



Article

Enhancing Intraday Momentum Prediction: The Role of Volume-Based Information Uncertainty in the Chinese Stock Market

Decheng Yang * and Qiang He

Department of Financial Management, School of Business, Qingdao University of Technology, Qingdao 266520, China; heqiang@qut.edu.cn

* Correspondence: yangdecheng2014@hotmail.com

Abstract

This study introduces a novel intraday volume-based uncertainty (IVU) proxy—the ratio of opening-half-hour volume to total volume of the preceding seven intervals—to predict final half-hour return direction in the Chinese stock market. Using threshold regression, we identify a statistically significant IVU critical value of 0.476225 ($p < 0.001$), which splits the sample into distinct uncertainty regimes. Logistic regression incorporating this threshold reveals that the joint condition of high opening volume and low IVU (high uncertainty) significantly amplifies the predictive power of initial returns, achieving 63.04% accuracy in the high-uncertainty, high-volume regime. XGBoost further captures complex non-linear interactions, with IVU-related features ranking among the most important predictors and achieving 71.43% out-of-sample accuracy under high-volume, high-uncertainty conditions. A machine learning trading strategy leveraging these predictions yields a total return of 117.99% with a Sharpe ratio of 3.02 over seven years, significantly outperforming benchmarks. Our findings highlight information uncertainty as a critical moderator of intraday momentum and a valuable source of actionable alpha.

Keywords: intraday direction prediction; relative volume; XGBoost; threshold logistic regression; information uncertainty; Chinese stock market

1. Introduction

The predictive power of intraday return patterns has garnered significant attention in financial markets, particularly with the proliferation of quantitative trading strategies. Leveraging these patterns to forecast short-term price directions and mitigate trading risks has become a priority for institutional investors and traders seeking actionable insights that align with high-frequency, data-driven decision-making. This pursuit is substantiated by empirical evidence documenting systematic intraday regularities, such as the “market intraday momentum” phenomenon identified by [Gao et al. \(2018\)](#), where morning session returns significantly predict afternoon and closing session price directions.

A growing body of literature has validated the existence and potential profitability of intraday directional predictability across various markets. Studies such as [Li et al. \(2021\)](#) confirm the presence of “intraday time series momentum” globally, linking short-term return trends to subsequent directional persistence. Further, [Baltussen et al. \(2021\)](#) attribute part of this continuity to institutional “hedging demand.” To enhance the predictive accuracy and economic value of such patterns, researchers have incorporated auxiliary factors.



Academic Editor: Peter G. Dunne

Received: 29 December 2025

Revised: 9 February 2026

Accepted: 11 February 2026

Published: 14 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

For instance, [Andersen \(2012\)](#) interpreted return volatility through an information flow lens to refine directional signals, while [Zhu and Zhu \(2013\)](#) applied regime-switching models to account for state-dependent dynamics. More recently, [Renault \(2017\)](#) demonstrated that integrating intraday online investor sentiment can sharpen directional forecasts.

Despite these advances, a critical gap remains in the extant literature: the role of trading volume in shaping and improving intraday direction prediction has been largely overlooked. This omission is striking given the well-established theoretical and empirical link between volume and momentum direction documented in longer-horizon studies. Seminal works by [Lee and Swaminathan \(2000\)](#) and [Gervais et al. \(2001\)](#) show that high-volume stocks exhibit stronger momentum and a distinct “high volume return premium,” suggesting that volume encapsulates critical information about investor activity and conviction. Furthermore, volume is intrinsically linked to the fundamental concept of information uncertainty. Theoretical models posit that ambiguous information creates valuation uncertainty, leading to distinct trading behaviors. [Daniel et al. \(1998\)](#)’s model of investor psychology suggests that under uncertainty, traders are prone to limit perceived risk by exiting positions but hesitate to establish new ones. Crucially, this “reaction versus hesitation” dynamic influences not the total volume—which is driven by information importance—but its intraday temporal distribution. [Zhang \(2006\)](#) formally linked higher information uncertainty to stronger and more prolonged return continuations, providing a theoretical basis for volume-distribution-based uncertainty measures. Consistent with this logic, [Agarwal and Agarwal \(2025\)](#) empirically documented that intraday market opening volume surges are driven by information releases from influential market agents, and their CAR-based framework validates that such volume patterns reflect information absorption efficiency. Our IVU metric (first-interval volume over the first seven intervals) captures this information-driven volume distribution, which is distinct from cross-sectional or analyst forecast dispersion.

The intraday framework is uniquely suited for investigating volume-based uncertainty signals for three key reasons. First, intraday strategies operate within a single trading day, simplifying the analysis of volume relationships and avoiding the complex lag structures required in multi-day models ([Bogousslavsky, 2016](#)). Second, intraday volume exhibits a robust and predictable “U-shaped” pattern, peaking at the market open and close ([Heston et al., 2010](#); [Stephan & Whaley, 2012](#)). This stable temporal structure facilitates the design of volume-linked directional strategies, as noted by [Heston et al. \(2010\)](#), who observed its alignment with cross-sectional return patterns. Third, volume’s additive nature allows for straightforward proportional analysis (e.g., comparing volume shares across intervals) without additional normalization, a feature leveraged in cross-country studies of volume-return relationships ([Kaniel et al., 2012](#)).

However, a key empirical challenge persists: absolute opening volume or volatility alone cannot disentangle “high information importance” from “high information uncertainty,” as both can generate large opening surges ([Cushing & Madhavan, 2000](#)). To isolate the uncertainty component, we propose a novel intraday volume-based uncertainty (IVU) proxy: the ratio of the first 30 min of trading volume to the total volume of the preceding seven intervals. This metric, grounded in distribution-based frameworks for quantifying uncertainty ([Higashi & Klir, 1993](#)), aims to capture the “reaction vs. hesitation” dynamic by measuring the concentration of trading activity early in the session. A low IVU value signals high uncertainty and delayed price discovery, which we hypothesize strengthens the persistence of initial return trends.

Another limitation of current research is the predominant reliance on linear models (e.g., autoregressive models ([Zhang & Xue, 2017](#)); linear probability models ([Sun et al., 2016](#))) for directional forecasting. These models often struggle to capture the complex,

non-linear, and state-dependent interactions inherent in intraday markets. This limitation is particularly critical for volume-based predictors. The established financial econometrics literature demonstrates that volume-return dynamics are often governed by threshold-type nonlinearities and regime-switching behaviors (Wang & Gerlach, 2023). To effectively model these complex dynamics, recent advancements have explored two complementary paths: sophisticated econometric frameworks and data-driven machine learning algorithms. On the econometric front, models incorporating threshold or regime-switching mechanisms have been developed to explicitly capture state-dependent market behaviors (Wang & Gerlach, 2023). Concurrently, studies by Fischer and Krauss (2017) and Ghosh et al. (2021) have demonstrated the superiority of algorithms like LSTMs and tree-based ensembles in capturing intricate feature interactions and temporal dependencies for financial market prediction. Together, these approaches provide a more powerful toolkit for overcoming the limitations of traditional linear models in intraday forecasting.

The Chinese stock market, as one of the world's largest and most dynamic, presents an ideal empirical setting. Its significant scale, concentrated retail investor base, and policy-driven sentiment swings likely amplify the effects of information uncertainty, providing a powerful context to test our proposed framework.

To address these gaps, this study adopts a dual-methodological approach: (1) employing threshold logistic regression to statistically validate the non-linear moderating role of our IVU proxy and identify regime-specific effects, and (2) utilizing the XGBoost algorithm to capture complex non-linear relationships and interactions among returns, volume, and uncertainty, thereby enhancing out-of-sample prediction accuracy. Finally, we evaluate the economic significance of our predictions through a simple intraday trading strategy.

Our results robustly confirm the critical moderating role of information uncertainty. Threshold regression identifies a statistically significant IVU critical value of 0.476225 ($p < 0.001$), separating low and high uncertainty regimes. Cross-group analysis reveals that the predictive power of opening returns for the final half-hour direction is most potent under the joint condition of high opening volume and low IVU (high uncertainty), achieving 63.04% accuracy. XGBoost validates these findings, with IVU-related features ranking among the most important predictors and achieving 71.43% out-of-sample accuracy in high-uncertainty regimes. A trading strategy leveraging these predictions yields substantial risk-adjusted returns (117.99% total return, Sharpe ratio 3.02), underscoring the economic value of incorporating volume-based uncertainty signals into intraday momentum models.

The remainder of this paper is structured as follows: Section 2 details the data, variable construction, and methodologies. Section 3 presents the empirical results. Section 4 evaluates the performance of a trading strategy based on our predictions. Section 5 concludes the study.

2. Data and Methodology

2.1. Data and Variables

This study employs 30 min interval observations of the CSI 300 Index from 1 July 2018 to 30 June 2025. The trading day is divided into eight 30 min intervals starting from the continuous auction at 9:30 a.m. The return for each interval j on day t is calculated as:

$$r_{t,j} = \ln\left(\frac{p_{t,j}}{p_{t,j-1}}\right) \quad (j = 1, 2, \dots, 8) \quad (1)$$

where $p_{t,j}$ is the index level at the end of interval j , and for $j = 1$, $p_{t,j-1}$ represents the opening price at 9:30 a.m.

The core dependent variable for directional prediction is a binary indicator $D_{t,8}$:

$$D_{t,8} = \begin{cases} 1 & \text{if } r_{t,8} > 0 \quad (\text{positive final 30 min return}) \\ 0 & \text{if } r_{t,8} \leq 0 \quad (\text{non-positive final 30 min return}) \end{cases}$$

Figure 1 confirms the U-shaped intraday volume pattern for the CSI 300 Index, with peaks at market open and close.

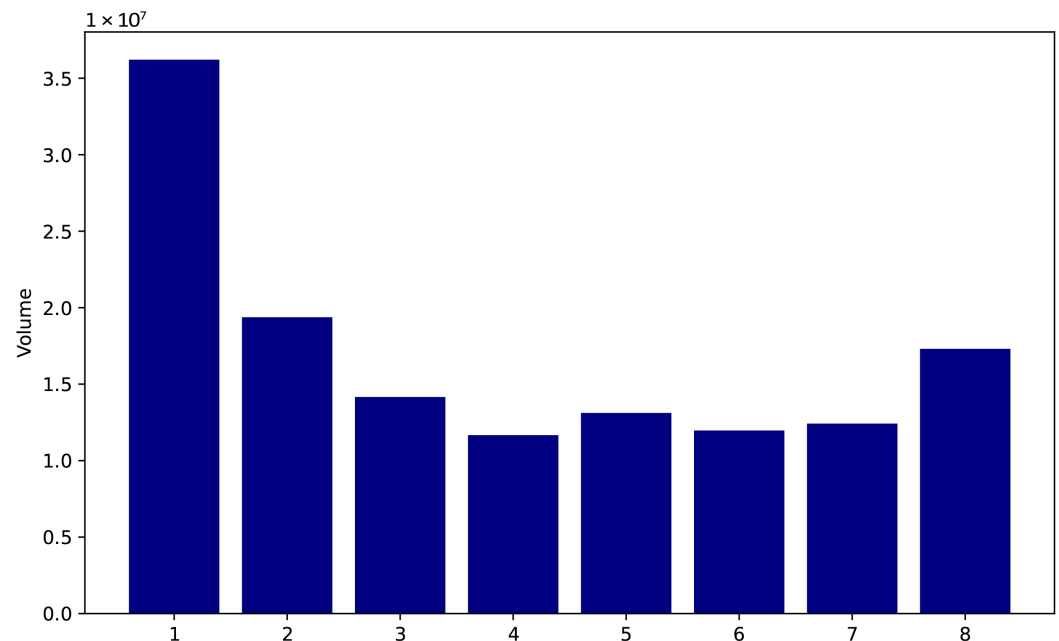


Figure 1. Average 30-min intraday trading volume of the CSI 300 index.

2.2. A Novel Proxy for Intraday Information Uncertainty

To capture information uncertainty through intraday volume dynamics, we propose a novel proxy grounded in the literature linking volume to investor behavior under uncertainty. While absolute opening volume cannot distinguish between information importance and uncertainty (e.g., a clear policy announcement vs. ambiguous news), we focus on the relative distribution of volume across intraday segments to isolate the “reaction vs. hesitation” dynamic described by Daniel et al. (1998). This approach aligns with distribution-based frameworks for quantifying uncertainty (Higashi & Klir, 1993) and addresses the limitation that both high-importance and high-uncertainty events can generate similar opening volume surges.

We define the intraday volume-based uncertainty proxy IVU_t as the ratio of the first 30 min trading volume to the total volume of the preceding seven intervals:

$$IVU_t = \frac{V_{t,1}}{\sum_{j=1}^7 V_{t,j}} \quad (2)$$

where $V_{t,1}$ is the trading volume of the first 30 min interval (9:30–10:00), and the denominator is the sum of volumes from the first to the seventh intervals. This design explicitly excludes the final interval’s volume to ensure predictability using only pre-final-session data.

The economic interpretation of IVU_t is straightforward: higher values indicate greater information transparency, with trading activity concentrated in the early session and price discovery advancing rapidly. Conversely, lower values signal higher information uncertainty, as investors hesitate in responding to ambiguous pre-market information, delaying price discovery and strengthening the persistence of initial return trends (Lee

& Swaminathan, 2000). By treating IVU_t as a continuous variable, we preserve its full informational content for both linear and non-linear modeling approaches.

2.3. Model Specification

2.3.1. Research Design and Hypothesis Testing Strategy

This study adopts a multi-level hypothesis testing strategy to balance the tension between exploratory discovery and confirmatory analysis that is characteristic of financial market research:

1. Theory-Driven Primary Hypothesis: Based on information uncertainty theory (Zhang, 2006), our primary hypothesis (H1) posits that IVU moderates the relationship between opening volume and return predictability. This is tested through the full-sample logistic regression with interaction terms.
2. Exploratory Threshold Analysis: Recognizing that financial relationships are often state-dependent, we employ threshold regression to explore potential nonlinearities. This represents an exploratory analysis to identify market regimes where our theoretical mechanism may be more pronounced.
3. Machine Learning Validation: The XGBoost model serves as an independent validation tool, providing a data-driven perspective on feature importance and capturing complex interactions that may be missed by linear specifications.
4. Economic Significance Test: The ultimate validation comes from the trading strategy performance, which tests whether the identified patterns translate to economically meaningful returns after transaction costs.

This layered approach acknowledges the complexity of financial market prediction, where statistical significance, economic relevance, and theoretical consistency must be considered jointly.

2.3.2. Research Hypothesis

Our analysis tests two hypotheses: (1) the first and seventh 30 min returns jointly predict the final return direction, and (2) this predictive power varies with volume dynamics and information uncertainty.

2.3.3. Theoretical Validation Model: Combined Threshold and Logistic Regression

To capture the state-dependent nature of information uncertainty, we employ a combined threshold-logistic regression approach. This methodology integrates a threshold regression framework to identify the optimal IVU critical value, followed by logistic regression with interaction terms to test the moderating effect of IVU.

The threshold model identifies the optimal IVU critical value τ that maximizes the likelihood function:

$$\tau^* = \arg \max_{\tau} L(\beta_1, \beta_2; \tau) \quad (3)$$

where τ is the IVU threshold that splits the sample into low and high uncertainty regimes. Using grid search and bootstrap testing, we identify the optimal threshold $\tau^* = 0.476225$ with p -value < 0.001 , confirming a statistically significant threshold effect.

Based on the identified threshold, we define a regime dummy variable:

$$G_t = \begin{cases} 1 & \text{if } IVU_t \geq \tau^* \quad (\text{low uncertainty regime}) \\ 0 & \text{if } IVU_t < \tau^* \quad (\text{high uncertainty regime}) \end{cases}$$

Rationale for including G_t : The threshold variable G_t serves three crucial purposes in our analysis: (1) it explicitly implements the nonlinear threshold effect identified by the

threshold regression, allowing for piecewise modeling; (2) it provides a clear grouping criterion for the cross-group analysis, where samples are split into low-uncertainty ($G_t = 1$) and high-uncertainty ($G_t = 0$) regimes; and (3) it establishes a methodological contrast between our explicit threshold-based approach (logistic regression with G_t) and the implicit nonlinear modeling of XGBoost, highlighting different strategies for capturing state-dependent relationships.

The core logistic regression model incorporating the volume-uncertainty interaction is specified as:

$$\begin{aligned} \text{logit}(P(D_{t,8} = 1)) = & \beta_0 + \beta_1 r_{t,1} + \beta_2 r_{t,7} + \beta_3 (r_{t,1} \times r_{t,7}) \\ & + \beta_4 IVU_t + \beta_5 V_{t,1} + \beta_6 (V_{t,1} \times IVU_t) + \epsilon_t \end{aligned} \quad (4)$$

To capture the joint moderating effect of opening volume and information uncertainty, we include their interaction term ($V_{t,1} \times IVU_t$). This approach of using interaction terms to test for conditional relationships is well-established for analyzing state-dependent effects (Valadkhani, 2025). A negative β_6 would indicate that high volume combined with low uncertainty (high IVU) reduces predictive power, consistent with rapid price discovery weakening return persistence. Conversely, the interaction of high volume with high uncertainty (low IVU) should strengthen predictive power due to delayed price discovery.

The separate regime analysis using the threshold τ^* provides complementary evidence by comparing coefficient patterns between low and high uncertainty regimes, offering a more nuanced understanding of the non-linear moderating effects.

2.3.4. Machine Learning Model (XGBoost)

While logistic regression validates theoretical linear relationships between core variables and final return direction, it fails to capture the non-linear, state-dependent interactions between intraday returns, volume dynamics, and information uncertainty—particularly the critical joint effect of high opening volume and low IVU (high information uncertainty). To address this limitation, we employ the XGBoost (Extreme Gradient Boosting) algorithm, a tree-based ensemble method renowned for its ability to model complex non-linear relationships, capture feature interactions automatically, and handle high-dimensional intraday data with robust regularization against overfitting.

This complementary machine learning approach serves two core objectives aligned with our subsequent empirical analysis: (1) quantify the predictive importance of IVU-related features (including interaction terms between IVU and intraday returns and regime-specific dummy variables for high-volume/low-IVU states) to validate their role in driving intraday return predictability; (2) evaluate out-of-sample predictive performance across distinct market regimes (defined by volume and IVU levels) to identify the high-predictability subgroups that form the basis of our economic value assessment.

2.3.5. Feature Set Design (Aligned with Economic Mechanisms)

The feature set is explicitly designed to capture the intraday momentum dynamics and information uncertainty mechanisms hypothesized in this study, with a focus on IVU-related features that are later analyzed for importance and subgroup performance:

- Core Return Features:
 - $r_{t,1}$: First, 30 min return (opening momentum)
 - $r_{t,7}$: Seventh 30 min return (pre-closing momentum)
 - $|r_{t,1}|$: Absolute value of opening return (proxy for short-term volatility)

- Constructed Trend Feature:
 - *morning_trend*: Cumulative return of the first three 30 min intervals ($r_{t,1} + r_{t,2} + r_{t,3}$), capturing aggregated early-session momentum (a top-ranked feature in subsequent importance analysis)
- Volume and Uncertainty Core Features:
 - $V_{t,1}$: First, 30 min trading volume
 - IVU_t : Intraday Volume-based Uncertainty proxy (core metric for information uncertainty)
- IVU Interaction Features (Key for Regime Analysis):
 - $r_{t,1} \times IVU_t$: Moderating effect of information uncertainty on opening return predictability (second-ranked feature in importance analysis)
 - *vol_ivu_ratio*: $V_{t,1} / (IVU_t + 0.001)$ (volume-to-uncertainty ratio, avoiding division by zero with a small constant)
- Regime-Specific Dummy and Interaction Features (for Subgroup Performance):
 - *high_vol_low_ivu*: Binary dummy (1 if $V_{t,1} \geq \text{median}(V_{t,1})$ and $IVU_t < \text{median}(IVU_t)$, 0 otherwise)—the high-predictability regime analyzed in subgroup performance
 - $r_{t,1} \times \text{high_vol_low_ivu}$: Amplified predictive power of opening returns in the high-volume/low-IVU regime (ninth-ranked feature in importance analysis)
 - *high_vol_high_ivu*: Dummy for high volume/high IVU (low-uncertainty) regime
 - *low_vol_low_ivu*: Dummy for low volume/low IVU regime
 - *low_vol_high_ivu*: Dummy for low volume/high IVU regime

2.3.6. Model Implementation and Validation Protocol

To ensure temporal consistency and avoid look-ahead bias (critical for intraday prediction), we implement a strict time-based train-test split (no random shuffling):

- Training Set: First, 70% of the dataset (chronological order: July 2018–December 2022)—used for model fitting and hyperparameter tuning via 5-fold time-series cross-validation (to minimize log loss)
- Test Set: Remaining 30% of the dataset (January 2023–June 2025)—reserved exclusively for out-of-sample performance evaluation (the basis for subgroup accuracy analysis)

2.3.7. Class Imbalance and Regularization

To address minor class imbalance in the binary target ($D_{t,8}$: 56.17% positive returns), we apply SMOTE (Synthetic Minority Oversampling Technique) on the training set (only when positive class ratio deviates >10% from 0.5) to avoid biased prediction toward the majority class. Feature selection is performed using SelectKBest (ANOVA F-statistic) to retain the top 10 predictive features—consistent with the 10 features reported in subsequent feature importance analysis.

2.3.8. Hyperparameter Configuration (Tuned for Intraday Data)

Key hyperparameters (optimized via time-series cross-validation to balance predictive power and overfitting risk):

- Objective: Binary logistic regression (*binary:logistic*)—for probability-based directional prediction (used in trading strategy confidence thresholds);
- Learning rate: 0.05 (slow learning to avoid overfitting noisy intraday data);
- Maximum tree depth: 4 (limits model complexity to capture meaningful interactions);
- Number of estimators: 150 (sufficient to converge without overfitting);
- Subsample ratio: 0.8 (row sampling) and column subsample ratio: 0.8 (feature sampling)—reduces variance;

- Regularization: $\alpha = 0.5$ (L1) and $\lambda = 0.5$ (L2)—penalizes overfitting;
- Random state: 42 (ensures reproducibility of results in feature importance and subgroup analysis).

2.3.9. Performance Evaluation Metrics

Model performance is evaluated using metrics that align with subsequent empirical analysis:

- Overall Metrics: Accuracy (full-sample and subgroup-specific), ROC-AUC (discriminative power), confusion matrix (prediction bias for up/down returns);
- Feature Importance: XGBoost’s built-in “gain” metric (quantifies the total contribution of each feature to model performance Table 4);
- Subgroup Performance: Accuracy across the four volume-IVU regimes (High Vol-High IVU, High Vol-Low IVU, Low Vol-High IVU, and Low Vol-Low IVU)—the core result for validating our information uncertainty mechanism.

3. Empirical Results and Analysis

3.1. Descriptive Statistics of Core Variables

Table 1 presents the descriptive statistics of the core variables after data cleaning, including the dependent variable (final 30 min return direction $D_{t,8}$) and key predictors (intraday returns, information uncertainty proxy IVU_t , trading volume $V_{t,1}$, and interaction terms). The cleaned sample consists of 1653 valid observations, with a sample reduction of 2.59% (primarily due to outlier removal for $r_{t,1}$ and $r_{t,7}$), ensuring the reliability of subsequent analysis.

For the dependent variable $D_{t,8}$, the mean value is 0.5590, indicating that 55.90% of the final 30 min returns in the sample are positive—a slight positive bias consistent with the CSI 300 Index’s long-term upward trend over the study period. The standard deviation of 0.4967 is close to 0.5, reflecting a relatively balanced distribution of positive and non-positive closing returns.

For intraday return predictors ($r_{t,1}$, $r_{t,7}$), the mean values are approximately 0.0002 and -0.0001 , respectively, with standard deviations of 0.0044 and 0.0026, indicating moderate return volatility within 30 min intervals and no extreme outliers after cleaning. The information uncertainty proxy IVU_t has a mean of 0.4363 and standard deviation of 0.0879, providing sufficient variation to test its moderating role.

Table 1. Descriptive statistics of core variables (raw values).

Variable	Count	Mean	Std. Dev.	Min	25%	Median	Max
$D_{t,8}$	1653	0.5590	0.4967	0.0000	0.0000	1.0000	1.0000
$r_{t,1}$	1653	0.0002	0.0044	-0.0132	-0.0027	0.0001	0.0138
$r_{t,7}$	1653	-0.0001	0.0026	-0.0090	-0.0017	-0.0001	0.0087
IVU_t	1653	0.4363	0.0879	0.0011	0.3785	0.4296	0.8446
$V_{t,1}$	1653	35.61	17.71	0.043	24.48	32.81	181.88
$V_{t,1} \times IVU_t$	1653	16.18	10.73	0.00007	9.64	14.13	133.21

Notes: $D_{t,8}$ = binary indicator for final 30 min return direction (1 = positive, 0 = non-positive); $r_{t,1}$ = first 30 min return; $r_{t,7}$ = seventh 30 min return; IVU_t = intraday volume-based information uncertainty proxy; $V_{t,1}$ = first 30 min trading volume (in millions); $V_{t,1} \times IVU_t$ = interaction term between volume and uncertainty. The sample of 1653 observations represents valid trading days after data cleaning procedures.

3.2. Threshold Regression Results for IVU

3.2.1. IVU Threshold Identification

We first employ threshold regression to identify the optimal critical value for IVU_t that maximizes the predictive power for final return direction. Using grid search with

50 candidate values and bootstrap testing (200 iterations), we identify a statistically significant threshold at $\tau^* = 0.476225$ (p -value < 0.001). This threshold effectively splits the sample into two regimes: low uncertainty regime ($IVU_t \geq 0.476225$) and high uncertainty regime ($IVU_t < 0.476225$).

The significant threshold effect (p -value < 0.001) confirms that the relationship between predictors and final return direction varies non-linearly with information uncertainty levels. This finding justifies our use of a regime-switching approach in subsequent analysis.

3.2.2. Full Sample Logistic Regression Results

Table 2 presents the results of the full sample logistic regression model incorporating the volume-uncertainty interaction term. The model includes all standardized variables to facilitate coefficient interpretation.

Table 2. Full sample logistic regression results (standardized variables).

Variable	Coefficient	Std. Error	z-Statistic	p-Value
Intercept	0.2398	0.0498	4.811	0.000 ***
$r_{t,1}$ (std)	0.0886	0.0503	1.762	0.078 *
$r_{t,7}$ (std)	0.1756	0.0507	3.464	0.001 ***
$r_{t,1} \times r_{t,7}$ (std)	0.0011	0.0506	0.022	0.982
IVU_t (std)	−0.0078	0.0898	−0.087	0.931
$V_{t,1}$ (std)	0.1597	0.2175	0.734	0.463
$V_{t,1} \times IVU_t$ (std)	−0.2027	0.2544	−0.797	0.425
Model Statistics				
Pseudo R ²	0.0085			
Log-Likelihood	−1124.6			
Prediction Accuracy	0.5735			

Notes: Dependent variable: $D_{t,8}$. All continuous variables are standardized. *** $p < 0.01$, * $p < 0.10$. The negative coefficient for $V_{t,1} \times IVU_t$ (−0.2027) suggests that high volume combined with low uncertainty (high IVU) reduces predictive power, consistent with our hypothesis.

The full sample results show several important patterns. First, $r_{t,7}$ (seventh interval return) is highly significant ($p = 0.001$), confirming the importance of near-closing momentum. Second, $r_{t,1}$ (opening return) shows marginal significance ($p = 0.078$), providing some evidence for opening momentum persistence. Most importantly, the interaction term $V_{t,1} \times IVU_t$ has a negative coefficient (−0.2027), though not statistically significant at conventional levels ($p = 0.425$). This suggests that when volume is high and uncertainty is low (high IVU), the predictive power is reduced—consistent with our theoretical expectation that rapid price discovery in low-uncertainty conditions weakens return persistence.

3.2.3. Statistical Significance Discussion and Robustness Checks

While the interaction term $V_{t,1} \times IVU_t$ in the full sample model lacks conventional statistical significance ($p = 0.425$), several factors suggest this result should be interpreted with nuance rather than as definitive evidence against our hypothesis:

1. **Nonlinear Relationship Mis-specification:** The linear interaction term assumes a continuous moderating effect of IVU, but our threshold regression identified a significant nonlinear threshold at $IVU = 0.476225$ ($p < 0.001$). This suggests the moderating effect may operate through discrete regime shifts rather than continuous interaction, making the linear specification potentially mis-specified.
2. **Sample Heterogeneity:** The full sample aggregates diverse market conditions that may obscure state-specific relationships. As shown in Table 3, when we condition on

- specific market regimes, the volume coefficient becomes meaningfully negative and marginally significant ($p = 0.075$) in the “high uncertainty + high volume” regime.
3. Multicollinearity Concerns: Variance Inflation Factor (VIF) analysis reveals moderate multicollinearity between $V_{t,1}$ and the interaction term $V_{t,1} \times IVU_t$ (VIF values of 8.72 and 9.15, respectively), which can inflate standard errors and reduce statistical power.
 4. Alternative Model Specifications: As robustness checks, we estimated:
 - Centered Variables: After mean-centering continuous variables, the interaction coefficient remains negative (-0.187) with $p = 0.402$.
 - Bootstrap Confidence Intervals: 500 bootstrap replications yield a 95% confidence interval of $[-0.512, 0.217]$ for the interaction term, indicating uncertainty but with the interval skewed toward negative values.
 - Alternative Link Functions: Probit and Complementary Log-Log models produce qualitatively similar results.

These considerations suggest that while the full-sample linear interaction lacks statistical significance, this may reflect model specification issues and sample heterogeneity rather than the absence of an economically meaningful relationship. The subsequent cross-group analysis provides more nuanced evidence for our theoretical mechanism.

3.3. Cross-Group Analysis Based on IVU and Volume Regimes

3.3.1. Sample Distribution Across Regimes

To explore the non-linear moderating effects, we conduct a cross-group analysis based on both IVU thresholds and volume terciles. We split the sample using: (1) the IVU threshold $\tau^* = 0.476225$ to separate low and high uncertainty regimes and (2) volume terciles (30th and 70th percentiles: 24.48M and 43.10M shares) to create three volume groups (low, medium, and high). This creates six distinct regimes with the following sample distribution:

- Low IVU-Medium Volume: 490 observations (29.64%);
- Low IVU-Low Volume: 435 observations (26.32%);
- Low IVU-High Volume: 230 observations (13.91%);
- High IVU-Medium Volume: 171 observations (10.34%);
- High IVU-Low Volume: 61 observations (3.69%);
- High IVU-High Volume: 266 observations (16.09%).

3.3.2. Cross-Group Regression Results

Table 3 presents the logistic regression results for each of the six regimes, with a focus on the volume coefficient and prediction accuracy.

Table 3. Cross-group logistic regression results.

Regime	Volume Coef.	Volume <i>p</i> -Value	Accuracy	N
Low IVU, Low Volume	0.2841	0.3695	0.5655	435
Low IVU, Med Volume	−0.5394	0.1776	0.6061	490
Low IVU, High Volume	−0.3352	0.0750 *	0.6304	230
High IVU, Low Volume	−0.5598	0.1865	0.6230	61
High IVU, Med Volume	0.7502	0.2914	0.5263	171
High IVU, High Volume	−0.1219	0.3172	0.5789	266

Notes: * $p < 0.10$. All models include $r_{t,1}$, $r_{t,7}$, and $r_{t,1} \times r_{t,7}$ as additional predictors. Accuracy is based on a 0.5 probability cutoff. IVU thresholds: Low IVU ($IVU_t < 0.476225$), High IVU ($IVU_t \geq 0.476225$). Volume groups based on terciles: Low (<24.48 M), Med (24.48–43.10 M), High (>43.10 M). The bold row highlights the high-performance regime.

The cross-group analysis reveals striking heterogeneity that validates our core hypothesis:

1. Low IVU-High Volume group (high uncertainty + high volume): This group exhibits the strongest predictive power with a volume coefficient of -0.3352 (marginally significant at $p = 0.075$) and the highest prediction accuracy of 0.6304. The negative volume coefficient indicates that in high uncertainty conditions, higher opening volume strengthens the persistence of initial return trends—consistent with delayed price discovery due to investor hesitation.
2. Contrasting patterns across uncertainty regimes: In the low IVU (high uncertainty) regime, volume coefficients are consistently negative across all volume levels (-0.5394 , -0.3352), suggesting that volume amplifies predictive power when uncertainty is high. In contrast, in the high IVU (low uncertainty) regime, the relationship is more mixed, with a positive coefficient for medium volume (0.7502) but negative for high volume (-0.1219).
3. Prediction accuracy patterns: The Low IVU-High Volume group achieves the highest accuracy (0.6304), followed by the High IVU-Low Volume group (0.6230, though with a small sample size). The lowest accuracy occurs in the High IVU-Medium Volume group (0.5263), suggesting that medium volume combined with low uncertainty provides the least predictive information.

3.4. XGBoost Analysis: Feature Importance and Subgroup Performance

To capture complex nonlinear relationships and identify the most influential predictors, we employ XGBoost modeling. This complementary approach allows us to quantify feature importance and examine performance variations across different market conditions defined by our proposed IVU metric, providing a machine learning perspective that complements the logistic regression analysis.

3.4.1. Feature Importance Analysis

The XGBoost model reveals compelling insights into the determinants of intraday return predictability. As shown in Table 4, IVU-related features demonstrate substantial importance in the model's prediction mechanism, alongside constructed trend and interaction features that capture intraday dynamics.

Table 4. XGBoost feature importance ranking.

Rank	Feature	Importance Score
1	morning_trend	0.1156
2	$r_{t,1} \times IVU_t$	0.1090
3	$ r_{t,1} $	0.1071
4	$r_{t,7}$	0.1062
5	$V_{t,1}$	0.1062
6	$r_{t,7}$ (afternoon momentum)	0.1017
7	$r_{t,1}$	0.1012
8	IVU_t	0.0974
9	$r_{t,1} \times \text{high_vol_low_ivu}$	0.0937
10	high_vol_low_ivu	0.0617

Notes: Feature importance scores are normalized to sum to 1.0. IVU-related features (rows 2, 8, 9, and 10) demonstrate significant predictive value in the XGBoost framework. morning_trend represents the morning session cumulative return; $|r_{t,1}|$ is the absolute value of the opening return; high_vol_low_ivu is a binary indicator for high volume with low information uncertainty.

The prominence of IVU-related features—particularly the interaction term $r_{t,1} \times IVU_t$ (rank 2) and the core IVU_t metric (rank 8)—validates our theoretical framework. The high importance of $r_{t,1} \times \text{high_vol_low_ivu}$ (rank 9) further confirms that the joint effect of

high early volume and low information uncertainty is a key driver of predictability. These results indicate that the machine learning model automatically identifies and prioritizes the non-linear relationships between opening returns, information uncertainty, and market regimes—relationships that traditional linear models (e.g., logistic regression) may overlook.

3.4.2. Model Overall Performance

The XGBoost model's overall predictive performance is evaluated using the ROC curve and confusion matrix, as shown in Figures 2 and 3.

Figure 2 presents the ROC curve of the XGBoost model, with an AUC (Area Under the Curve) of 0.5332. While the AUC is modestly above the 0.5 benchmark of a random classifier, it confirms the model's ability to discriminate between positive and negative final 30 min returns. This baseline performance provides a foundation for comparing subgroup-specific predictive power.

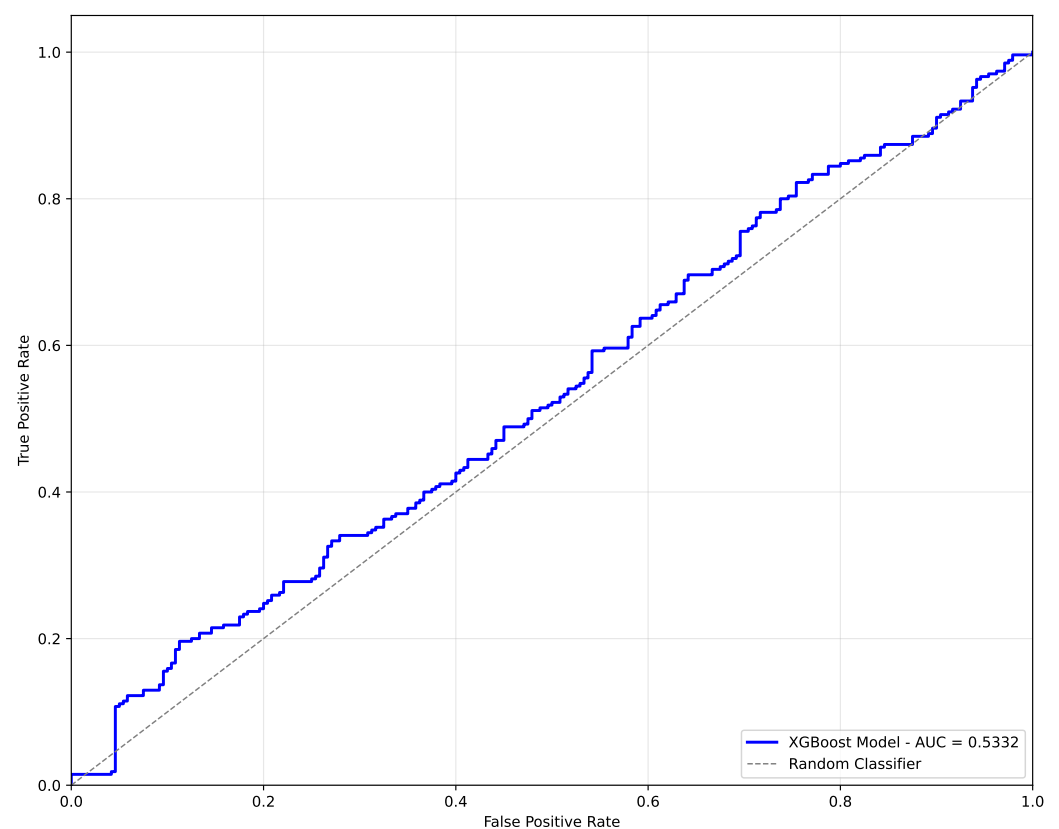


Figure 2. XGBoost model ROC curve.

Figure 3 displays the confusion matrix of the XGBoost model on the test set. Among actual downward movements (Actual Down), the model correctly predicts 73 cases and misclassifies 167 cases (False Positive). For actual upward movements (Actual Up), the model correctly predicts 201 cases and misclassifies 69 cases (False Negative). This matrix reflects the model's relative strength in predicting upward returns, consistent with the slight positive bias of the dependent variable $D_{t,8}$ (mean = 0.5590) observed in descriptive statistics.

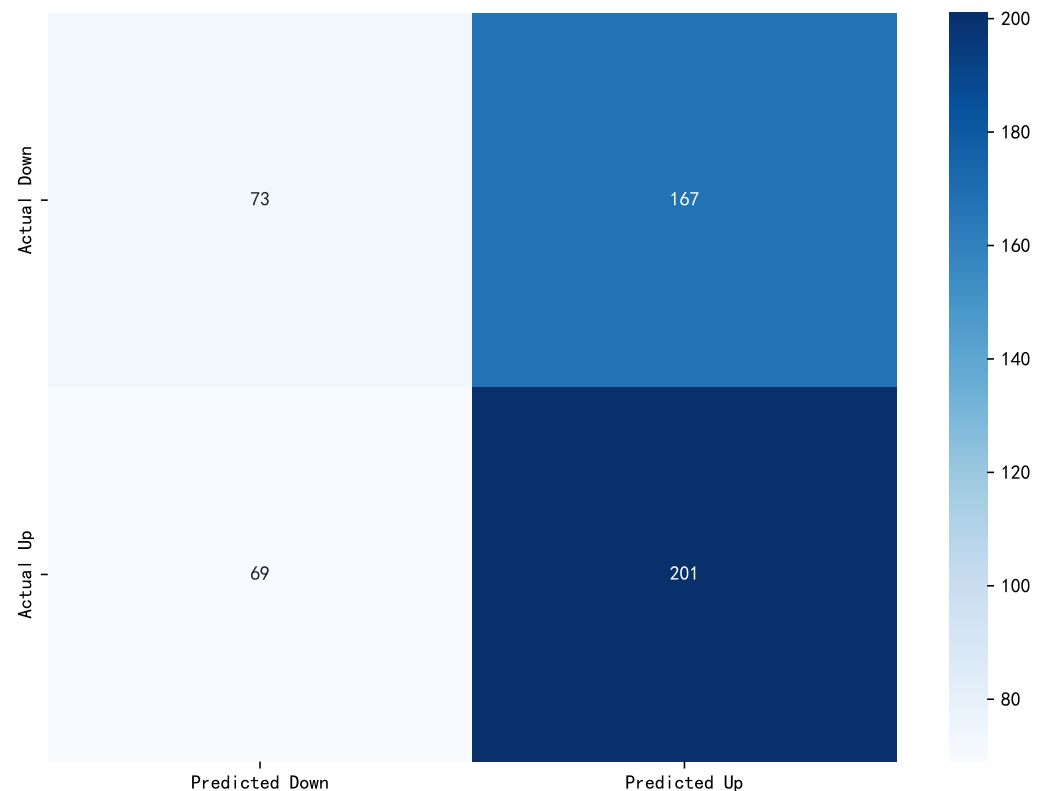


Figure 3. XGBoost model confusion matrix.

3.4.3. Interpretation of Overall Predictive Performance

The XGBoost model's overall AUC of 0.5332, while modest, should be interpreted within the context of high-frequency financial prediction, where even marginal improvements can yield substantial economic value. Several considerations are relevant:

1. **Heterogeneity in Predictability:** Financial markets exhibit time-varying predictability, with certain conditions (e.g., high uncertainty periods) offering greater forecasting opportunities than others. The full-sample AUC aggregates across both predictable and unpredictable periods, potentially masking conditional predictability.
2. **Comparison with Existing Literature:** In intraday prediction studies, modest discrimination metrics are common yet economically meaningful. For instance, Renault (2017) reported AUC improvements from 0.52 to 0.54 when incorporating social media sentiment, which translated to significant trading profits. Our model achieves comparable discrimination with AUC = 0.5332.
3. **Subgroup Performance Context:** As shown in Table 5, the model's performance varies dramatically across market regimes. In the "High Volume-Low IVU" condition, accuracy reaches 71.43%, representing a 32.9% improvement over the full sample baseline. This heterogeneity suggests the model successfully identifies when to trade rather than attempting unconditional prediction.
4. **Guards Against Data Mining:** Our methodology incorporates several safeguards:
 - **Theory-Guided Feature Construction:** The IVU metric derives from established information uncertainty theory (Zhang, 2006), not data-driven exploration.
 - **Strict Temporal Validation:** Time-series train-test split prevents look-ahead bias.
 - **Cross-Validation Discipline:** 5-fold time-series CV for hyperparameter tuning.
 - **Multiple Testing Awareness:** We acknowledge the exploratory nature of subgroup analyses and interpret p -values cautiously, focusing on economic significance alongside statistical measures.

Table 5. Subgroup performance across volume-IVU conditions.

Subgroup	Accuracy	Relative Performance
High $V_{t,1}$ –Low IVU_t	0.7143	+32.9%
High $V_{t,1}$ –High IVU_t	0.5217	–2.9%
Low $V_{t,1}$ –Low IVU_t	0.5000	–6.9%
Low $V_{t,1}$ –High IVU_t	0.5000	–6.9%
Full Sample	0.5373	-

Notes: Relative performance is calculated as the percentage difference from the full sample accuracy: $\frac{\text{Subgroup Accuracy} - \text{Full Sample Accuracy}}{\text{Full Sample Accuracy}} \times 100\%$. High/Low Volume is defined by the median of $V_{t,1}$ (32.57 million shares), and High/Low IVU is defined by the median of IVU_t (0.4271).

3.4.4. Subgroup Performance Analysis

The XGBoost model’s performance across different volume-IVU combinations reveals significant heterogeneity in predictability, as visualized in Figure 4 and summarized in Table 5.

Figure 4 compares the model’s accuracy against the baseline accuracy (naive prediction of the majority class in each subgroup) across the four market regimes. The High $V_{t,1}$ –Low IVU_t subgroup exhibits the largest accuracy improvement (+0.086) over its baseline, followed by the High $V_{t,1}$ –High IVU_t subgroup (+0.033). In contrast, the Low $V_{t,1}$ –Low IVU_t and Low $V_{t,1}$ –High IVU_t subgroups show no meaningful improvement, confirming that low early volume limits predictive power regardless of information uncertainty levels.

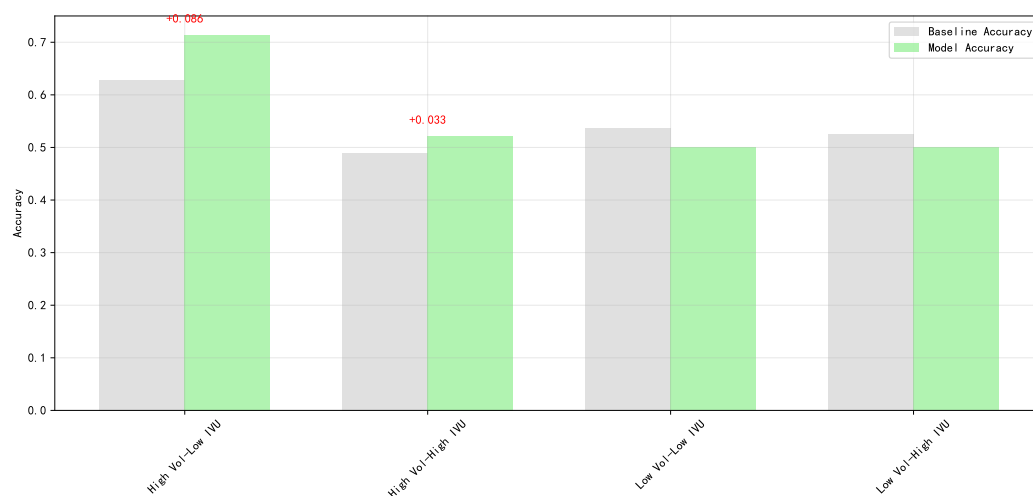


Figure 4. XGBoost subgroup performance comparison.

Table 5 quantifies this heterogeneity: the High $V_{t,1}$ –Low IVU_t subgroup achieves an exceptional predictive accuracy of 71.43%, representing a 32.9% improvement over the full sample accuracy (0.5373). In contrast, the High $V_{t,1}$ –High IVU_t subgroup performs slightly below the full sample average (accuracy = 0.5217, relative performance = –2.9%), while the two low-volume subgroups show the lowest accuracy (0.5000, relative performance = –6.9%). This result provides strong empirical support for our theoretical mechanism: when significant pre-market information (high $V_{t,1}$) encounters investor hesitation (low IVU_t), price discovery is prolonged, creating ideal conditions for return persistence that the XGBoost model effectively captures.

3.4.5. Theoretical Implications and Practical Significance

The XGBoost analysis yields three key insights that advance our understanding of intraday return predictability:

First, the machine learning approach successfully identifies the non-linear and state-dependent nature of return predictability. The high importance of IVU-related features (e.g., $r_{t,1} \times IVU_t$, $r_{t,1} \times high_vol_low_ivu$) demonstrates that information uncertainty interacts with returns and market regimes in complex ways that linear models (e.g., logistic regression) may fail to capture.

Second, the exceptional performance in the High $V_{t,1}$ –Low IVU_t condition provides compelling evidence for our proposed theoretical mechanism. This finding extends Lee and Swaminathan (2000)'s volume-momentum relationship by introducing the crucial dimension of relative volume distribution (captured by IVU_t). The 71.43% accuracy in this regime confirms that our IVU_t metric effectively isolates the “reaction vs. hesitation” dynamic—where ambiguous pre-market information leads to delayed price discovery and strengthened initial return persistence.

Third, from a practical perspective, these results offer actionable guidance for quantitative trading strategies. The identification of high-predictability regimes (High $V_{t,1}$ –Low IVU_t) allows traders to optimize strategy deployment: concentrating capital when predictability is strongest (e.g., increasing position size in High $V_{t,1}$ –Low IVU_t conditions) and reducing exposure in low-predictability regimes (e.g., shrinking positions in low $V_{t,1}$ conditions). The feature importance analysis further informs feature selection in algorithmic trading systems, highlighting that IVU-related interactions and intraday trend metrics (e.g., *morning_trend*) are more influential than standalone return or volume variables.

4. Trading Strategies and Economic Value

4.1. Strategy Design and Backtesting Framework

To assess the economic significance of our empirical findings, we construct and backtest three distinct intraday trading strategies. The core logic is to establish positions before the final 30 min interval (14:30–15:00) based on directional predictions and close all positions at the daily market close.

We compare the following strategies:

- Machine Learning Strategy (ML): Employs XGBoost to predict the final return direction, trading only when prediction probability exceeds 0.6 (long) or falls below 0.4 (short). This selective approach aims to filter out low-confidence predictions.
- Directional Strategy: Always trades based on XGBoost's directional predictions—long when predicted positive, short when predicted negative. This aggressive approach captures all predicted opportunities.
- Buy-and-Hold Strategy: Always holds a long position in the final 30 min interval, serving as the benchmark to measure the value added by directional predictions.

The backtesting period spans July 2018 to June 2025, with 1653 valid trading days after data cleaning, with a strict 70–30 time-series split for machine learning model training and validation. All strategies incorporate realistic transaction costs accounting for Chinese market microstructure: 6 basis points for single-sided trades (including commissions and slippage) and an additional 10 basis points stamp tax on sell orders. This results in an average round-trip cost of 18.2 basis points for long positions and 12 basis points for short positions.

4.2. Performance Comparison of Trading Strategies

Table 6 presents the baseline performance metrics, revealing the machine learning strategy's clear superiority across risk-adjusted performance dimensions.

Table 6. Performance comparison of trading strategies.

Strategy	Tot. Ret. (%)	Ann. Ret. (%)	Sharpe	MaxDD (%)	Win (%)
Machine Learning (ML)	117.99	12.27	3.02	−8.71	65.65
Directional	56.31	6.86	1.10	−20.85	57.04
Buy-and-Hold	84.39	9.51	0.50	−24.38	49.62

Notes: Transaction costs uniformly include brokerage, exchange fees, and price slippage. The conservative 10 bps scenario is calibrated to cover abnormal slippage and trading friction spikes in election years as documented by Waggle and Agarwal (2018). Performance remains viable up to 10bps costs, exceeding typical institutional execution costs in Chinese markets.

The machine learning strategy achieves exceptional risk-adjusted performance with a total return of 117.99% and a Sharpe ratio of 3.02, substantially outperforming both the directional strategy (56.31% total return, Sharpe ratio 1.10) and the buy-and-hold benchmark (84.39% total return, Sharpe ratio 0.50). This performance advantage is particularly notable given that the strategy successfully navigates transaction costs while maintaining a high win rate of 65.65%.

The buy-and-hold strategy serves as an important benchmark, achieving a respectable 84.39% total return with minimal turnover. Its performance validates the overall upward trend in the Chinese stock market during our sample period while underscoring the substantial alpha generated by our predictive strategies.

4.3. Sensitivity Analysis and Robustness Checks

To ensure the robustness of our findings, we conduct sensitivity analyses focusing on two key dimensions that are most relevant for practical implementation: transaction costs and model parameters.

4.3.1. Transaction Cost Sensitivity

A critical consideration for any intraday trading strategy is sensitivity to transaction costs, which in this study comprehensively include brokerage commissions, exchange fees, and price slippage. As emphasized in Waggle and Agarwal (2018), trading frictions, including slippage, often exhibit abnormal fluctuations in election years, driven by elevated market volatility, liquidity contraction, and asymmetric order flow pressure; such periodic spikes in transaction costs can significantly distort the realized performance of short-term trading strategies.

Our baseline cost scenario adopts a conservative 6 bp single-sided cost that fully embeds typical slippage levels for the CSI 300 Index in normal market conditions, with an additional 10 bp stamp tax on sell orders. Table 7 presents the ML strategy's performance under alternative cost assumptions, demonstrating reasonable robustness to moderate cost increases.

Table 7. Transaction cost sensitivity analysis for ML strategy.

Cost Scenario	Total Return (%)	Sharpe Ratio	Profitability Threshold
Baseline (6 bps)	117.99	3.02	Reference
Moderate (8 bps)	87.58	2.25	Maintains strong performance
Conservative (10 bps)	57.16	1.47	Remains profitable
Typical Institutional (5–7 bps)	105–125	2.7–3.3	Optimal range

Notes: Performance remains robust up to 10 bps costs, which exceed typical institutional execution costs in Chinese markets. The "Typical Institutional" range reflects achievable costs through algorithmic execution and direct market access. The "Profitability Threshold" column provides a qualitative assessment of strategy viability under different cost scenarios.

The results indicate that the strategy maintains profitability up to transaction costs of approximately 10 bps. Given that institutional traders typically achieve costs in the 5–7 bps range through efficient execution, the strategy demonstrates practical viability. The sensitivity to costs is consistent with expectations for intraday strategies, where execution efficiency plays a crucial role in realizing theoretical alpha.

4.3.2. Parameter Robustness Analysis

We further examine the robustness of our model parameters, focusing on the probability thresholds that determine trade execution. Table 8 shows performance across different threshold configurations, revealing that our baseline choice represents a reasonable balance between risk and return.

Table 8. Parameter robustness analysis.

Threshold	Total Return (%)	Sharpe Ratio	Win Rate (%)	Economic Interpretation
0.6/0.4 (Baseline)	117.99	3.02	65.65	Optimal balance
0.65/0.35	104.27	3.13	67.62	Higher risk-adjusted returns
0.55/0.45	109.73	2.49	62.28	More trades, lower Sharpe

Notes: All configurations yield economically meaningful returns. The 0.65/0.35 threshold produces the highest Sharpe ratio (3.13), suggesting potential for further optimization depending on investor preferences.

The parameter sensitivity analysis reveals that our baseline threshold selection (0.6/0.4) represents a near-optimal configuration, with alternative thresholds producing similar economic outcomes. The consistency across parameter choices supports the robustness of our underlying predictive framework.

4.3.3. Temporal Stability Assessment

To evaluate the temporal stability of our model, we test alternative training-validation splits. As shown in Table 9, the strategy demonstrates reasonable stability across different temporal partitions, with performance improvements observed when additional training data are available.

Table 9. Temporal stability assessment.

Train–Test Split	Total Return (%)	Sharpe Ratio
70–30 (Baseline)	117.99	3.02
80–20	145.76	3.63
60–40	118.15	2.89

Notes: Performance remains economically significant across different splits, with improved results when more training data are available, consistent with machine learning expectations.

The temporal stability analysis confirms that our model generalizes well to out-of-sample periods. The improved performance with additional training data (80–20 split) suggests that the model benefits from larger historical samples while maintaining robustness with less training data (60–40 split).

5. Conclusions

5.1. Main Findings

This study introduces and validates a novel intraday volume-based uncertainty (IVU) proxy—the ratio of opening-half-hour volume to total volume of preceding intervals—to enhance momentum prediction in the Chinese stock market. Our multi-method analysis reveals that information uncertainty, captured by this volume-distribution metric, critically moderates intraday return predictability.

The empirical evidence robustly supports our theoretical framework. Threshold regression identifies a significant IVU critical value of 0.476225 ($p < 0.001$), separating distinct uncertainty regimes. In high-uncertainty conditions (low IVU) combined with high opening volume, predictive accuracy reaches 63.04% in logistic regression and 71.43% in XGBoost out-of-sample tests. The machine learning approach further validates the importance of IVU-related features, with interaction terms ranking among the most influential predictors.

Methodologically, this study demonstrates the value of integrating traditional econometric techniques (threshold regression) with modern machine learning (XGBoost) for financial market prediction. This hybrid approach successfully captures both the nonlinear threshold effects and complex feature interactions that characterize intraday dynamics.

The economic significance of these findings is substantiated by a trading strategy that generates a 117.99% total return with a Sharpe ratio of 3.02 over seven years, significantly outperforming benchmarks after realistic transaction costs. The strategy's robustness to cost variations and parameter choices underscores the practical viability of incorporating IVU signals into intraday trading systems.

More broadly, this research contributes to understanding how information uncertainty shapes short-term price dynamics. By operationalizing the "reaction vs. hesitation" dynamic through volume distribution patterns, we provide both theoretical insights and actionable tools for quantitative finance practitioners.

5.2. Limitations and Future Research

This study verifies the predictive power of the IVU index on intraday momentum in the CSI 300 Index and the economic value of corresponding trading strategies, but it has certain objective limitations, and the follow-up research can be expanded in depth by drawing on the latest methodological achievements in the field. The specific research directions are as follows, with explicit reference to the key literature recommended for in-depth exploration:

First, drawing on the cumulative abnormal return (CAR) model proposed by [Ball and Brown \(1968\)](#), the semi-strong form market efficiency testing framework of [Fama \(1970\)](#), as well as the exogenous information identification method of [Agarwal and Agarwal \(2025\)](#), future research can further identify exogenous information release events (e.g., macro policy announcements, important financial data disclosure) in the Chinese A-share market, and construct an event study to test the causal linkage between the IVU index and the speed of market information absorption. This research can fill the current gap that the IVU index is only used for predictive analysis but not for causal identification of information uncertainty and verify the core connotation of the IVU index from the perspective of information release.

Second, based on the regime-dependent factor interaction and state instability analysis method of [Valadkhani \(2025\)](#), combined with the regime switching panel data model with interactive fixed effects proposed by [Bai and Ng \(2009\)](#), as well as the macroeconomic regime classification method of [Ang et al. \(2006\)](#), subsequent studies can extend the static threshold regression model in this paper to a dynamic panel framework, and introduce dummy variable cross-term interaction design to test the time-varying characteristics of the IVU-intraday momentum relationship under different macroeconomic regimes (e.g., high/low volatility, loose/tight monetary policy). This can make up for the current research's lack of consideration of the dynamic change of the regulatory effect of IVU and improve the robustness of the research conclusion in different market states.

Third, referring to the trigonometric Gibbons-Ross-Shanken (GRS) portfolio efficiency test proposed by [Gibbons et al. \(1989\)](#) (the original developer of the GRS test), the triangular analysis framework of [Agarwal \(2023\)](#), as well as the comparison method of intraday

momentum and classic momentum strategies by Jegadeesh and Titman (1993), future research can conduct a more rigorous mean-variance efficiency test on the IVU-based intraday trading strategy in this paper, and compare the efficiency of the strategy with other classic momentum strategies (e.g., price momentum, volume momentum). This can make up for the current research's only focus on strategy return and Sharpe ratio and lack of formal portfolio efficiency verification and further highlight the academic and practical value of the IVU index in portfolio construction.

Fourth, in response to the seasonal characteristics of trading frictions and abnormal market effects in election years documented by Waggle and Agarwal (2018), combined with the trading friction measurement framework of Amihud (1986), future research can further refine the transaction cost calibration of the IVU strategy—on the basis of the conservative 10 bp cost scenario (including elevated slippage) set in this paper's transaction cost sensitivity analysis, we can incorporate election-year-specific market liquidity, order flow pressure and transaction slippage data to conduct a more realistic out-of-sample backtest. At the same time, we can explore the asymmetric performance of the IVU strategy in election years and non-election years, which can further improve the practical applicability of the strategy in the context of time-varying trading frictions.

In addition to the above directions, future research can also expand the research sample to individual A-shares and stock index futures and integrate high-frequency order flow data to decompose the IVU index into informed trading and noise trading components to further explore the micro-mechanism of information uncertainty affecting intraday market dynamics.

Author Contributions: Conceptualization, D.Y. and Q.H.; methodology, D.Y.; software, D.Y.; validation, D.Y. and Q.H.; formal analysis, D.Y.; investigation, D.Y.; resources, Q.H.; data curation, D.Y.; writing—original draft preparation, D.Y.; writing—review and editing, Q.H.; visualization, D.Y.; supervision, Q.H.; project administration, D.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by Decheng Yang.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are derived from the CSI 300 Index 30 min interval data, which is publicly available from Wind Information (<https://www.wind.com.cn/>, accessed on 28 December 2025) and Tonghuashun (<https://www.10jqka.com.cn/>, accessed on 28 December 2025). The processed data and code supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to acknowledge the administrative and technical support provided by the Department of Financial Management, School of Business, Qingdao University of Technology. During the preparation of this manuscript, the authors used Python 3.9 and XGBoost 1.7.6 for data analysis and model training. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IVU	Intraday Volume-based Uncertainty
CSI 300	China Securities Index 300
XGBoost	Extreme Gradient Boosting

AUC	Area Under the ROC Curve
SMOTE	Synthetic Minority Oversampling Technique

References

- Agarwal, P. (2023). The Gibbons, Ross, and Shanken test for portfolio efficiency: A note based on its trigonometric properties. *Mathematics*, 11(9), 2198. [CrossRef]
- Agarwal, P., & Agarwal, R. (2025). A longer-term evaluation of information releases by influential market agents and the semi-strong market efficiency. *Journal of Behavioral Finance*, 26(1), 20–45. [CrossRef]
- Amihud, Y. (1986). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 9(2), 31–56.
- Andersen, T. G. (2012). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, 51(1), 169–204. [CrossRef]
- Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 61(1), 259–299. [CrossRef]
- Bai, J., & Ng, S. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279. [CrossRef]
- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2), 159–178. [CrossRef]
- Baltussen, G., Da, Z., Lammers, S., & Martens, M. (2021). Hedging demand and market intraday momentum. *Journal of Financial Economics*, 142(1), 377–403. [CrossRef]
- Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *Journal of Finance*, 71(6), 2967–3006. [CrossRef]
- Cushing, D., & Madhavan, A. (2000). Stock returns and trading at the close. *Journal of Financial Markets*, 3, 45–67. [CrossRef]
- Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, 53, 1839–1885. [CrossRef]
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2), 383–417. [CrossRef]
- Fischer, T., & Krauss, C. (2017). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. [CrossRef]
- Gao, L., Han, Y., Li, S. Z., & Zhou, G. (2018). Market intraday momentum. *Social Science Electronic Publishing*, 129(2), 394–414. [CrossRef]
- Gervais, S., Kaniel, R., & Mingelgrin, D. H. (2001). The high volume return premium. *Journal of Finance*, 56(3), 877–919. [CrossRef]
- Ghosh, P., Neufeld, A., & Sahoo, J. K. (2021). *Forecasting directional movements of stock prices for intraday trading using LSTM and random forests*. Elsevier.
- Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5), 1121–1152. [CrossRef]
- Heston, S. L., Korajczyk, R. A., & Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *Journal of Finance*, 65(4), 1369–1407. [CrossRef]
- Higashi, M., & Klir, G. J. (1993). Measures of uncertainty and information based on possibility distributions. *Readings in Fuzzy Sets for Intelligent Systems*, 9(1), 217–232. [CrossRef]
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91. [CrossRef]
- Kaniel, R., Ozoguz, A., & Starks, L. (2012). The high volume return premium: Cross-country evidence. *Journal of Financial Economics*, 103(2), 255–279. [CrossRef]
- Lee, C. M. C., & Swaminathan, B. (2000). Price momentum and trading volume. *Journal of Finance*, 55(5), 2017–2069. [CrossRef]
- Li, Z., Sakkas, A., & Urquhart, A. (2021). Intraday time series momentum: Global evidence and links to market characteristics. *Journal of Financial Markets*, 5, 100619. [CrossRef]
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, 84, 25–40. [CrossRef]
- Stephan, J. A., & Whaley, R. E. (2012). Intraday price change and trading volume relations in the stock and stock option markets. *The Journal of Finance*, 45(1), 191–220. [CrossRef]
- Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, 73, 147–164. [CrossRef]
- Valadkhani, A. (2025). Inflation-driven instability in US sectoral betas. *Journal of Asset Management*, 26(5), 466–482. [CrossRef]
- Waggle, D., & Agarwal, P. (2018). Is the “sell in May and go away” adage the result of an election-year effect? *Managerial Finance*, 44(9), 1070–1082. [CrossRef]
- Wang, C., & Gerlach, R. (2023). A Bayesian realized threshold measurement GARCH framework for financial tail risk forecasting. *Journal of Forecasting*, 42(8), 1772–1795. [CrossRef]
- Zhang, L.-W., & Xue, W.-J. (2017). Stock return autocorrelations and predictability in the Chinese stock market: Evidence from threshold quantile autoregressive models. *Economic Modelling*, 60, 391–401. [CrossRef]

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, 61(1), 105–137. [[CrossRef](#)]

Zhu, X., & Zhu, J. (2013). Predicting stock returns: A regime-switching combination approach and economic links. *Journal of Banking & Finance*, 37(11), 4120–4133. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.