is highest in small caps on average (0.685), whereas TCI is maximised in large caps (0.797), reflecting differences in participation intensity and the count-versus-volume composition of trading across size buckets.

Industry results preserve the same ranking of targets while revealing informative dispersion in levels. For TCI, the highest average R^2 are observed in *Money/Financials* (0.807), *Shops* (0.782), and *Manufacturing* (0.779), with *Telecom*, *Utilities*, and *Nondurables* in the lower part of a still-elevated range (down to about 0.734). TVI peaks in *Money/Financials* (0.705), followed by *Manufacturing* (0.673) and *Shops* (0.670), and is relatively lower in *Nondurables*, *Telecom*, and the *Unknown* residual category. For returns, the strongest averages occur in *Other* (0.200), *Business Equipment* (0.194), *Health* (0.192), and *Manufacturing* (0.188), while *Utilities*, *Energy*, and *Money/Financials* sit at the weaker end (roughly 0.146–0.149). The inter-industry R^2 variations are economically significant and these sectoral patterns are consistent with differences in the timing and concentration of news, the structure of liquidity provision, and intra-industry trading conventions.

Embedded variable-importance diagnostics corroborate the behavioral reading. For the two order-flow targets, the complementary imbalance measure is the dominant predictor on average: TCI forecasts load most heavily on the same-period TVI, and TVI forecasts load most heavily on TCI, with secondary contributions from buy–sell ratio metrics and morning-session imbalances. For returns, the most influential signals combine market-wide return factors, intraday momentum measured from the open to 1 p.m., and contemporaneous order-flow variables, indicating that nonlinearity arises primarily through interactions between participation intensity and broader market conditions rather than through a single stand-alone predictor. Together, these importance patterns explain why RF gains are concentrated in flow targets and are more modest, but still present, for returns.

From a practical perspective, the RF results imply that intraday decision rules should be conditioned on the predicted state of order flow. The high and persistent R^2 for TCI and TVI across years, security types, and industries make these forecasts reliable state variables for sizing and timing. The stronger return predictability in small caps and certain industries suggests where directional exposure is most likely to pay off, while the consistently weak return R^2 for ETFs cautions against applying the same return-oriented rules to ETF universes. Finally, these findings are not driven by thin subsamples: average test set sizes per subgroup–year are large, and the rolling design ensures that all estimates rely strictly on information available at the forecast origin.

In sum, the Random Forest extracts rich nonlinear structure from midday information. It delivers very strong forecasts for afternoon order-flow imbalances and nontrivial predictability for returns, with coherent time variation and economically interpretable cross-sectional differences. Within the same predictor space, the Random Forest model extends the linear benchmark by revealing interaction-driven predictability that Lasso cannot capture. The improvement in accuracy which was economically significant, does not stem from new information, but from the model's ability to combine existing variables—such as the VWAP gap, turnover, and order imbalances—in nonlinear and conditional ways. Interpreted through the execution lens, this means that the market's response to a given VWAP deviation depends jointly on liquidity depth and trade imbalance intensity. For example, a positive deviation may lead to reversion when liquidity is thin but continuation when trading pressure remains buyer-dominant. These results thus support the core hypothesis that midday VWAP deviations contain predictive information whose strength varies with microstructural context, providing a data-driven foundation for adaptive VWAP execution rules that adjust to prevailing conditions.

5 VWAP Importance Analysis

5.1 Experimental Design and Metric

To quantify the incremental value of VWAP-related information, we perform a controlled exclusion test. In the baseline specifications (Sections 4.1 and 4.2), the predictor set includes a transformed morning VWAP defined as

$$\text{Relative VWAP} \equiv \log(\text{VWAP}_{\text{morning}}) - \log(P_{\text{open}}), \tag{5.1}$$

where $\text{VWAP}_{\text{morning}}$ is computed from the open to 1:00 p.m. and P_{open} is the opening price. We refer to this variable as *Relative VWAP*. We then remove Relative VWAP from the feature set and re-estimate the models with no other change to the rolling design, preprocessing, or hyperparameters.

To compare forecast accuracy on a consistent scale across years and subgroups, we compute a *relative out-of-sample performance* measure:

$$\text{Rel}\,R^2 = \frac{R^2_{\text{with RVWAP}} - R^2_{\text{without RVWAP}}}{R^2_{\text{with RVWAP}}}, \tag{5.2}$$

so that positive values indicate improvements from including Relative VWAP and negative values indicate deterioration. Because return R^2 can be small, we interpret relative changes alongside the underlying levels and report absolute differences in the supplementary figures where relevant.

This section directly tests the thesis's execution-motivated hypothesis by examining whether midday deviations from VWAP, conditional on liquidity and order-flow information, contain forward-looking predictive value for noon-to-close returns and imbalances. The analysis quantifies VWAP's incremental contribution to predictive accuracy through controlled inclusion–exclusion tests within the rolling Lasso and Random Forest frameworks. This design allows us to determine whether VWAP acts as an actionable state variable for adjusting execution aggressiveness at 1:00 p.m. rather than a descriptive benchmark.

5.2 Lasso: Results and Discussion

The Lasso results indicate that Relative VWAP contributes *modestly and heterogeneously* to out-of-sample predictability. For **afternoon returns**, several industries exhibit small but systematic gains from including Relative VWAP (Business Equipment, Chemicals, Durables, Health, and Manufacturing), with intermittent negatives in Telecom, Utilities, and Nondurables concentrated in 2020 (Figure 10). These patterns suggest that Relative VWAP carries information that complements intraday momentum and order-flow variables in sectors where trading is more inventory- or benchmark-driven, while the large 2020 drawdowns are consistent with the pandemic-related microstructure break that altered intraday participation and execution practices.

For **order-flow targets**, effects are near zero on average and concentrated around the 2019–2021 window. Trade *volume* imbalance shows mostly negligible changes, with a few sector-year improvements (e.g., Money/Financials in 2020; Manufacturing in 2024) and several transient declines in 2020 for Health and Nondurables (Figure 5.4). Trade *count* imbalance exhibits a similar picture: small positives in Business Equipment and Chemicals offset by sizeable negatives in 2020 for Nondurables and Telecom (Figure 5.5). Taken together, these results imply that Relative VWAP adds limited incremental information for forecasting the *state of flow* once other midday indicators are present, and that its marginal contribution is sensitive to regime shifts.

Cross-sectional averages by **market capitalisation** and **security type** confirm the sector patterns. The cap-size summary shows modest average gains for returns, with minimal average impact on the order-flow measures (Figure 5.2). By type, *common stocks* display slightly larger relative improvements for return than *ETFs*, while effects for the flow targets remain close to zero in both universes (Figure 5.1). This is consistent with Relative VWAP being more informative when idiosyncratic inventory pressure and benchmark-

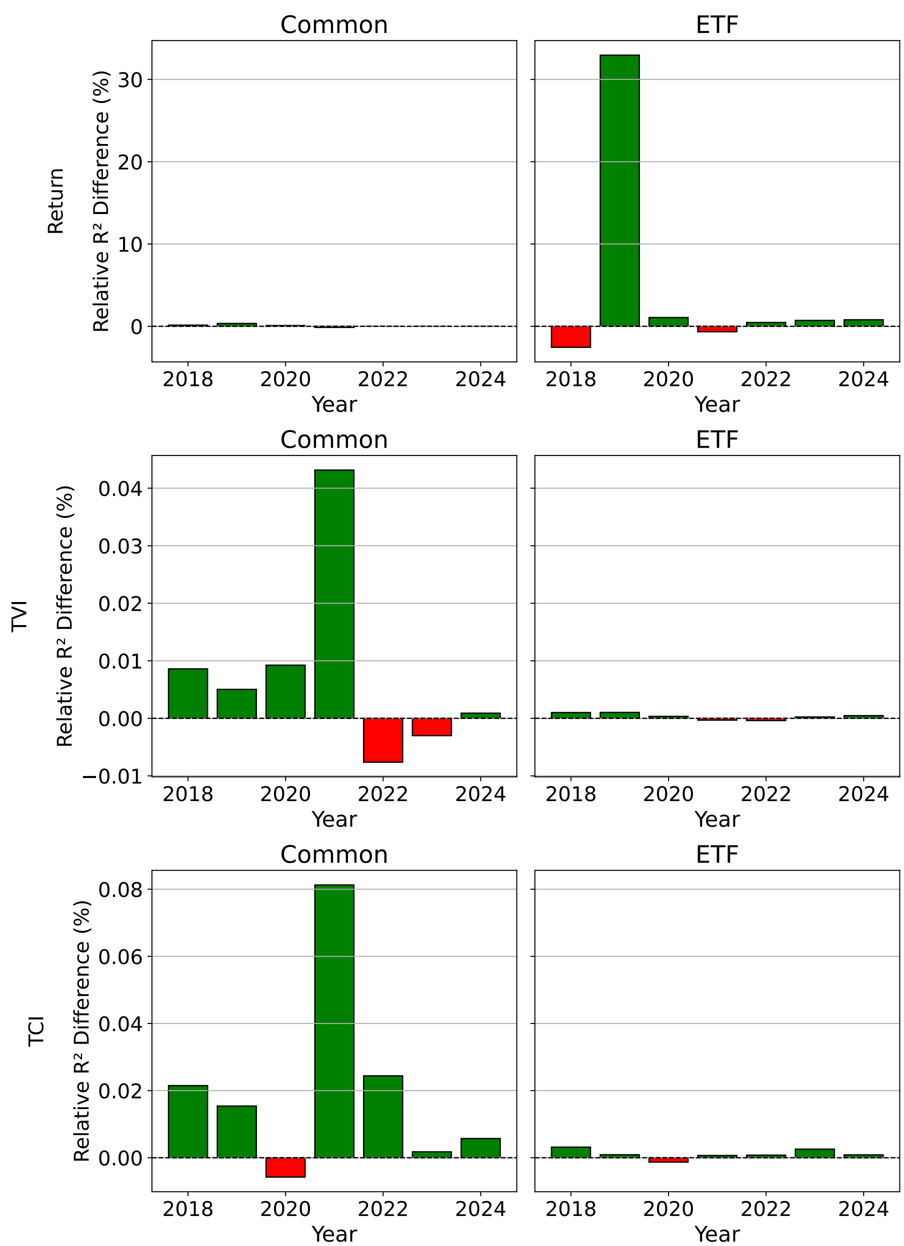


Figure 5.1: Relative out-of-sample R^2 gain from including VWAP in the Lasso model, disaggregated by security type (common stock vs. ETF) and target variable (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

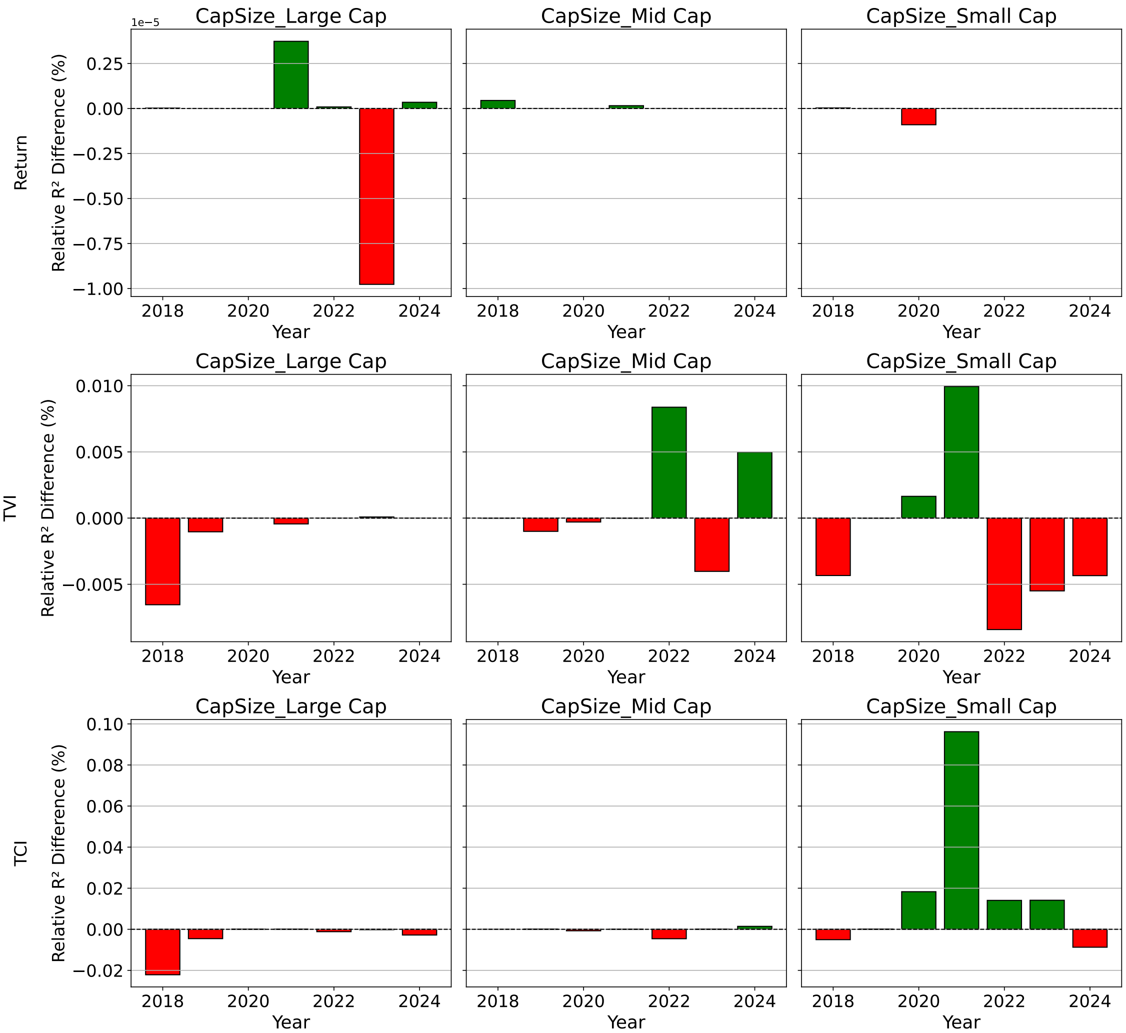


Figure 5.2: Relative out-of-sample R^2 gain from including VWAP in the Lasso model, disaggregated by market capitalization (small, mid and large cap) and target variable (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

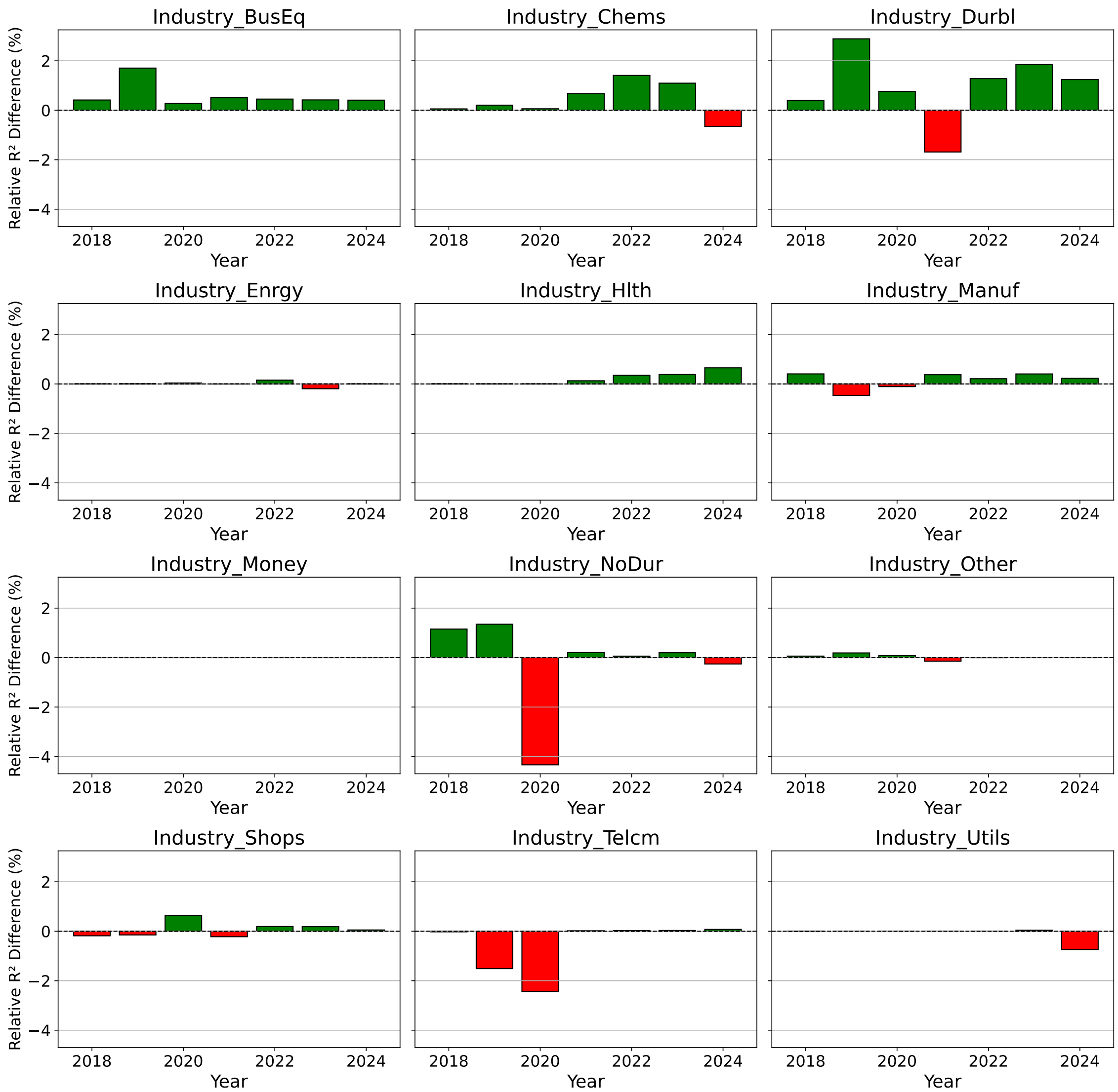


Figure 5.3: Relative out-of-sample R^2 gain from including VWAP in the Lasso model for forecasting returns, disaggregated by industry groups, for test years 2018–2024.

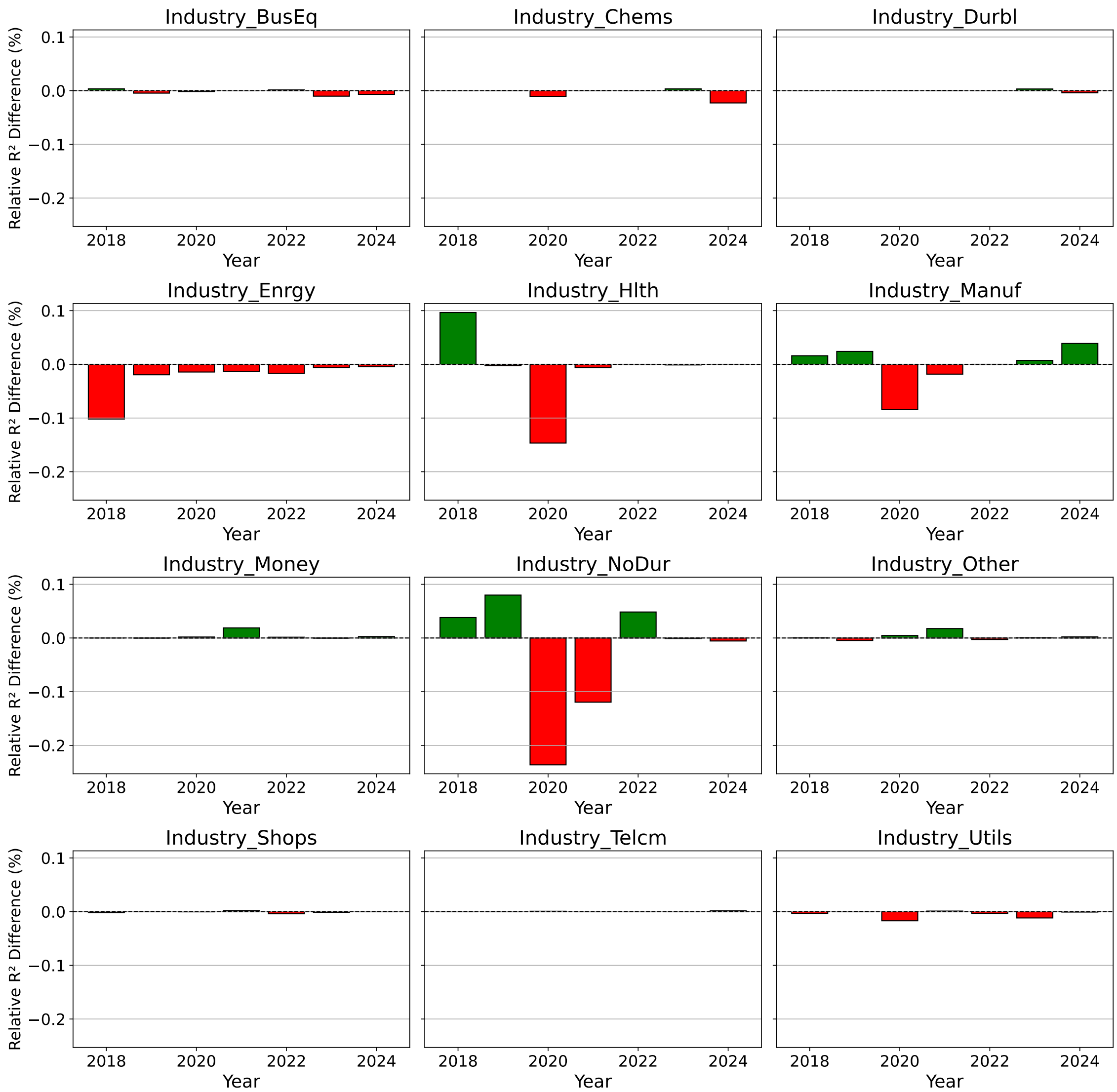


Figure 5.4: Relative out-of-sample R^2 gain from including VWAP in the Lasso model for forecasting Trade Volume Imbalance (TVI), disaggregated by industry groups, for test years 2018–2024.

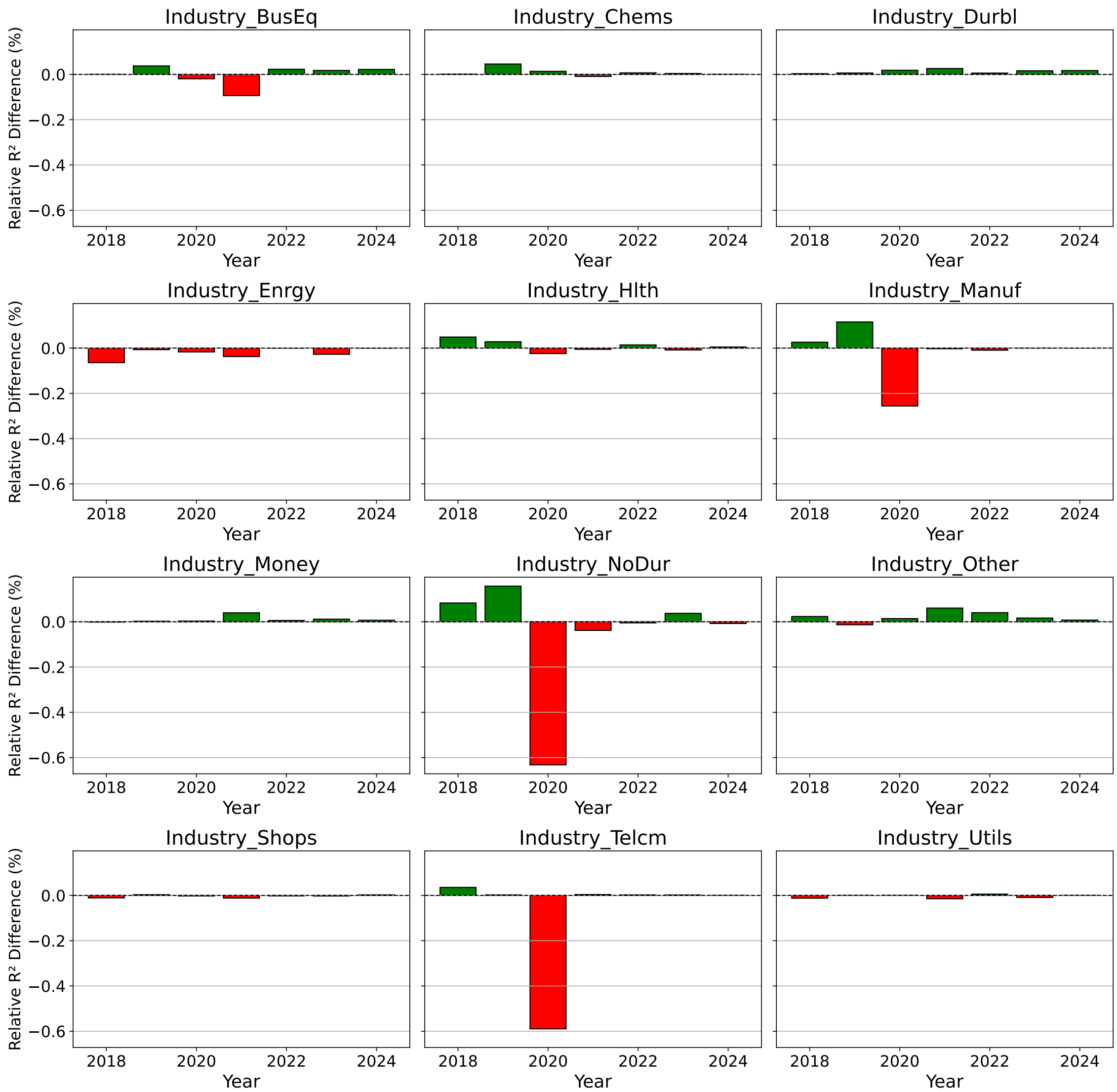


Figure 5.5: Relative out-of-sample R^2 gain from including VWAP in the Lasso model for forecasting Trade Count Imbalance (TCI), disaggregated by industry group, for test years 2018–2024.

tracking behaviours matter, and less so where mechanical creation/redemption flows dominate.

Overall, the Lasso exclusion test indicates that Relative VWAP provides at most a small, context-dependent increment for return predictability and adds little to flow predictability (TCI/TVI) once other midday variables are included. The clustering of negative contributions in 2020 suggests regime sensitivity of VWAP-based signals. The weak and unstable VWAP coefficients imply that the link between midday price positioning and afternoon outcomes is not well captured by additive linear effects. For execution, this cautions against linear schedule rules keyed solely to the VWAP gap: scaling participation in proportion to the gap is unlikely to capture the nonlinear feedback among liquidity, order flow, and execution urgency that drives post-noon performance.

5.3 Random Forest: Results and Discussion

We repeat the exclusion test for the Random Forest model, reporting relative out-of-sample R^2 gains from including Relative VWAP by subgroup and year. Since Random Forest can capture nonlinearities and interactions that Lasso cannot, we expect clearer and more structured effects if VWAP contributes predictive information beyond order-flow and momentum indicators.

The corrected Random Forest results show that the incremental contribution of Relative VWAP is strongest and most stable for return forecasts, but far weaker for order-flow variables. In Figure 5.6, common stocks consistently display small but positive gains in returns, while ETFs exhibit volatile patterns: strong negatives in 2019, offset by moderate positives in other years. This contrast suggests that VWAP embeds information more aligned with the trading dynamics of individual equities than with the aggregated baskets of ETFs.

By market capitalisation (Figure 5.7), small-cap returns benefit the most, with repeated positive gains throughout the period. Mid-cap effects are less pronounced but remain positive in several years. Large caps show sharp positive impacts early in the

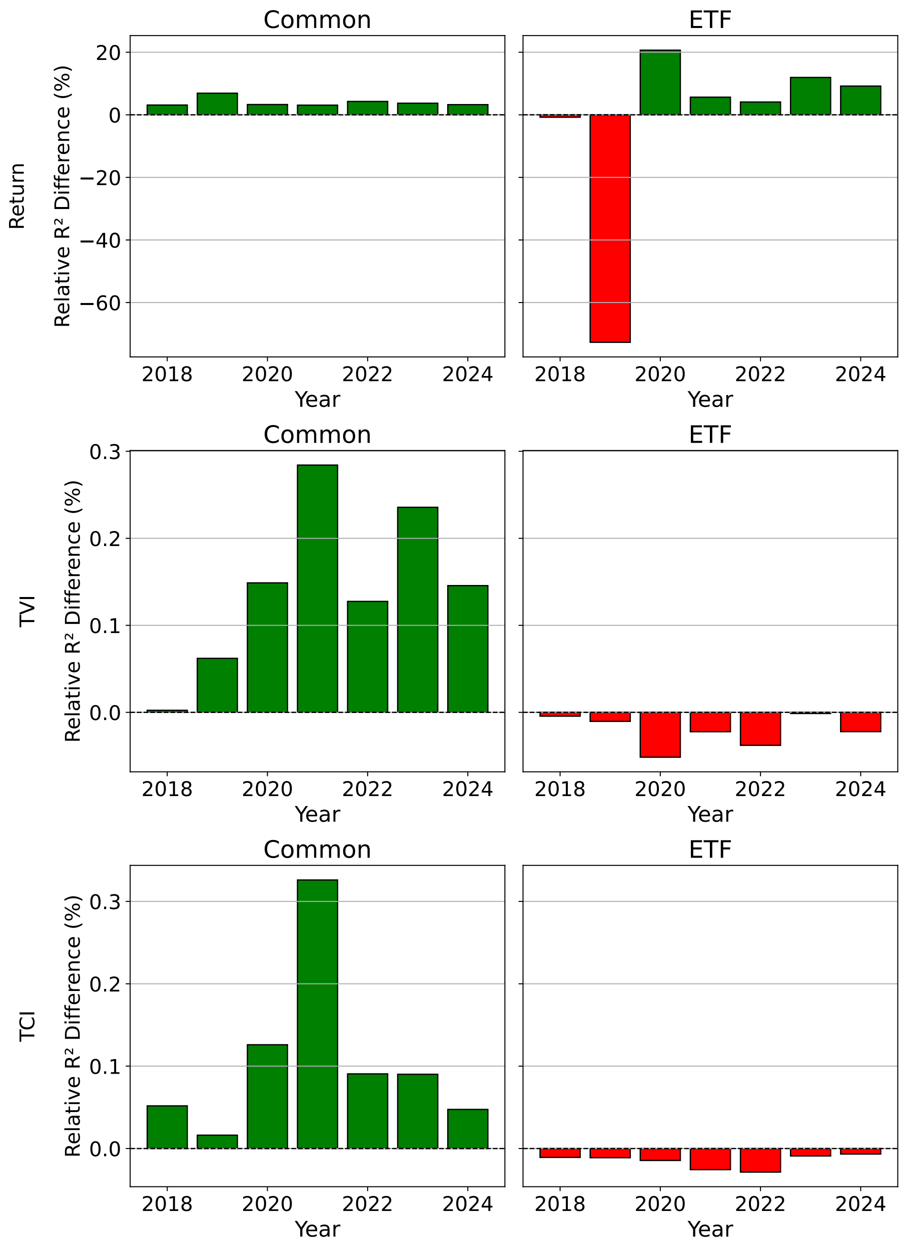


Figure 5.6: Relative out-of-sample R^2 gain from including VWAP in the Random Forest model, disaggregated by security type (common stock vs. ETF) and target variable (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

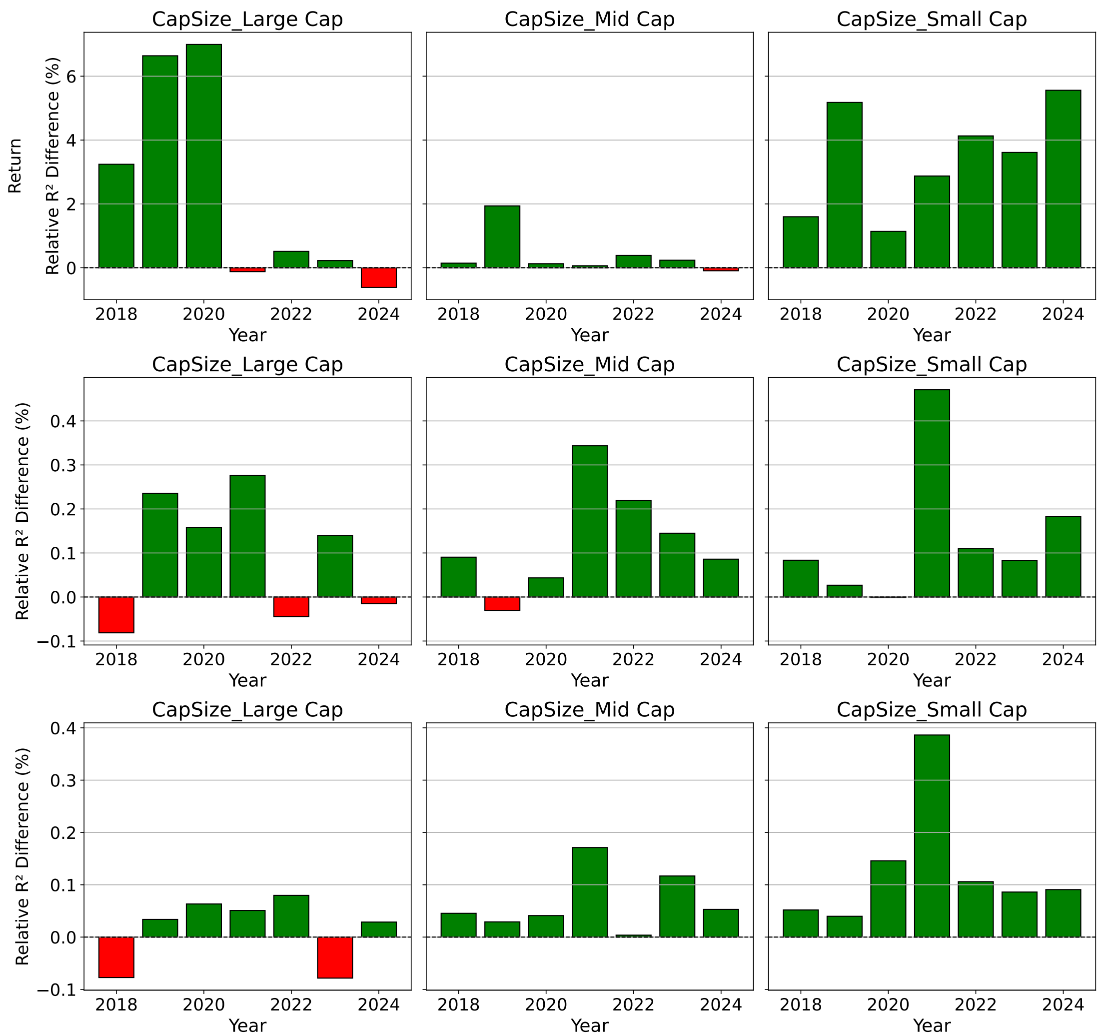


Figure 5.7: Relative out-of-sample R^2 gain from including VWAP in the Random Forest model, disaggregated by market capitalization (small, mid and large cap) and target variable (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

Figure 5.8: Relative out-of-sample R^2 gain from including VWAP in the Random Forest model for forecasting returns, disaggregated by industry group, for test years 2018–2024.

Figure 5.9: Relative out-of-sample R^2 gain from including VWAP in the Random Forest model for forecasting Trade Volume Imbalance (TVI), disaggregated by industry group, for test years 2018–2024.

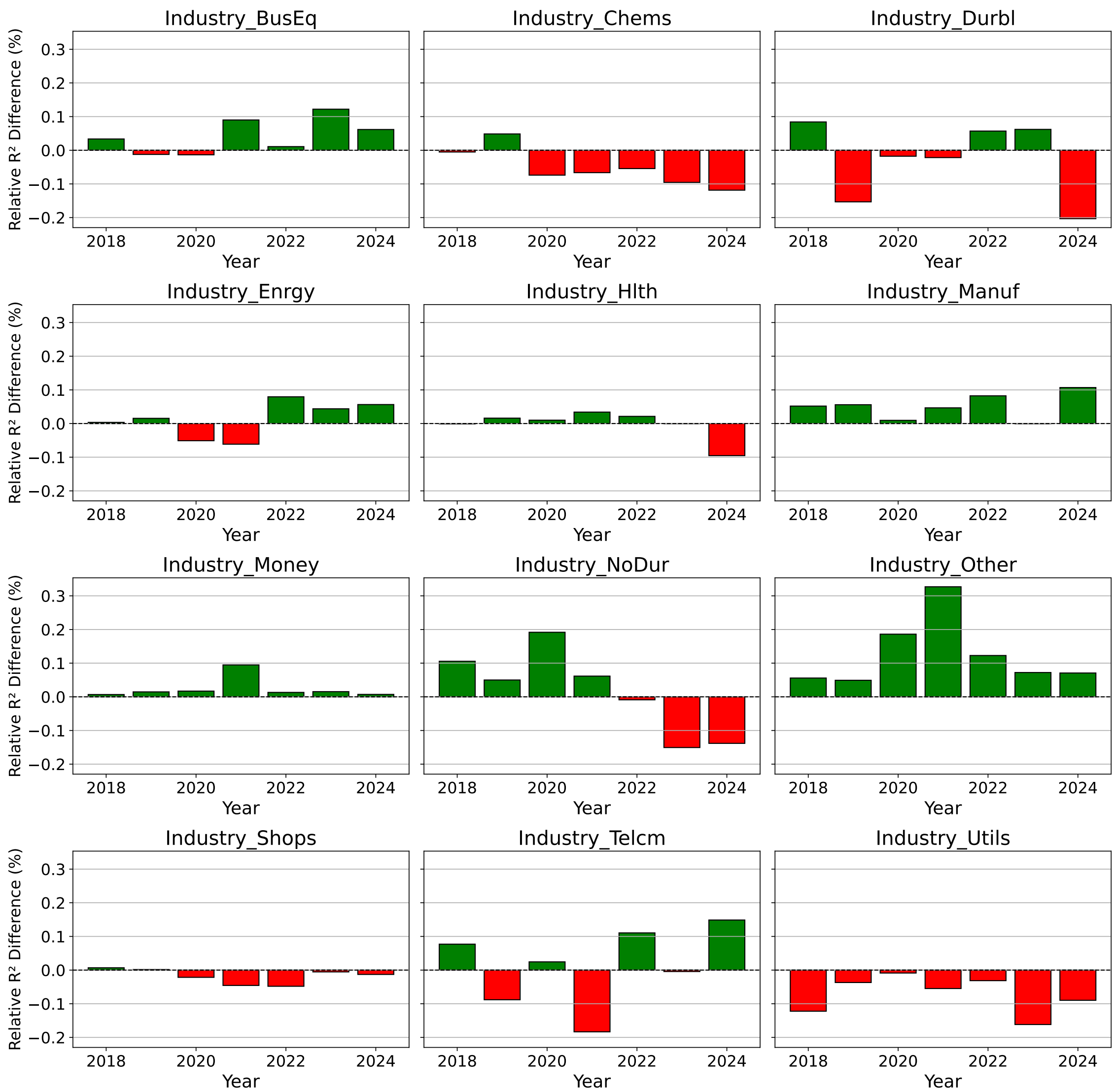


Figure 5.10: Relative out-of-sample R^2 gain from including VWAP in the Random Forest model for forecasting Trade Count Imbalance (TCI), disaggregated by industry group, for test years 2018–2024.

sample (2019–2020), though later years are mixed. The stronger VWAP contribution in smaller firms is intuitive: their prices are more sensitive to intraday order imbalances, and deviations from VWAP may better capture latent demand pressures in less liquid names.

At the industry level (Figure 5.8), Relative VWAP enhances return forecasts in nearly all sectors which is economically significant, but the magnitude varies. Financials Money, Energy, and selected cyclical industries (e.g., Durables, Manufacturing) show especially large improvements, with some gains exceeding 10% in particular years. In contrast, defensive sectors such as Utilities and Telecoms exhibit much weaker effects. This pattern indicates that VWAP-related pressure is most informative in industries prone to larger intraday swings and liquidity fluctuations.

For trade volume imbalance (Figure 5.9), VWAP provides limited incremental value. Most industry-year cells are close to zero or even negative, with only scattered positives (e.g., Utilities and Energy). The results imply that VWAP does not systematically enhance forecasts of volume-based imbalances once Random Forest already accounts for direct order-flow measures.

The case of trade count imbalance (Figure 5.10) is similarly mixed. While certain industries (e.g., Other, Telecom, Shops) display occasional positive gains, many others record flat or negative contributions, particularly in later years. Unlike for returns, VWAP appears to add little incremental predictive power to trade count dynamics, and in some cases it may even introduce noise.

In summary, the Random Forest evidence suggests that VWAP's incremental value is more visible in return forecasts, most notably for small-cap names and cyclical/financial sectors, whereas additions for flow targets (TVI/TCI) are modest and sometimes negative. Relative to Lasso, where effects were muted or unstable, the pattern appears stronger when the model can accommodate interactions and state dependence, which may reflect thresholds with turnover/imbalance, regime variation (e.g., 2020), or collinearity that down-weights VWAP in the penalized linear setting.

Taken together, these results indicate that the midday VWAP deviation may carry information about latent demand or supply that is more consequential for returns than for the structure of trade imbalances. This pattern may reflect that the VWAP deviation's association with returns and flows is not purely linear and likely operates through interactions and state dependence; models that capture interactions, such as Random Forest, seem useful for this purpose, while conclusions should be stress-tested across regimes, feature sets, and transformations before being embedded in adaptive VWAP tactics.

6 Equity Clustering

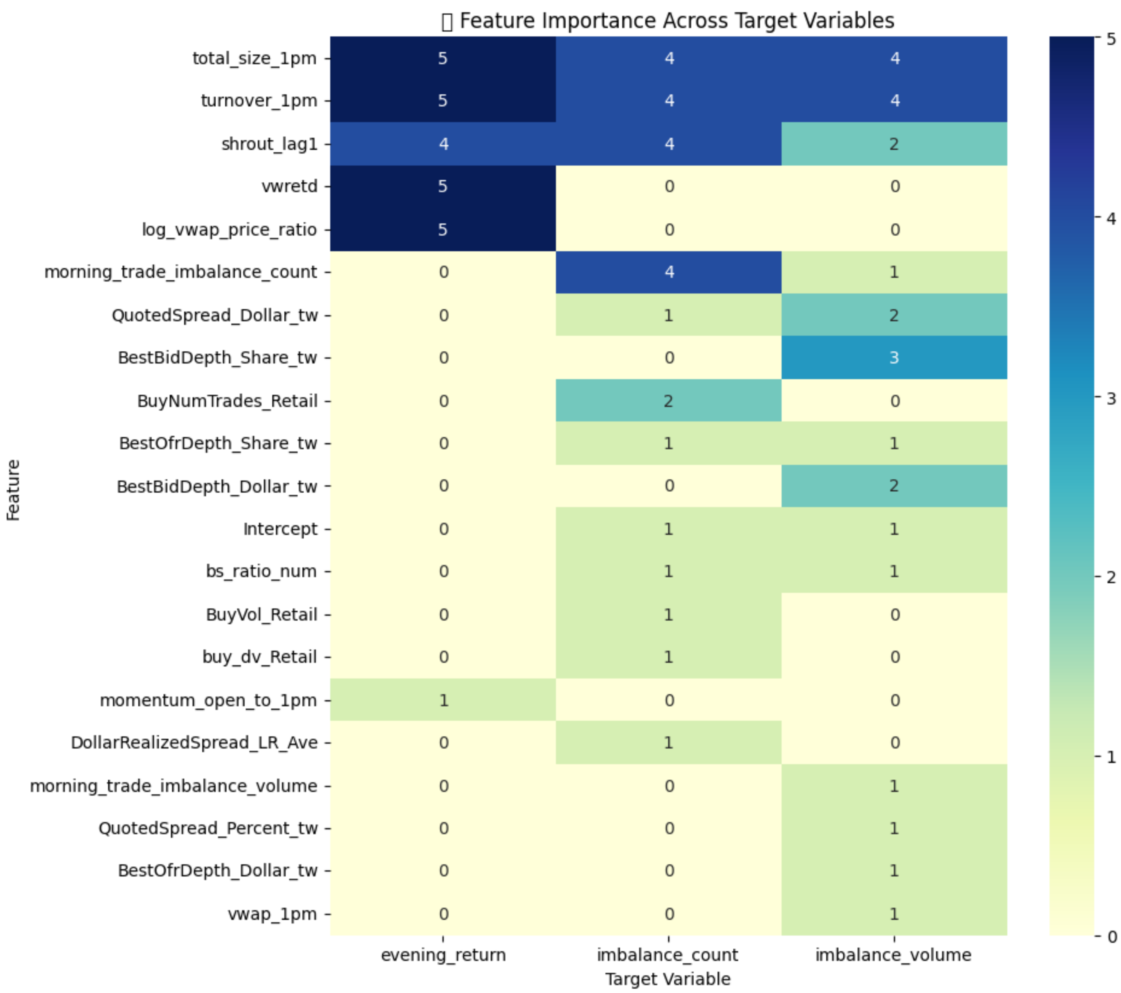
To uncover structural similarities among equities and identify the features that differentiate their predictive behavior, stocks are clustered on per-ticker LASSO coefficient profiles. The procedure is run separately for each predictive target: (i) the return, (ii) the trade count imbalance (TCI), and (iii) the trade volume imbalance (TVI). For each target and ticker, a LASSO model is estimated with a standardized pipeline (mean imputation, feature scaling, and cross-validated penalization over a logarithmic α grid). The resulting coefficients (including the intercept) form a matrix with tickers in rows and features in columns; equities with insufficient observations or unstable cross-validation are excluded.

For each target, I compute pairwise Euclidean distances on the coefficient matrix and apply hierarchical agglomerative clustering with average linkage. The dendrogram is truncated at $K = 5$ clusters to obtain interpretable groups. I report cluster assignments, cluster centroids (mean coefficients by cluster), and the top-five features per cluster based on the average absolute weight. As a quality diagnostic, I compute the silhouette score using Euclidean distances; values are positive across targets, indicating non-trivial separation in coefficient space. Dendrograms and 2-D PCA projections colored by cluster are provided in the Online Appendix.

6.1 Cross-Target Signals from the Clustering Runs

Table 6.1 summarizes the features most frequently selected by LASSO across all equities for each target (higher counts indicate broader relevance). Three patterns stand out. First, `total_size_1pm` and `turnover_1pm` are consistently influential for all targets, highlighting the central role of midday trading intensity in both price and order-flow dynamics. Second, return prediction loads more on price-level indicators (e.g., morning return `wret` and `log_vwap_price_ratio`), whereas imbalance predictions rely more on microstructure/liquidity variables (e.g., `morning_trade_imbalance_count`, depth and spread measures). Third, many signals are target-specific: variables that matter for return are often weak for TCI/TVI and vice-versa, implying that predictive archetypes differ by objective and motivating separate clusterings rather than a single pooled one.

Table 6.1: Heatmap of the most important features in clustering, shown separately for return, trade count imbalance (TCI), and trade volume imbalance (TVI) targets. Darker cells indicate features selected more consistently across tickers, highlighting both common predictors (e.g., total_size_1pm, turnover_1pm) and target-specific signals. Find the variable names and descriptions in table 2



6.2 Economic Interpretation of Clusters

Across targets, centroid profiles reveal distinct predictive archetypes. One group is dominated by liquidity and depth variables (microstructure-driven behavior), another by price level and VWAP-related signals (momentum/price-anchoring dynamics), and others by mixed intensity measures (turnover and total size). These differences are economically meaningful and actionable: combining clusters in portfolio construction diversifies exposure to complementary sources of predictability (price-level vs. order-flow signals) and reduces dependence on any single mechanism.

All inputs (ticker-level betas, cluster assignments, centroids, silhouette diagnostics, dendrograms, PCA maps, and top-feature tables) are exported by the clustering scripts and archived with the thesis materials to ensure full replicability.

From the perspective of the midday execution problem, these clusters reveal which subsets of equities exhibit similar responses to VWAP deviations and order-flow pressures. Stocks grouped by comparable midday turnover, liquidity depth, and share outstanding characteristics reflect distinct execution environments that a trader confronts when deciding whether to accelerate or delay completion of an order. In this sense, the clustering results operationalize the thesis hypothesis by translating statistical heterogeneity into execution heterogeneity: equities with persistent high-liquidity or large-cap signatures tend to display weaker predictive links between VWAP gaps and afternoon drift, whereas thinly traded or high-turnover clusters show stronger sensitivity. Hence, the clustering analysis directly connects model-based predictability to the practical question of how midday information can guide adaptive VWAP execution, by showing that equities with similar microstructure signatures share consistent midday-to-close response patterns.

7 Backtesting

This chapter evaluates the practical value of the intraday forecasts by simulating daily, rules-based long–short portfolios over the 2024 calendar year. The design mirrors the production script used in our experiments and adheres strictly to out-of-sample evaluation with a rolling three-year estimation window. The objective is to assess whether cross-sectional rankings implied by models trained on midday information translate into positive payoffs after 1:00 p.m. Specifically, we form daily long and short legs by sorting stocks on each model's predicted signal, measure realized performance using the subsequent return from 1:00 p.m. to the close, and compare the symmetry of the long and short legs as well as the performance of the market-neutral spread. This procedure is repeated across all analysis subsets, security type, capitalization buckets, and industries, and for each training target: return, TCI, and TVI.

7.1 Methodology and Implementation

The master dataset contains equity–day observations with timestamped features/targets and identifiers. For backtesting, rows with valid dates, the relevant training target, and the realized afternoon return are retained. Subsets are constructed using the same logic as in the empirical analyses: common stocks ($\texttt{shrcd} \in \{10, 11\}$), ETFs ($\texttt{shrcd} = 73$), capitalization buckets via the `cap_size` attribute, and industry buckets via `major_industry`. Observations from unknown or empty categories are excluded. For each (*subset*, *target*) pair, we estimate a Lasso regression within a fixed preprocessing pipeline consisting of mean imputation, standardization, and a Lasso model with a fixed penalty α taken from previously optimized summary files, thereby ensuring no hyperparameter tuning on test data. The daily re-estimation uses a rolling window of the preceding three years, ending strictly before the test day, with a minimum of 250 training days. Where necessary, up to four multiplicative bumps of α are applied to ensure convergence while keeping the parameter scale close to the original.

On each test day, eligible stocks in the cross section are scored to obtain a *Predicted* value, and two legs are formed using symmetric quantile thresholds. Using the default $q = 0.10$, the long leg contains stocks in the top decile of scores, while the short leg contains those in the bottom decile. Equal-weight mean realized afternoon returns are computed for each leg, and the daily market-neutral spread is defined as the difference between the long and short returns. When the training target is TCI or TVI, the evaluation still uses realized returns, thus directly testing whether predicted imbalances contain incremental information about subsequent price movements. Performance is evaluated using the annualized Sharpe ratio, excess Sharpe ratio relative to a 3.98% annual risk-free rate, and annualized mean return, with calculations based on daily means and standard deviations scaled to a 252-day year:

$$\text{Sharpe} = \frac{\mu}{\sigma}\sqrt{252}, \qquad \text{Excess Sharpe} = \frac{\mu - r_f/252}{\sigma_{(x-r_f/252)}}\sqrt{252}, \qquad \text{Annualized Return} = 252\,\mu.$$

The implementation adheres to principles designed for rigor and reproducibility: the master file is loaded once; dates and numeric fields are validated; infinities and missing values are removed; fixed α values are taken from prior optimization to prevent information leakage; models are refit each test day in strict time order; and the predictor set is rebuilt programmatically by selecting numeric columns and excluding identifiers and the current target. Quantile legs are reconstructed every day from the test cross section, with missing predictions or returns dropped for that date. The main parameters follow the production defaults: a 3.98% annual risk-free rate, decile-based legs, at least 250 training days, a Lasso iteration cap of 50,000 with tolerance 10^{-4}, and up to four α adjustments upon convergence warnings.

This backtest operationalizes the midday execution choice: if the VWAP gap at 1:00 p.m. and morning TCI/TVI encode latent afternoon pressure, then forecasts built only on noon-available inputs should improve the accelerate vs. pace decision for finishing the schedule. Forming positions at 1:00 p.m. is thus a proxy for moving an execution trajectory forward when predicted return/imbalance is favorable, and holding back when it is not—directly testing whether the noon information set is actionable for VWAP-oriented execution.

7.2 Results and Performance Analysis

Across all (*subset*, *target*) pairs with optimized α, the long–short decile strategy delivers strikingly strong 1:00 p.m.–to–close performance in 2024. Median spread Sharpe ratios (annualized) by training target are extremely elevated for TCI and TVI and remain very high even for the return-trained model, with corresponding median annualized spread returns in the multi-hundred-percent range. In levels, average daily spread means cluster around one percentage point while median spread standard deviations are substantially lower, producing very high risk-adjusted scores. The pattern is broad-based: order-flow targets generally dominate the return target at the subset level, consistent with midday imbalance signals carrying richer information about near-term price pressure than contemporaneous returns. Two segments illustrate the range: diversified common stocks and several industry buckets record extreme spread Sharpes and annualized spreads, whereas ETFs show the weakest outcomes, still positive but far more modest.It is notable that the processing time for this procedure was approximately eight days.

Across the subsets, annualised Sharpe ratios ranged from roughly 6 to 90, and annualised mean returns from about 0.24 to 3.4. While these figures indicate very strong in-sample risk-adjusted performance, they are unusually high for an intraday market-neutral setting; potential reasons for these high values are discussed in section 7.3.

The long and short legs are both highly active and contribute meaningfully to the spread, but the short side is, on average, stronger in magnitude. Median daily long-leg means are positive while the median absolute short-leg means are slightly larger, and the absolute short-leg Sharpe typically exceeds the long-leg Sharpe. This mild asymmetry suggests the signals are not just identifying “winners” but also systematically locating intraday overvaluation or pressure that corrects into the close. At the cross-sectional level, industry stratification appears beneficial: most industries show TCI/TVI beating the return-targeted model, with only a few exceptions where the return target edges out. The dispersion across subsets is economically plausible: broad, diversified universes can generate lower portfolio volatility through cross-sectional averaging, mechanically inflating Sharpe for a given spread mean, while narrower or more liquid universes such as ETFs deliver lower but still positive performance.

Viewed through the execution lens, the positive and asymmetric portfolio payoffs indicate that noon VWAP deviations and order-flow signals anticipate the direction of slippage: a positive (negative) return or buy-pressure forecast supports front-loading (deferring/being passive) to reduce expected VWAP shortfall. In other words, the statistical lift converts into a concrete execution rule, map forecast sign and magnitude into aggressiveness, closing the loop from prediction to schedule control.

7.3 Practical Considerations and Limitations

The backtest results should be interpreted with caution, as several simplifying assumptions likely exaggerate the reported Sharpe ratios and annualised returns. The return magnitudes are exceptionally high for intraday strategies and should be viewed as upper-bound estimates rather than attainable real-world outcomes. The following limitations outline the main reasons for potential inflation in these figures.

The backtest does not include explicit transaction costs, even though each portfolio leg trades twice per day—once at 1:00 p.m. to open and once at the close to unwind. Even modest per-trade costs such as commissions, half-spreads, market-impact slippage, or exchange fees can accumulate rapidly in a high-turnover intraday strategy. To a first approximation, if c denotes the *per-side* cost for a single buy or sell (inclusive of execution slippage), then the net daily spread is

$$r_d^{\Delta,\text{net}} \approx r_d^{\Delta} - 2c_{\text{long}} - 2c_{\text{short}} \approx r_d^{\Delta} - 4c,$$

With raw mean spreads near one percent per day, even a stylised 5 bps per-side cost could remove roughly 20 bps of daily profit. In smaller-capitalisation equities, costs would likely be higher due to limited depth and greater market impact.

The backtest assumes that trades occur exactly at 1:00 p.m. and 4:00 p.m. using contemporaneous spot prices. In reality, executions depend on queue position, available liquidity, and market volatility at those times. Achieving perfect fills at those precise prices is rarely possible. For example, a trader attempting to buy at 1:00 p.m. may experience price drift or partial fills before the order completes, especially in volatile names.

Slippage—the difference between the intended and actual execution price—can be substantial for thinly traded securities. The strategy implicitly assumes full immediacy and neglects the adverse price movement that often accompanies large trades in less-liquid names. For instance, submitting a market order to sell a mid-cap stock with a wide bid–ask spread could push the price down several ticks before completion, reducing realised profits relative to the modelled outcome.

The simulation presumes that all stocks in the universe are continuously tradeable. In practice, some securities can become temporarily untradeable due to halts, auction freezes, or insufficient liquidity. A more realistic backtest would apply a feasibility filter removing names that fall below minimum thresholds of dollar volume or quoted depth within the 12:55–1:05 p.m. window.

Prices are treated as if transactions occur at the prevailing spot rate, ignoring the bid–ask spread. This assumption effectively grants free liquidity to the trader. For example, if a stock trades at $100.00 bid and $100.10 ask, the model assumes the trader can buy at $100.00 and sell at $100.10 without cost—an unrealistic simplification that biases returns upward.

The strategy treats short selling as frictionless, omitting borrow fees, locate costs, and regulatory constraints such as uptick rules or inventory scarcity. In reality, some securities—especially small-cap or hard-to-borrow names—may be only partially shortable or entirely unavailable. Borrow rates can vary widely and meaningfully reduce the profitability of short positions. Notably, the larger portion of returns in this study arises from the short leg, which implies that unmodelled borrowing and locate costs could materially compress realised performance.

Equal-weighting across top and bottom deciles is operationally challenging at intraday horizons. Precise dollar neutrality requires fractional shares or frequent re-scaling, and small deviations can create unintended exposure drift during the afternoon. Moreover, daily open–close turnover is effectively 200% per position, and the composition of the deciles can change from day to day, implying high implementation turnover and capacity limits. Executing large notional volumes across many mid- and small-cap equities near 1:00 p.m. could impose non-trivial market impact.

Although the rolling-window design enforces strict time ordering and uses only pre-1:00 p.m. information for predictions, subtle optimism biases may remain. These include survivorship bias in the stock universe, inadvertent leakage through engineered variables, or predictors indirectly reflecting post 1:00 p.m. information (for example, VWAPs updated intraday). Such effects, while small, could elevate reported forecast accuracy and performance.

Taken together, these limitations underscore that the headline Sharpe ratios and annualised returns likely represent optimistic upper bounds. The reported magnitudes should therefore be interpreted as indicative of signal strength rather than achievable profit. Incorporating realistic transaction-cost debits, liquidity filters, shorting constraints, and bid–ask spreads would yield more moderate yet still informative estimates. Two immediate robustness extensions are recommended: (i) a transaction-cost sensitivity study applying per-trade costs of 3–10 bps and recomputing net Sharpe ratios, and (ii) a feasibility filter excluding hard-to-borrow or illiquid names. If the relative ranking of predictability—TCI $\geq$ TVI $\gg$ return—remains intact under these adjustments, the qualitative conclusions would continue to hold while the magnitudes better reflect implementable strategies.

Cumulative spread return plots provide visual diagnostics of performance dynamics. Placeholders below can be used to include final exhibits once produced:

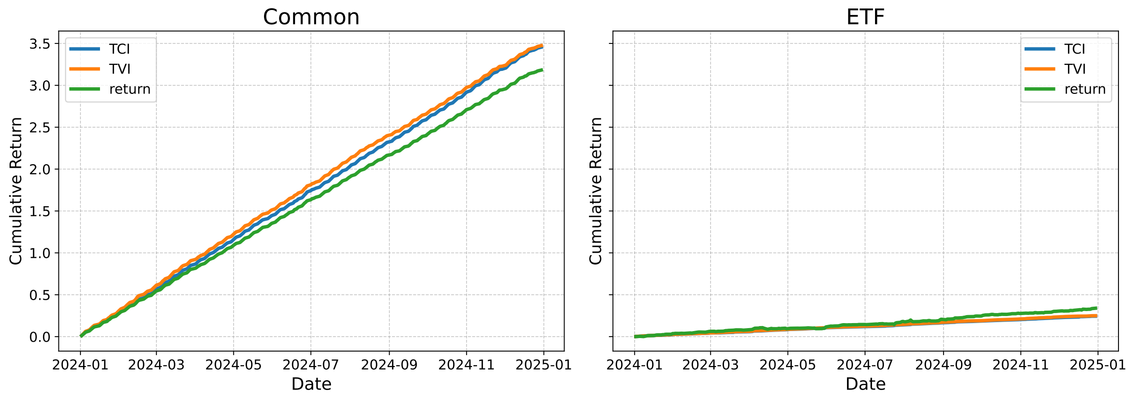


Figure 7.1: Cumulative returns for equally weighted long–short portfolios (top 10%, bottom 10%) over 2024, rebalanced daily. The left panel shows results for Common equities and the right panel for ETFs, with performance reported for three targets: evening return, evening trade count imbalance (TCI), and evening trade volume imbalance (TVI).

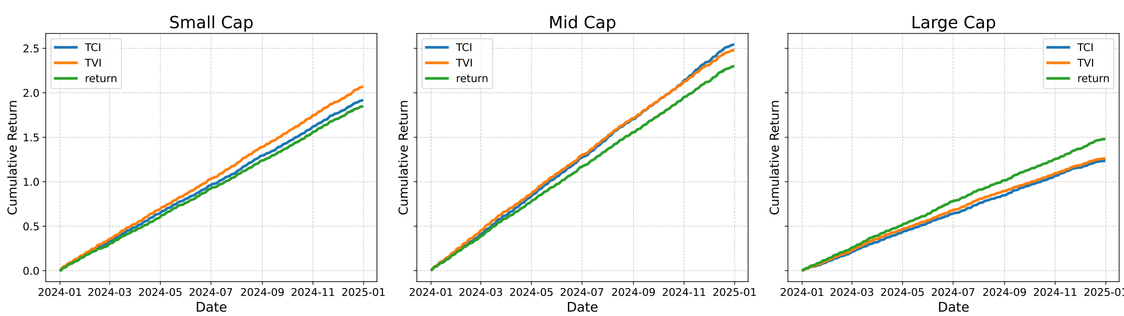


Figure 7.2: Cumulative returns for equally weighted long–short portfolios (top 10%, bottom 10%) over 2024, rebalanced daily. Performance is reported for three market capitalization tiers—small-cap, mid-cap, and large-cap—using the same three predictive targets: evening return, evening trade count imbalance (TCI), and evening trade volume imbalance (TVI). The plots cover January 2024 to January 2025, illustrating differences in predictive target performance across capitalization segments

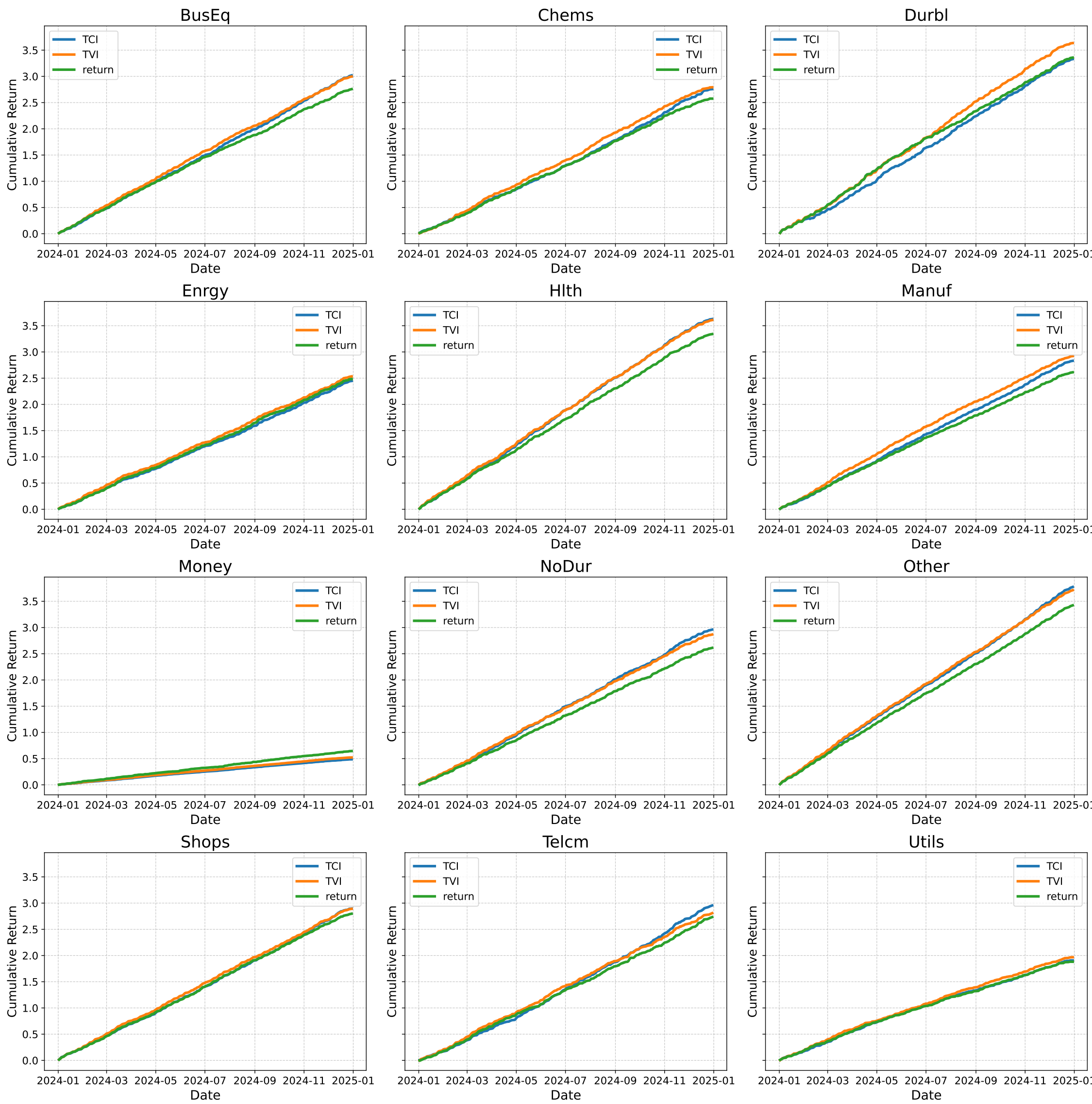


Figure 7.3: Cumulative returns for equally weighted long–short portfolios (top 10%, bottom 10%) over 2024, rebalanced daily. Performance is shown separately for each industry group—BusEq, Chems, Durbl, Enrgy, Hlth, Manuf, Money, NoDur, Other, Shops, Telcm, and Utils—using three predictive targets: evening return, evening trade count imbalance (TCI), and evening trade volume imbalance (TVI). This figure highlights variation in predictive target effectiveness and stability across industries from January 2024 to January 2025

8 Conclusion

This thesis investigated whether midday market indicators contain useful information for forecasting afternoon market dynamics, focusing on three specific targets: the return from 1:00 p.m. to the market close, trade count imbalance (TCI), and trade volume imbalance (TVI). The study also examined how such forecasts can guide the 1:00 p.m. execution decision, when traders must determine how aggressively to complete remaining orders under uncertainty. To address this, a rolling out-of-sample forecasting framework was developed using both Lasso and Random Forest models, incorporating stratifications by equity type, market capitalization, and industry. The analysis further assessed the incremental role of VWAP through exclusion tests and identified structural similarities among equities using clustering analysis.

The results show that order-flow variables (TCI and TVI) are considerably more predictable than returns across most years and subsamples. This difference likely reflects the underlying persistence of intraday liquidity demand and supply imbalances that evolve more smoothly than prices themselves. In contrast, price movements incorporate the cumulative effect of diverse trading motives and external shocks, making them harder to forecast over short horizons. Therefore, the stronger predictability of TCI and TVI implies that the statistical structure of order flow carries economically relevant information about subsequent market behavior, even when prices appear noisy. TCI showing a marginal yet consistent advantage over TVI. This small but robust edge suggests that trade-count asymmetry—reflecting the direction and persistence of trading activity—may provide a cleaner signal of latent liquidity pressure than size-weighted volume measures, which can be distorted by block prints or irregular trade sizes. Industry-level stratification also improved forecast accuracy, suggesting that firms within the same sector share common liquidity regimes and trading routines that shape their short-term dynamics. This finding supports the use of industry-aware modeling to reduce noise and exploit recurring intraday patterns.

The analysis of VWAP's predictive role revealed distinct behaviors between modeling frameworks. In the Lasso setting, excluding VWAP produced negligible changes in R^2, indicating weak and inconsistent effects. However, in the Random Forest framework, VWAP added measurable predictive power, particularly for returns. This divergence suggests that VWAP contributes to predictability primarily through nonlinear interactions—for instance, the informativeness of a VWAP deviation may depend on the security's liquidity, turnover, or concurrent imbalance conditions. While linear models cannot capture such context-dependent effects, ensemble-based methods are more flexible and can learn threshold-driven relationships. This helps explain why VWAP's incremental effect appears only in nonlinear models and varies across years, reflecting changing market regimes and the evolving nature of algorithmic execution behavior.

The clustering results provided further evidence of structural heterogeneity across equities. Total size by 1:00 p.m., turnover by 1:00 p.m., and shares outstanding emerged as the most important features distinguishing equity clusters. These variables capture stable differences in midday trading intensity and market depth, which, in turn, affect the predictability of both order flow and prices. Larger, more liquid securities tend to exhibit smoother intraday liquidity replenishment and lower price sensitivity to temporary imbalances, whereas smaller or thinner names react more sharply to localized order shocks. This clustering insight highlights the potential value of conditioning forecasting models—and by extension, execution strategies—on stock-specific microstructure characteristics rather than treating all securities as homogeneous.

The backtesting analysis linked statistical predictability to economic relevance. Long—short portfolios constructed from model forecasts generated substantial and persistent excess returns during the 2024 test period, with Sharpe ratios ranging from approximately 6 to 90 and annualized returns between 0.24 and 3.4 across subsets. These values demonstrate that, under idealized conditions, forecast signals could translate into economically meaningful outcomes. However, these figures should be interpreted with caution. Several simplifying assumptions likely inflate performance: transaction costs were not debited despite two trades per day; executions were assumed to occur precisely at 1:00 p.m. and 4:00 p.m. using spot prices; slippage was ignored, particularly for less liquid names; and the analysis did not consider trade feasibility, bid–ask spreads, or short-selling costs, even though the short leg contributed most of the excess return. In practice, these factors would substantially reduce realized profitability, though the relative ranking of signals—TCI outperforming TVI, both well above returns—is expected to remain valid. After accounting for frictions, the observed Sharpe ratios and returns likely represent upper bounds rather than achievable outcomes.

These empirical findings collectively inform how forecasts can be translated into actionable execution guidance. Noon-to-close forecasts of return or imbalance provide directional cues for VWAP pacing: positive forecasts imply latent buying pressure and may justify accelerating when behind schedule, whereas negative forecasts indicate selling pressure and suggest easing participation when ahead. Because Random Forest models capture nonlinear dependencies between VWAP deviations, liquidity, and imbalance, they can identify conditions in which these signals are more reliable. This operational interpretation reconnects the statistical results to the trader's decision problem: the models serve not only as forecasting tools but also as decision aids that help calibrate execution speed to evolving intraday conditions.

Despite encouraging results, several limitations remain. The feature set was deliberately restricted to liquidity, order-flow, and VWAP-related measures observable by midday, excluding other potentially informative variables such as limit-order book depth, news sentiment, or high-frequency volatility indicators. While Lasso and Random Forest capture complementary forms of structure, they do not span the full space of modern forecasting tools; more advanced nonlinear or hybrid methods—such as gradient boosting, extremely randomized trees, or recurrent neural networks—could capture additional temporal and cross-sectional dependencies. Although the rolling-window design mitigates look-ahead bias, it cannot entirely remove optimism from survivorship effects or regime-specific patterns. Furthermore, capacity and scalability constraints were not explicitly modeled: executing meaningful notional volumes near 1:00 p.m. in mid- or small-cap equities could produce significant market impact. Finally, the backtesting framework, while informative, remains an initial proof of concept and should be extended to account for transaction costs, execution constraints, and dynamic position sizing.

Future research could extend this work in several directions. Integrating richer intraday data—such as limit-order book depth, real-time imbalance indicators, and volatility forecasts—would enhance model granularity. Expanding the methodological scope to include regime-switching or online learning approaches could improve adaptability under changing market conditions. Incorporating transaction-cost modeling, liquidity filters, and borrowing costs would yield more realistic assessments of net profitability. Finally, embedding the forecasting layer directly into a live or simulated VWAP strategy would enable direct evaluation of execution quality improvements attributable to predictive signals.

Taken together, this research contributes to both the academic literature on intraday market predictability and the practical design of algorithmic execution strategies. It suggests that order-flow imbalances are more reliably forecastable than short-term returns, that industry-aware and cluster-based modeling can improve stability and precision, and that VWAP's incremental information is primarily nonlinear and conditional on market context. Most importantly, it bridges the gap between statistical forecasting and trader decision-making by showing how midday signals can inform execution pacing in a realistic framework. While the reported magnitudes should be interpreted conservatively, the study demonstrates a consistent connection between measurable intraday predictability and the economic logic of adaptive VWAP execution.

Bibliography

Admati, Anat R. and Paul Pfleiderer (Apr. 2015). "A Theory of Intraday Patterns: Volume and Price Variability". In: *The Review of Financial Studies* 1.1, pp. 3–40. ISSN: 0893-9454. DOI: 10.1093/rfs/1.1.3. eprint: https://academic.oup.com/rfs/article-pdf/1/1/3/24434731/010003.pdf. URL: https://doi.org/10.1093/rfs/1.1.3.

Almgren, Robert F. and Neil Chriss (2000). "Optimal execution of portfolio transactions". In: *Journal of Risk* 3.2. First published Winter 2000/2001, pp. 5–39. DOI: 10.21314/JOR.2001.041. URL: https://www.risk.net/journal-risk/2161150/optimal-execution-portfolio-transactions.

Andersen, Torben G. and Tim Bollerslev (1998). "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts". In: *International Economic Review* 39.4, pp. 885–905. ISSN: 00206598, 14682354. URL: http://www.jstor.org/stable/2527343 (visited on 11/03/2025).

Avellaneda, Marco and Jeong-Hyun Lee (2010). "Statistical Arbitrage in the US Equities Market". In: *Quantitative Finance* 10.7, pp. 761–782. DOI: 10.1080/14697680903124632.

Biais, Bruno, Pierre Hillion, and Chester Spatt (1995). "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse". In: *The Journal of Finance* 50.5, pp. 1655–1689. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2329330 (visited on 11/03/2025).

Bini, B.S. and Tessy Mathew (2016). "Clustering and Regression Techniques for Stock Prediction". In: *Procedia Technology* 24, pp. 1248–1255. ISSN: 22120173. DOI: 10.1016/j.protcy.2016.05.104. (Visited on 08/11/2025).

Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327. DOI: 10.1016/0304-4076(86)90063-1. URL: https://www.sciencedirect.com/science/article/pii/0304407686900631.

Campbell, John Y. and Samuel B. Thompson (Nov. 2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" In: *The Review of Financial Studies* 21.4, pp. 1509–1531. ISSN: 0893-9454. DOI: 10.1093/rfs/hhm055. eprint: https://academic.oup.com/rfs/article-pdf/21/4/1509/24453399/hhm055.pdf. URL: https://doi.org/10.1093/rfs/hhm055.

Chauhan, Lakshay, John Alberg, and Zachary C. Lipton (July 2020). *Uncertainty-Aware Lookahead Factor Models for Quantitative Investing*. DOI: 10.48550/arXiv.2007.04082. arXiv: 2007.04082 [q-fin]. (Visited on 08/11/2025).

Chen, Xiao et al. (Dec. 2024). "A Multi-Feature Stock Price Prediction Model Based on Multi-Feature Calculation, LASSO Feature Selection, and Ca-LSTM Network". In: *Connection Science* 36.1, p. 2286188. ISSN: 0954-0091, 1360-0494. DOI: 10.1080/09540091.2023.2286188. (Visited on 08/11/2025).

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2002). "Order imbalance, liquidity, and market returns". In: *Journal of Financial Economics* 65.1, pp. 111–130. URL: https://EconPapers.repec.org/RePEc:eee:jfinec:v:65:y:2002:i:1:p:111-130.

Cont, R., A. Kukanov, and S. Stoikov (June 2013). "The Price Impact of Order Book Events". In: *Journal of Financial Econometrics* 12.1, pp. 47–88. ISSN: 1479-8417. DOI: 10.1093/jjfinec/nbt003. URL: http://dx.doi.org/10.1093/jjfinec/nbt003.

Corberán-Vallet, Ana et al. (Apr. 2023). "A New Approach to Portfolio Selection Based on Forecasting". In: *Expert Systems with Applications* 215, p. 119370. ISSN: 09574174. DOI: 10.1016/j.eswa.2022.119370. (Visited on 08/11/2025).

Ding, H. (2024). "Evaluating the Effectiveness of Elastic Net Model in Predicting Stock Closing Price". In: *TNS* 51.1, pp. 172–178. DOI: 10.54254/2753-8818/51/2024CH0198. URL: https://doi.org/10.54254/2753-8818/51/2024CH0198.

Engle, Robert F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007. URL: http://www.jstor.org/stable/1912773.

Engle, Robert F. and Jeffrey R. Russell (1998). “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data”. In: *Econometrica* 66.5, pp. 1127–1162. URL: http://www.jstor.org/stable/2999632.

Fallahi, Alireza (Oct. 2023). “Forecasting the Intraday Trading Volume of Stocks”. Master’s Thesis. Montréal, Canada: Concordia University. URL: https://spectrum.library.concordia.ca/id/eprint/XXXX/.

Fischer, Thomas and Christopher Krauss (2018). “Deep learning with long short-term memory networks for financial market predictions”. In: *European Journal of Operational Research* 270.2, pp. 654–669. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2017.11.054. URL: https://www.sciencedirect.com/science/article/pii/S0377221717310652.

Gao, Lei et al. (Aug. 2018). “Market Intraday Momentum”. In: *Journal of Financial Economics* 129.2, pp. 394–414. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2018.05.009. (Visited on 08/02/2025).

Gao, X. et al. (2021). “The role of hybrid models in financial decision-making: Forecasting stock prices with advanced algorithms”. In: *Egyptian Informatics Journal*. Describes ensemble learning blending tree-based models such as RF and gradient boosting with regression-based approaches.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (1993). “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”. In: *The Journal of Finance* 48.5, pp. 1779–1801. DOI: 10.1111/j.1540-6261.1993.tb05128.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1993.tb05128.x.

Hallam, Mark and Jose Olmo (Oct. 2014). “Forecasting Daily Return Densities from Intraday Data: A Multifractal Approach”. In: *International Journal of Forecasting* 30.4, pp. 863–881. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2014.01.007. (Visited on 08/02/2025).

Hamilton, James D. (1989). “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”. In: *Econometrica* 57.2, pp. 357–384. URL: http://www.jstor.org/stable/1912559.

Harvey, Campbell R., Yan Liu, and Heqing Zhu (Oct. 2016). “… and the Cross-Section of Expected Returns”. In: *The Review of Financial Studies* 29.1, pp. 5–68. ISSN: 0893-9454. DOI: 10.1093/rfs/hhv059. eprint: https://academic.oup.com/rfs/article-pdf/29/1/5/24450794/hhv059.pdf. URL: https://doi.org/10.1093/rfs/hhv059.

Hasbrouck, Joel (1991). “Measuring the Information Content of Stock Trades”. In: *The Journal of Finance* 46.1, pp. 179–207. DOI: 10.1111/j.1540-6261.1991.tb03749.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1991.tb03749.x.

Heston, Steven L., Robert A. Korajczyk, and Ronnie Sadka (2010). “Intraday Patterns in the Cross-section of Stock Returns”. In: *The Journal of Finance* 65.4, pp. 1369–1407. DOI: 10.1111/j.1540-6261.2010.01573.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01573.x.

Huddleston, Dillon, Fred Liu, and Lars Stentoft (Mar. 2023). “Intraday Market Predictability: A Machine Learning Approach”. In: *Journal of Financial Econometrics* 21.2, pp. 485–527. ISSN: 1479-8409, 1479-8417. DOI: 10.1093/jjfinec/nbab007. (Visited on 08/02/2025).

Illa, Pavan Kumar, Balakesavareddy Parvathala, and Anand Kumar Sharma (2022). “Stock Price Prediction Methodology Using Random Forest Algorithm and Support Vector Machine”. In: *Materials Today: Proceedings* 56, pp. 1776–1782. ISSN: 22147853. DOI: 10.1016/j.matpr.2021.10.460. (Visited on 08/11/2025).

Kim, Kyoung-jae (2003). “Financial Time Series Forecasting Using Support Vector Machines”. In: *Neurocomputing* 55.1. Support Vector Machines, pp. 307–319. ISSN: 0925-2312. DOI: 10.1016/S0925-2312(03)00372-2. URL: https://www.sciencedirect.com/science/article/pii/S0925231203003722.

Kyle, Albert S. (1985). "Continuous Auctions and Insider Trading". In: *Econometrica* 53.6, pp. 1315–1335. ISSN: 00129682, 14680262. URL: `http://www.jstor.org/stable/1913210` (visited on 11/03/2025).

Ma, Lingjie and Larry Pohlman (July 2008). "Return Forecasts and Optimal Portfolio Construction: A Quantile Regression Approach". In: *The European Journal of Finance* 14.5, pp. 409–425. ISSN: 1351-847X, 1466-4364. DOI: `10.1080/13518470802042369`. (Visited on 08/11/2025).

Matías, José M. and Juan C. Reboredo (Mar. 2012). "Forecasting Performance of Nonlinear Models for Intraday Stock Returns". In: *Journal of Forecasting* 31.2, pp. 172–188. ISSN: 0277-6693, 1099-131X. DOI: `10.1002/for.1218`. (Visited on 08/02/2025).

Neba, C. et al. (2023). "Comparative Analysis of Stock Price Prediction Models: Generalized Linear Model (GLM), Ridge Regression, Lasso Regression, Elasticnet Regression, and Random Forest – A Case Study on Netflix". In: *International Journal of Innovative Science and Research Technology* 8.10.

Nelson, Daniel B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". In: *Econometrica* 59.2, pp. 347–370. URL: `http://www.jstor.org/stable/2938260`.

Rapach, David and Guofu Zhou (2013). "Forecasting Stock Returns". In: *Handbook of Economic Forecasting*. Ed. by Graham Elliott and Allan Timmermann. Vol. 2A. Elsevier, pp. 328–383. DOI: `10.1016/B978-0-444-53683-9.00006-2`.

Rastogi, Akshar et al. (Apr. 2021). "Stock Market Prediction with Lasso Regression Using Technical Analysis and Time Lag". In: *2021 6th International Conference for Convergence in Technology (I2CT)*. Maharashtra, India: IEEE, pp. 1–5. ISBN: 978-1-7281-8876-8. DOI: `10.1109/I2CT51068.2021.9417935`. (Visited on 08/11/2025).

Renugadevi, T. et al. (Dec. 2016). "Stock Market Prediction Using Hierarchical Agglomerative and K-Means Clustering Algorithm". In: *Indian Journal of Science and Technology* 9.48. ISSN: 0974-5645, 0974-6846. DOI: `10.17485/ijst/2016/v9i48/108029`. (Visited on 08/11/2025).

Roy, Sanjiban Sekhar et al. (2015). “Stock Market Forecasting Using LASSO Linear Regression Model”. In: *Afro-European Conference for Industrial Advancement*. Ed. by Ajith Abraham, Pavel Krömer, and Vaclav Snasel. Vol. 334. Cham: Springer International Publishing, pp. 371–381. ISBN: 978-3-319-13571-7 978-3-319-13572-4. DOI: 10.1007/978-3-319-13572-4_31. (Visited on 08/11/2025).

Sai, N. et al. (July 2023). “Nifty Stock Prediction Using Elasticnet and LSTM”. In: *IOSR Journal of Computer Engineering (IOSR-JCE)*. Ser. I 25.4, pp. 09–14. ISSN: 2278-0661. DOI: 10.9790/0661-2504010914. URL: https://doi.org/10.9790/0661-2504010914.

Schorno, R. (2022). “Investigating the Use of Ridge, LASSO and Elastic Net for Predicting the Direction of the US Stock Market”. Master’s Thesis. Erasmus School of Economics, Erasmus University Rotterdam.

Sheng, Z. et al. (2025). “A Multi-Feature Stock Index Forecasting Approach Based on LASSO Feature Selection and Non-Stationary Autoformer”. In: *Electronics* 14.10, p. 2059. DOI: 10.3390/electronics14102059. URL: https://doi.org/10.3390/electronics14102059.

Ślusarczyk, Damian and Robert Ślepaczuk (May 2025). “Optimal Markowitz Portfolio Using Returns Forecasted with Time Series and Machine Learning Models”. In: *Journal of Big Data* 12.1, p. 127. ISSN: 2196-1115. DOI: 10.1186/s40537-025-01164-z. (Visited on 08/11/2025).

Ta, Van-Dai, Chuan-Ming Liu, and Direselign Addis Tadesse (Jan. 2020). “Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading”. In: *Applied Sciences* 10.2, p. 437. ISSN: 2076-3417. DOI: 10.3390/app10020437. (Visited on 08/11/2025).

Tsay, Ruey S. (1998). “Testing and Modeling Multivariate Threshold Models”. In: *Journal of the American Statistical Association* 93.443, pp. 1188–1202. URL: http://www.jstor.org/stable/2669861.

Wang, Qingxue et al. (2021). “High-frequency Statistical Arbitrage Strategy Based on Stationarized Order Flow Imbalance”. In: *Procedia Computer Science* 187. 2020 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI2020, pp. 518–523. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2021.04.093. URL: https://www.sciencedirect.com/science/article/pii/S1877050921008917.

Wang, Xingtong, Wen Wang, and Shuya Zhang (2023). "Stock Price Return Prediction Based on Multifactorial Machine Learning Approaches". In: *Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022)*. Ed. by Daowen Qiu, Yusheng Jiao, and William Yeoh. Vol. 5. Dordrecht: Atlantis Press International BV, pp. 324–333. ISBN: 978-94-6463-029-9 978-94-6463-030-5. DOI: 10.2991/978-94-6463-030-5_34. (Visited on 08/11/2025).

Xu, Xu et al. (Dec. 2023). "Forecasting Stock Closing Prices with an Application to Airline Company Data". In: *Data Science and Management* 6.4, pp. 239–246. ISSN: 26667649. DOI: 10.1016/j.dsm.2023.09.005. (Visited on 08/11/2025).

Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu (1998). "Forecasting with Artificial Neural Networks: The State of the Art". In: *International Journal of Forecasting* 14.1, pp. 35–62. ISSN: 0169-2070. DOI: 10.1016/S0169-2070(97)00044-7. URL: https://www.sciencedirect.com/science/article/pii/S0169207097000447.

Zheng, J. et al. (2024). "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance". In: *arXiv preprint arXiv:2402.17194*. URL: https://arxiv.org/abs/2402.17194.

Appendix

Appendix A – Declaration of Generative AI Assistance

I declare that I used OpenAI ChatGPT (Version: GPT-4, [`https://chat.openai.com/`) and Google Gemini (Version 2.5, [`https://gemini.google.com/`) between February and August 2025 to support various stages of my thesis work. These tools were used strictly as assistive resources to enhance my understanding, reflection, and writing process, in accordance with HEC Montréal's responsible AI usage guidelines.

Specifically, I used generative AI tools for the following purposes:

- Idea generation and clarification: ChatGPT assisted me in exploring different ways of framing my research questions, identifying relevant concepts, and stimulating critical thinking when facing conceptual roadblocks.

- Literature review support: ChatGPT helped me review and extract relevant information from academic papers more efficiently.

- Formatting and structure guidance: ChatGPT assisted in organizing the chapters of my thesis according to HEC Montréal's academic conventions and formatting requirements.

- Learning technical tools and languages: ChatGPT supported my learning and mastery of tools and techniques such as SQL, SAS, remote access to SSD environments, SLURM job scheduling, and working with computing clusters and parallelization.

- Code cleaning and efficiency: ChatGPT and Gemini assisted in identifying inefficiencies, and improving the structure and performance of my Python scripts used for modeling and data preprocessing.

- Understanding coding errors and limitations: ChatGPT and Gemini helped me interpret and manage errors and limitations encountered during model development.
- Writing assistance: ChatGPT assisted in improving sentence structure, enhancing clarity, correcting overlooked spelling errors, and rephrasing parts of my thesis draft, while ensuring that the final content reflected my personal academic style.

All GenAI-generated content was carefully reviewed, critically evaluated, and revised to ensure alignment with my own understanding, analytical reasoning, and academic integrity. No confidential data, personal information, or unpublished research was shared with these tools at any time. The final content of this thesis represents my original work, with GenAI tools serving solely as assistants to facilitate my learning and expression.

Appendix B

Equations

$$\text{Momentum}_{\text{open-1pm}} = \frac{Price_{\text{1pm}} - Price_{\text{open}}}{Price_{\text{open}}} \tag{1}$$

$$\text{Turnover}_{\text{1pm}} = \frac{\text{Total Value}_{\text{1pm}}}{\text{Shares Outstanding} \times Price_{\text{close}}} \tag{2}$$

Tables

Table 1: Performance comparison between Lasso and Random Forest models across all forecasting targets (return, TCI, and TVI) and subsets (security types, market capitalization groups, and industry groups) over the test years (2018–2024).

Target	Grouping	Bucket	Lasso			Random Forest		
			MAE	MSE	R^2	MAE	MSE	R^2
Return	Security Type	Common	0.00860567	0.000144121	0.099656137	0.008064547	0.000129468	0.192133577
		ETF	0.003704785	3.81E-05	0.083691472	0.003728769	3.89E-05	0.067149668
	Market Cap	Large Cap	0.005667953	6.52E-05	0.129264079	0.005669224	6.68E-05	0.109368805
		Mid Cap	0.006994432	9.75E-05	0.127108244	0.006722482	9.23E-05	0.174193033
		Small Cap	0.00626812	9.33E-05	0.076738028	0.005667722	7.95E-05	0.213267852
	Industry	BusEq	0.007738247	0.000118053	0.130673907	0.007413707	0.000109689	0.193844167
		Chems	0.006982706	9.63E-05	0.145385288	0.006835355	9.32E-05	0.175626577
		Durbl	0.00777174	0.000117936	0.143221878	0.007611594	0.000112987	0.181149376
		Enrgy	0.00868018	0.000140686	0.1178551	0.008516161	0.00013625	0.149260836
		Hlth	0.008573194	0.000144799	0.111976988	0.008143308	0.00013192	0.192078094
		Manuf	0.007584997	0.000112695	0.132348023	0.007347197	0.000105864	0.187616701
		Money	0.004532822	5.34E-05	0.07809429	0.004333286	4.94E-05	0.149296026
		NoDur	0.007200492	0.00010564	0.104820002	0.006952968	9.87E-05	0.165331267
		Other	0.009761633	0.000179873	0.078788038	0.008895661	0.000156294	0.200158916
		Shops	0.007932379	0.000123	0.11304541	0.007663856	0.00011551	0.16882566
		Telcm	0.006607834	9.16E-05	0.116600083	0.006376402	8.59E-05	0.172298105
		Utils	0.005906981	7.19E-05	0.117241243	0.00581703	6.96E-05	0.146327426
TCI	Security Type	Common	0.079264822	0.017469997	0.574093582	0.046367323	0.00899254	0.78089457
		ETF	0.187975768	0.065973392	0.574343913	0.105932839	0.028491728	0.815845044
	Market Cap	Large Cap	0.041659115	0.004056269	0.569378053	0.02344407	0.001931305	0.796619409
		Mid Cap	0.067471714	0.01026879	0.545922774	0.042799985	0.005603362	0.752222089
		Small Cap	0.153836249	0.046724907	0.583930757	0.103439613	0.027308926	0.757577721
	Industry	BusEq	0.065436187	0.011463039	0.554676404	0.038472428	0.005742467	0.775203238

Target	Grouping	Bucket	Lasso			Random Forest		
			MAE	**MSE**	R^2	**MAE**	**MSE**	R^2
		Chems	0.056426331	0.007677147	0.578294196	0.034529958	0.004238945	0.767324836
		Durbl	0.063956326	0.01019607	0.579608076	0.040173396	0.005731034	0.764906836
		Enrgy	0.062335078	0.009871988	0.557547454	0.03926528	0.005239637	0.764298268
		Hlth	0.067265546	0.011876609	0.554217523	0.039594213	0.006161793	0.768578506
		Manuf	0.070450602	0.013360818	0.576899003	0.042113001	0.006963201	0.77871999
		Money	0.162711175	0.053704257	0.576553516	0.09458074	0.024412088	0.807259352
		NoDur	0.07199569	0.014433091	0.543934849	0.046231619	0.008314448	0.734271653
		Other	0.096326795	0.024695671	0.587718712	0.059392475	0.013728399	0.772271943
		Shops	0.062556715	0.01031503	0.573703194	0.03702229	0.005251042	0.781963353
		Telcm	0.067674523	0.011272607	0.554537109	0.044184953	0.006190424	0.754700232
		Utils	0.056345557	0.007768887	0.553277084	0.03519185	0.004208727	0.757764752
TVI	Security Type	Common	0.099330139	0.028072318	0.55299821	0.074202729	0.020296686	0.677320313
		ETF	0.240474118	0.115171034	0.557719109	0.177418079	0.076831505	0.704833066
	Market Cap	Large Cap	0.055426936	0.007993688	0.514932692	0.041114288	0.005610758	0.657664347
		Mid Cap	0.08551132	0.01700311	0.512565078	0.063774052	0.011939449	0.657212071
		Small Cap	0.180639279	0.065261275	0.5746234	0.141530292	0.048674462	0.685488083
	Industry	BusEq	0.084030421	0.019798163	0.523820922	0.063602179	0.014295451	0.655675649
		Chems	0.073677497	0.01353045	0.539520228	0.056380112	0.010109619	0.656434906
		Durbl	0.0813261	0.016740431	0.549856773	0.064031767	0.012949176	0.652104942
		Enrgy	0.079549567	0.016720487	0.527354895	0.061722545	0.012540184	0.646164935
		Hlth	0.087396499	0.020954822	0.522839702	0.067523823	0.015796715	0.640349713
		Manuf	0.088789402	0.021519483	0.552008893	0.067226127	0.015697014	0.672744175
		Money	0.204594085	0.089764556	0.560565771	0.150567959	0.060239555	0.70502522
		NoDur	0.093416658	0.02503156	0.516329173	0.074110479	0.01957291	0.621704271
		Other	0.117962333	0.038466148	0.573971508	0.092908852	0.030181265	0.668369054
		Shops	0.079795307	0.017606256	0.543637917	0.060241931	0.012745202	0.669553656

Target	Grouping	Bucket	Lasso			Random Forest		
			MAE	**MSE**	R^2	**MAE**	**MSE**	R^2
		Telcm	0.08529584	0.01919738	0.51887447	0.067245239	0.014851393	0.627993049
		Utils	0.073510316	0.013623427	0.509337117	0.056338476	0.010151377	0.633340707

Table 2: Summary of variables and their explanations

Variable Name	**Type**	**Description**
total_size_1pm	Int	Total volume of trades from market open to 1 p.m.
turnover_1pm	Float	Turnover from market open to 1 p.m.
shrount_lag1	Float	Shares Outstanding, lagged by 1 day
vwret	Float	Value-Weighted Return including dividends
log_vwap_price_ratio	Float	Relative VWAP from market open to 1 p.m. (Eq 5.1)
morning_trade_imbalance _count	Float	Trade count imbalance from market open to 1 p.m. (Eq 3.2)
QuotedSpread_Dollar_tw	Float	Time-Weighted Dollar Quoted Spread (During Market Hours)
BestBidDepth_Share_tw	Float	Time-Weighted Best Bid Share Depth (During Market Hours)
BuyNumTrades_Retail	Int	Total Number of Retail Buys (During Market Hours)
BestOffDepth_Share_tw	Float	Time-Weighted Best Offer Share Depth (During Market Hours)
BestBidDepth_Dollar_tw	Float	Time-Weighted Best Bid Dollar Depth (During Market Hours)
intercept	Float	Estimated intercept of the model
bs_ratio_num	Int	Absolute Percent Order Imbalance - Num of Trades
BuyVol_Retail	Float	Sum of Retail Buys, Volume in Shares (During Market Hours)
momentum_open_to_1pm	Float	Momentum from market open to 1 p.m. (Eq 1)
DollarRealizedSpread _LR_Ave	Float	Simple Averaged Dollar Realized Spread (Lee Ready)
morning_trade_imbalance _volume	Float	Trade Volume Imbalance from market open to 1 p.m. (Eq 3.3)
QuotedSpread_Percent_tw	Float	Time-Weighted Percent Quoted Spread (During Market Hours)
BestOfrDepth_Dollar_tw	Float	Time-Weighted Best Offer Dollar Depth (During Market Hours)
vwap_1pm	Float	Volume Weighted Average Price from market open to 1 p.m. (Eq 3.1)

Table 3: Fama–French 12 Industry Classification Descriptions

Industry Name	**Description**
NoDur	Consumer Nondurables – Food, Tobacco, Textiles, Apparel, Leather, Toys

Industry Name	Description
Durbl	Consumer Durables – Cars, TVs, Furniture, Household Appliances
Manuf	Manufacturing – Machinery, Trucks, Planes, Office Furniture, Paper, Commercial Printing
Enrgy	Oil, Gas, and Coal Extraction and Products
Chems	Chemicals and Allied Products
BusEq	Business Equipment – Computers, Software, and Electronic Equipment
Telcm	Telephone and Television Transmission
Utils	Utilities
Shops	Wholesale, Retail, and Some Services (Laundries, Repair Shops)
Hlth	Healthcare, Medical Equipment, and Drugs
Money	Finance
Other	Other – Mines, Construction, Building Materials, Transportation, Hotels, Business Services, Entertainment

Graphs

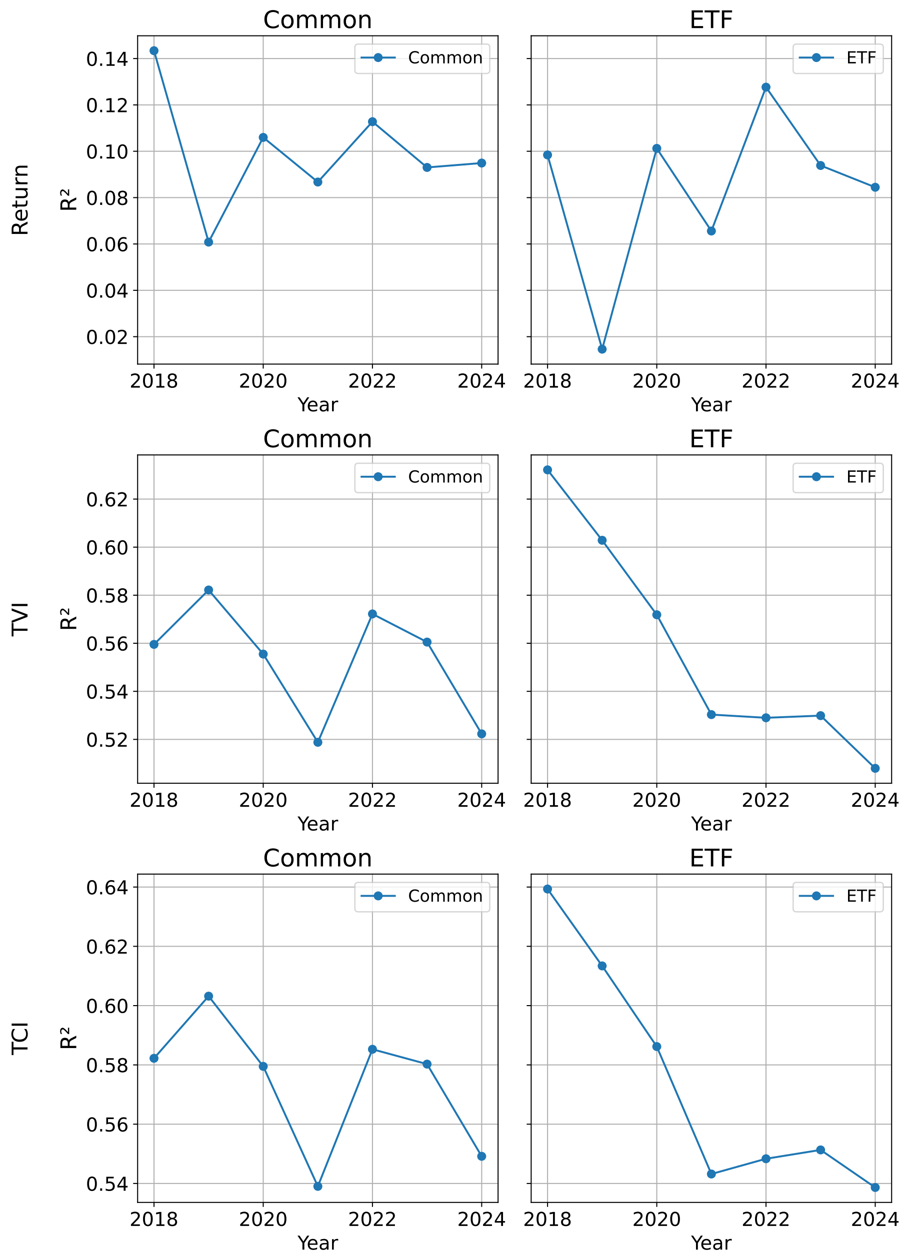


Figure 1: Out-of-sample R^2 comparison in the Lasso model, across security types (common stock vs. ETF) and target variables (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

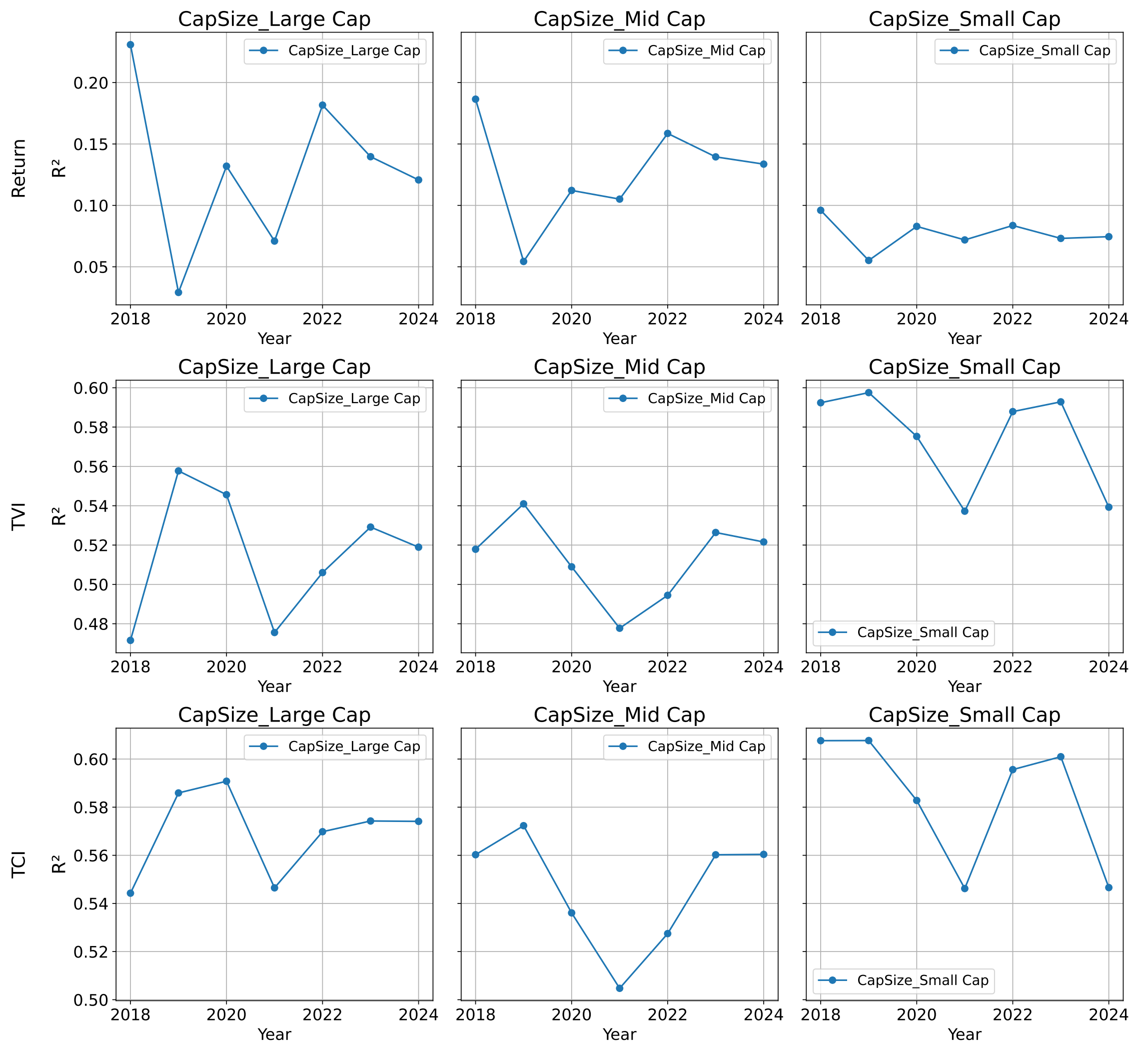


Figure 2: Out-of-sample R^2 comparison in the Lasso model, across market capitalization groups (small, mid and large cap) and target variables (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

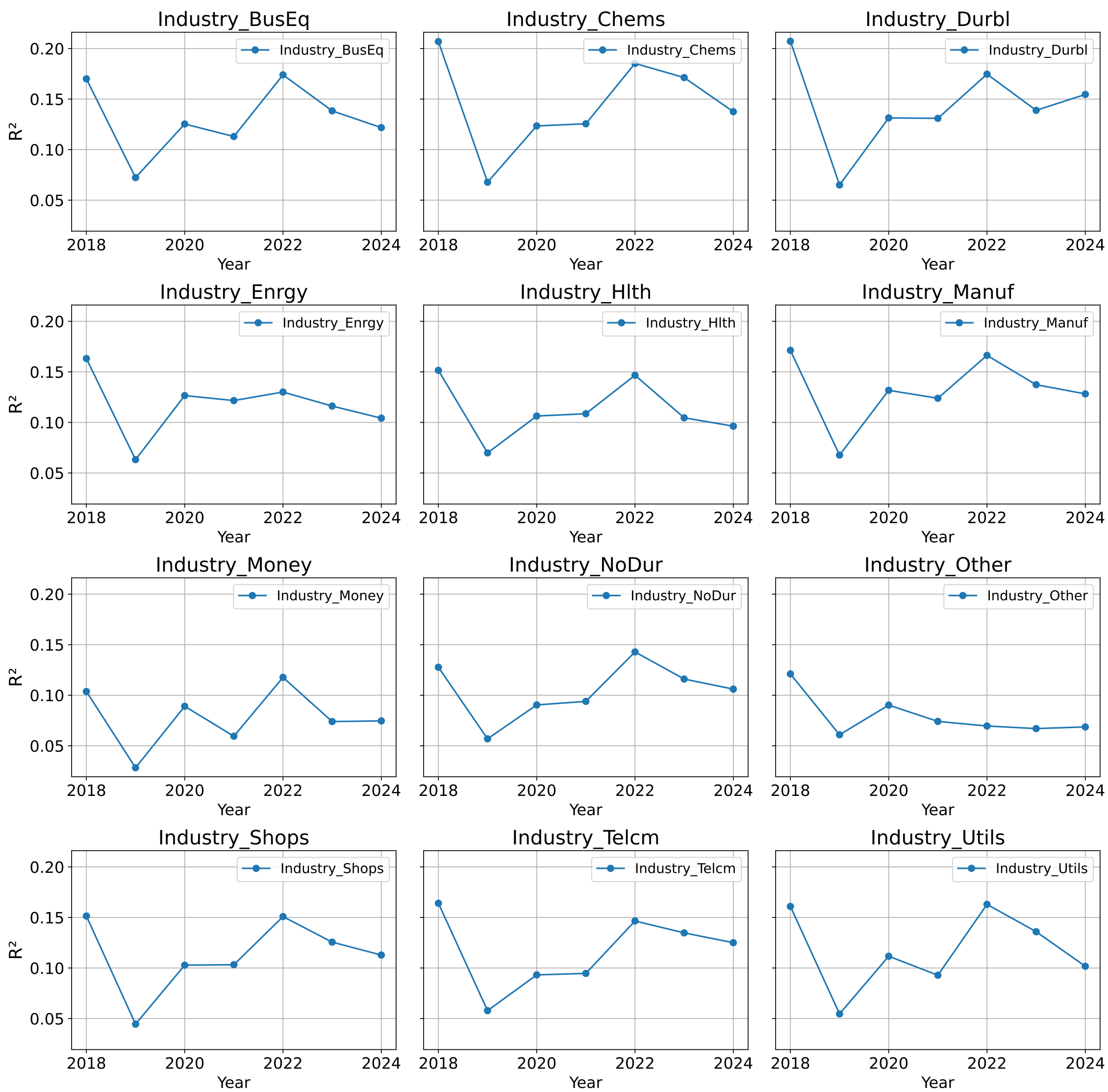


Figure 3: Out-of-sample R^2 comparison in the Lasso model for forecasting returns, disaggregated by industry groups, for test years 2018–2024.

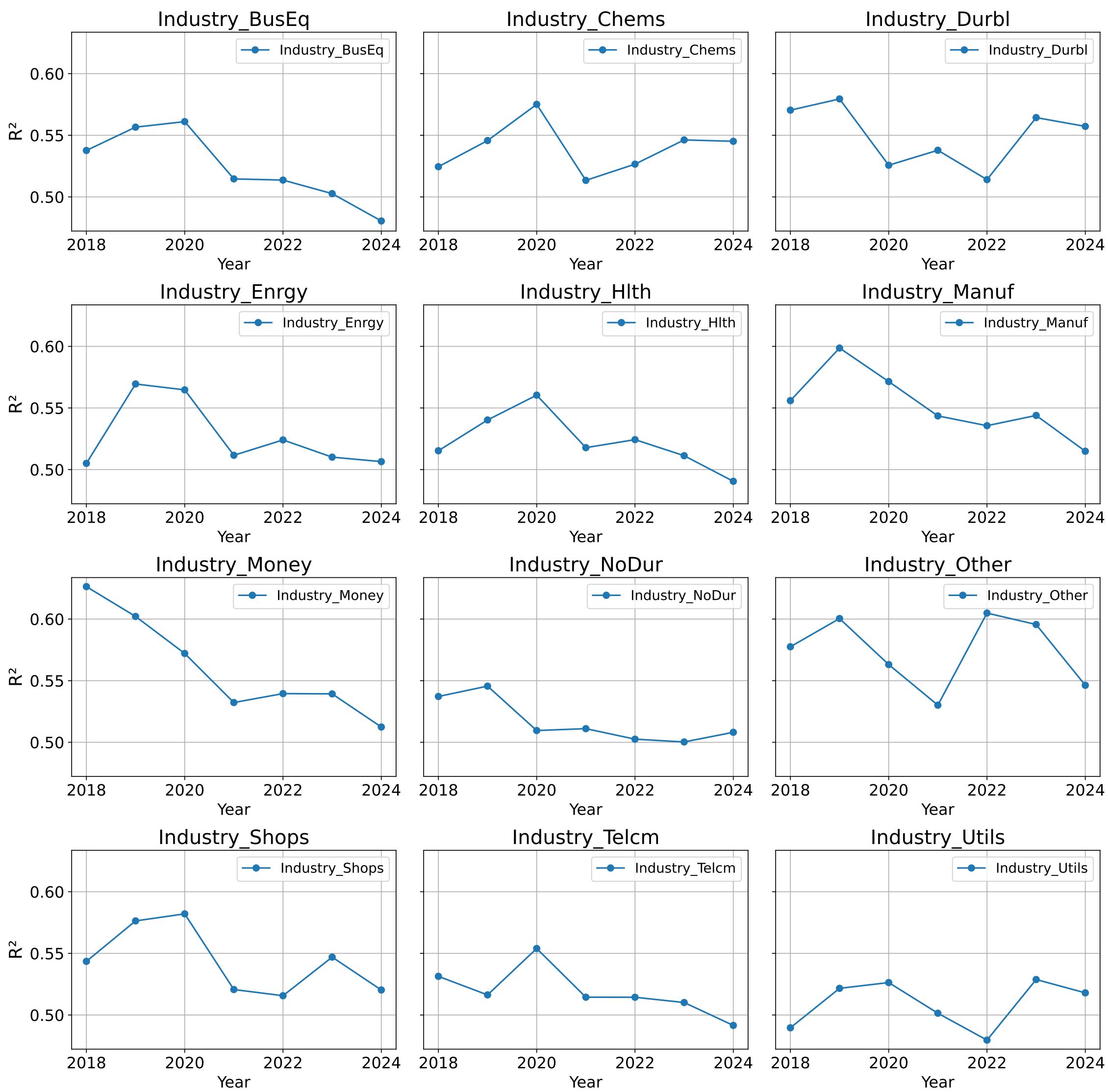


Figure 4: Out-of-sample R^2 comparison in the Lasso model for forecasting Trade Volume Imbalance (TVI), disaggregated by industry groups, for test years 2018–2024.

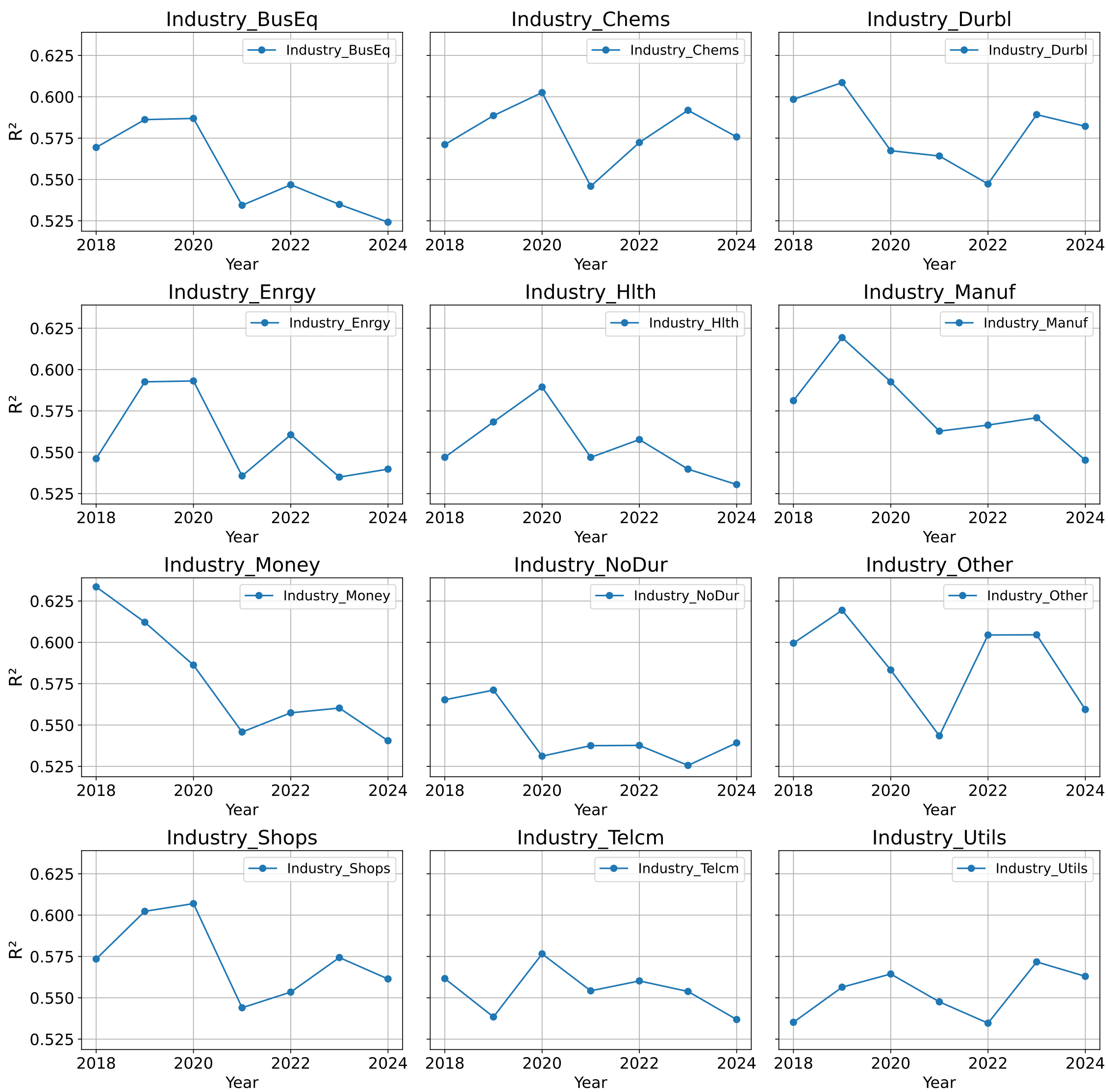


Figure 5: Out-of-sample R^2 comparison in the Lasso model model for forecasting Trade Count Imbalance (TCI), disaggregated by industry groups, for test years 2018–2024.

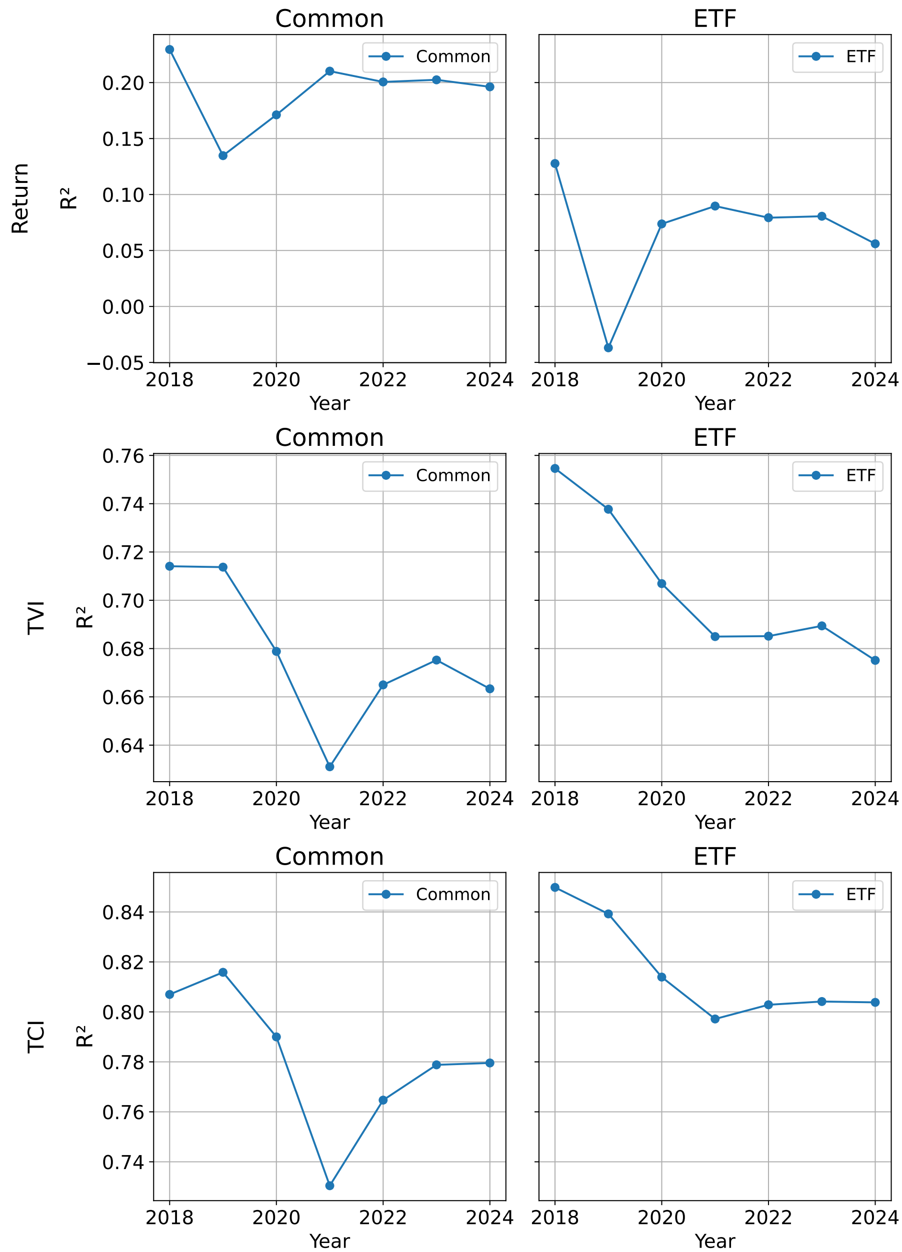


Figure 6: Out-of-sample R^2 comparison in the Random Forest model, across security types (common stock vs. ETF) and target variables (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

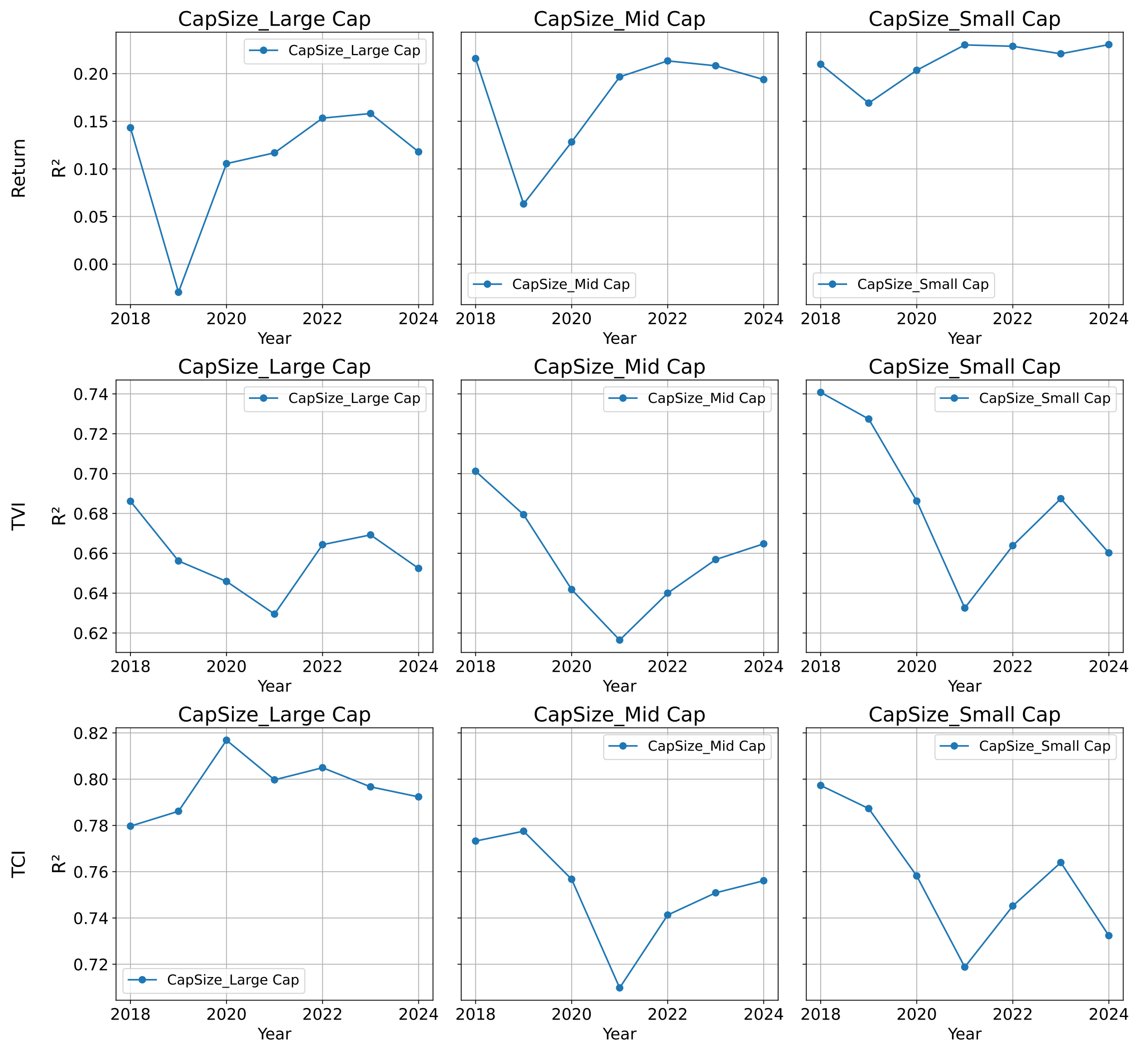


Figure 7: Out-of-sample R^2 comparison in the Random Forest model, across market capitalization groups (small, mid and large cap) and target variables (return, trade count imbalance, trade volume imbalance), for test years 2018–2024.

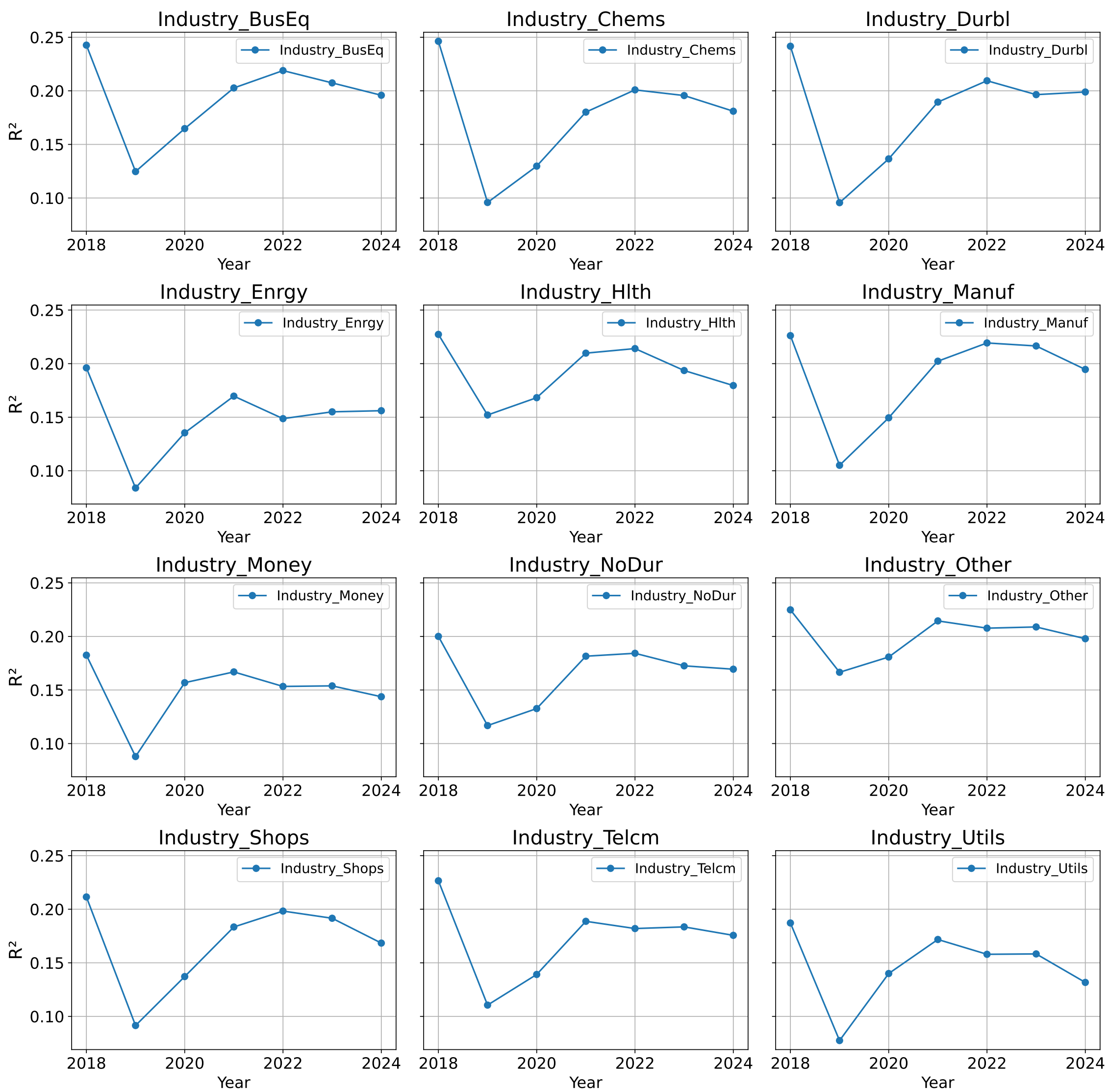


Figure 8: Out-of-sample R^2 comparison in the Random Forest model for forecasting returns, disaggregated by industry groups, for test years 2018–2024.

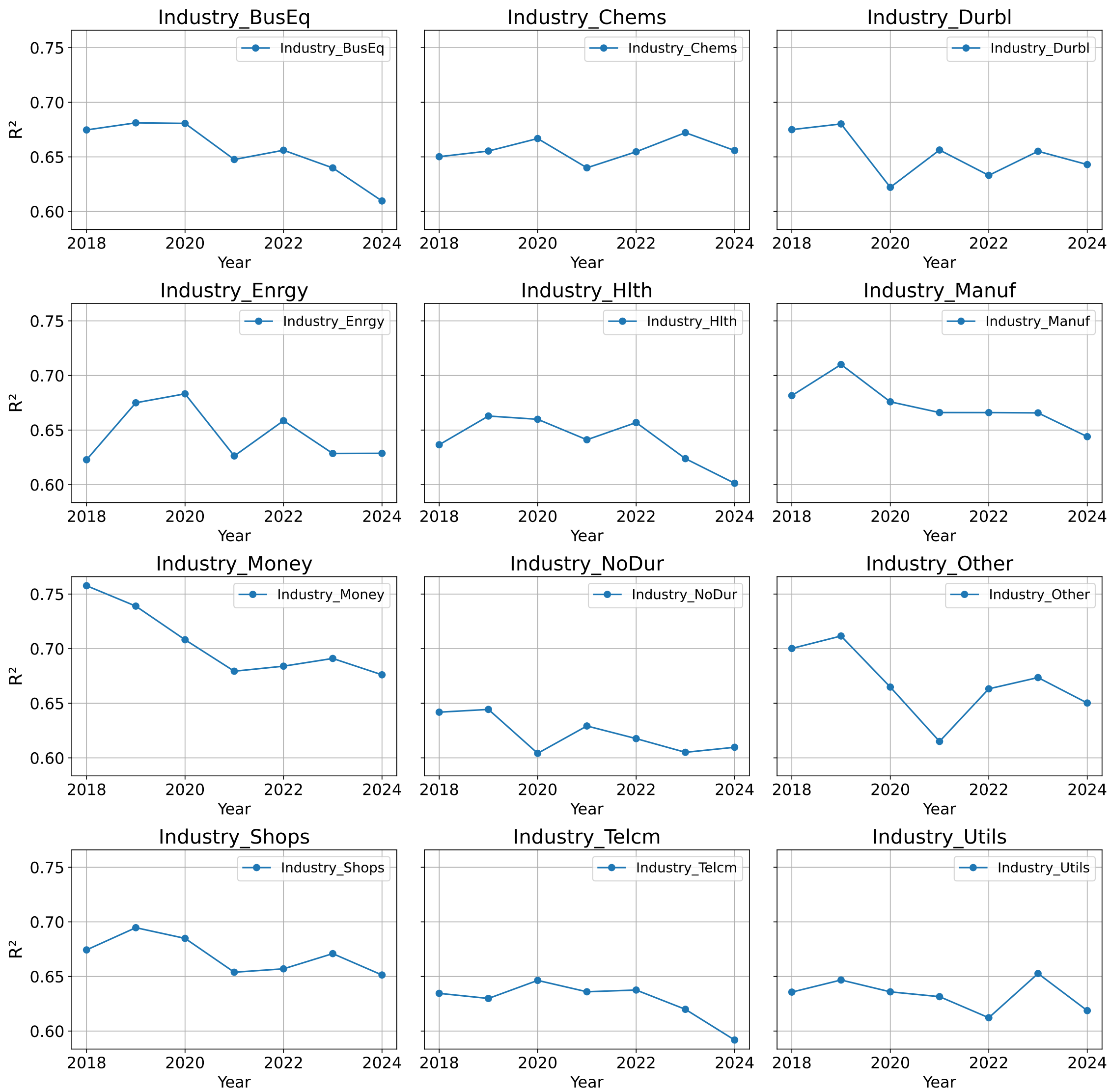


Figure 9: Out-of-sample R^2 comparison in the Random Forest model for forecasting Trade Volume Imbalance (TVI), disaggregated by industry groups, for test years 2018–2024.

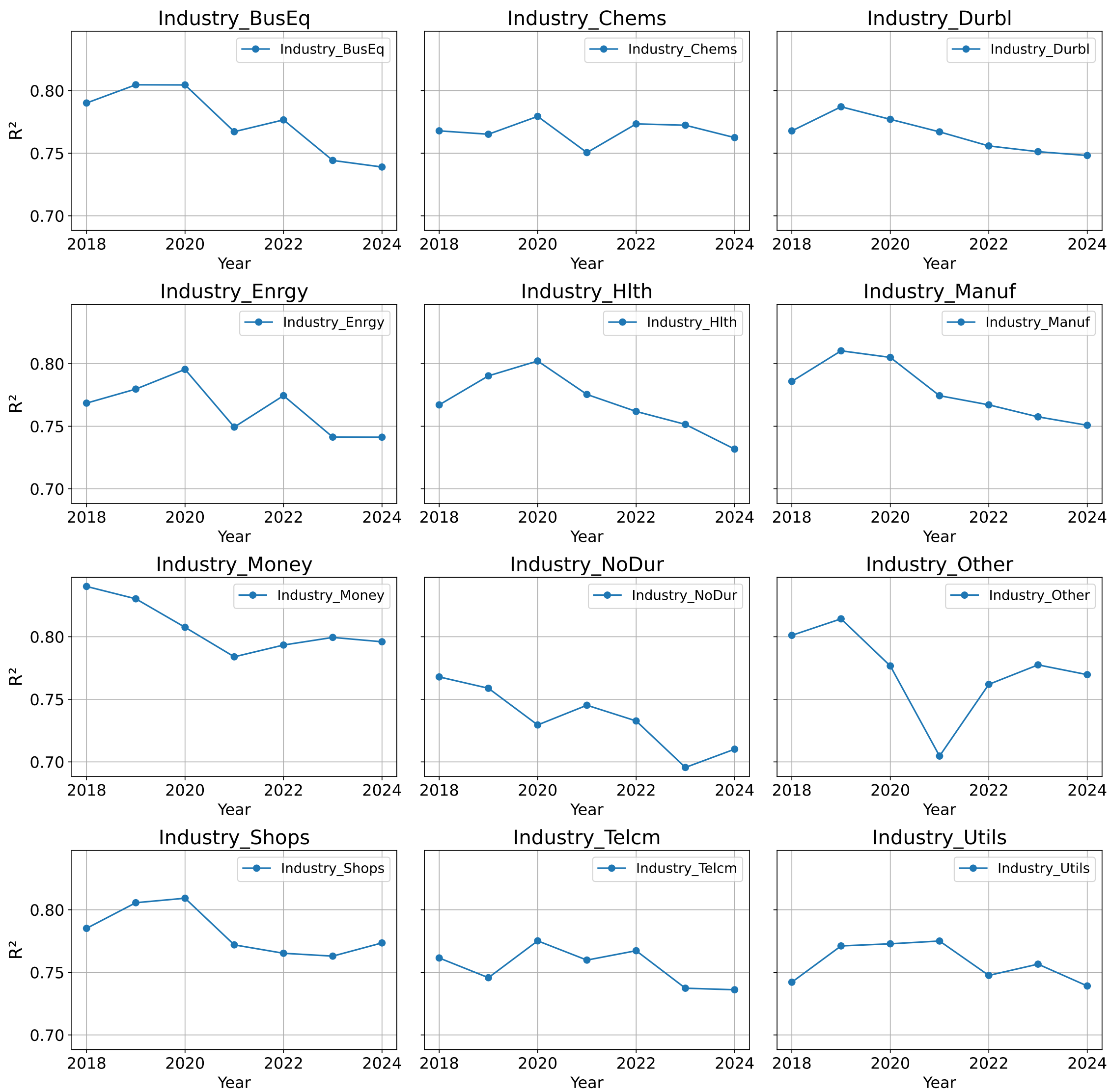


Figure 10:]
Out-of-sample R^2 comparison in the Random Forest model model for forecasting Trade Count Imbalance (TCI), disaggregated by industry groups, for test years 2018–2024.