# Bloomberg Capstone: Detecting Stock Anomalies Using Machine Learning

Shangqing Zhuang, Muyan Li, Xingyu Pan, Zebei Wang, Yu Zheng,

December 8, 2024

**Abstract**

The increasing accessibility of financial data sets and advancements in machine learning have opened new possibilities for detecting anomalies in stock markets. Our project explores two complementary approaches to identify stock return anomalies: a hybrid time-series LSTM-GARCH framework and a cross-sectional Isolation Forest model. The LSTM-GARCH model combines the predictive power of Long Short-Term Memory networks for time-series forecasting with the volatility estimation capabilities of GARCH, enabling the identification of anomalies in individual stock returns. The Isolation Forest model, on the other hand, employs an unsupervised learning algorithm to detect cross-sectional anomalies among stocks, leveraging features such as momentum, liquidity, and volatility indicators. Using data from the Russell 2000 index, our project evaluates the effectiveness of both methods in identifying significant deviations in stock behavior.

## 1 Introduction

In recent years, the increasing accessibility of vast financial datasets and the advancement of machine learning have opened new avenues for analyzing market anomalies, particularly in stock returns. Anomalies in stock returns can signal underlying risks, uncover trading opportunities, or even reflect broader market inefficiencies. Identifying these deviations is crucial for both investors and risk managers, as they can indicate hidden, impactful factors beyond standard financial metrics.

This paper investigates the effectiveness of two distinct machine learning models for detecting stock return anomalies. The first model, a hybrid LSTM-GARCH framework, captures temporal dependencies and volatility in individual stock time series [7]. The second model, an Isolation Forest algorithm, focuses on cross-sectional anomaly detection, utilizing features such as momentum indicators and market sentiment [8]. These approaches offer complementary perspectives—time-series analysis for individual stocks and cross-sectional analysis across multiple stocks—providing a comprehensive framework for identifying patterns in stock behavior under diverse market conditions.

An intuitive understanding of anomalies in time series data is the identification of points where the trajectory exhibits significant deviations from prior patterns. For stocks, this typically corresponds to days with sharp price increases or decreases. Following this definition of anomalies, we developed the LSTM-GARCH approach. This method focuses on a single stock's returns and identifies anomalies as returns that deviate significantly from what is considered reasonable. Using the stock's historical returns, the LSTM model estimates the center of the distribution, and the GARCH model estimates the standard deviation. A data point is classified as an anomaly if the actual return falls outside the estimated center $\pm$ a user-defined multiple of the estimated standard deviation. We found that a threshold of 2 to 2.5 standard deviations is appropriate. While this approach can effectively captures sudden price movements, it identifies anomalies driven by both macroeconomic and stock-specific factors, which may not suit cases where only stock-specific heterogeneity is of interest.

To address this limitation, we introduced a second approach: the Cross-Section Isolation Forest framework, which goes beyond stock returns to incorporate a broader set of features, enabling more comprehensive detection of anomalies driven by stock-specific heterogeneity. This unsupervised learning model identifies anomalies by detecting stocks that deviate significantly from their peers on a given day. Leveraging 22 features derived from Russell 2000 stock data—encompassing momentum, liquidity, and volatility—the model assigns anomaly scores, with lower scores indicating greater deviation. Stocks with anomaly scores a predetermined threshold are classified as anomalies. Our findings demonstrate that this approach effectively identifies stocks with higher variability in future returns and CAPM residuals compared to non-anomalous stocks, showcasing its capability to detect cross-sectional deviations that are not evident in individual stock analyses.

The structure of this paper is organized as follows: Section 2 provides relevant literature reviews covering LSTM-GARCH models, Isolation Forest in anomaly detection, and CAPM that will be covered in the analysis. In Section 3, we first introduce our data source and several variables that we are going to explore in the project. We also provide graphs to summarize the data and display the overall trends for the variable. Section 4 and Section 5 delve into data processing and discuss our methodology, presenting the key parameters and evaluation metrics chosen for each and covering the findings of our anomaly detection models with qualitative and quantitative analyzes of the detected anomalies for the LSTM-GARCH and Cross-Section Isolation Forest models, respectively. Both sections also provide specific observations on particular stocks and key dates to illustrate the models' real-world applicability. Section 6 introduces the design and implementation of an interactive surface to visualize anomalies for users. Finally, Section 7 summarizes the strengths and weaknesses of both models and suggests directions for future exploration in the detection of stock return anomalies.

# 2 Literature Review

## 2.1 Overview of LSTM and GARCH Models

In this method, we integrate LSTM and GARCH models to analyze and detect anomalies in financial returns. Specifically, the LSTM model is used to generate an expected return for each day based on historical data, while the GARCH model estimates a reasonable range of volatility for these returns. Any actual realized return falling outside this estimated range is classified as an anomaly.

In the following sections, we will conduct a literature review of both the LSTM and GARCH models to justify the rationale behind their application and the methods employed in this project.

### 2.1.1 LSTM for Stock Price Prediction

LSTM networks have emerged as a powerful tool for predicting stock returns due to their ability to capture long-term dependencies in time series data. Several studies have demonstrated the effectiveness of LSTM models in stock price forecasting.

Fischer and Krauss (2018) applied LSTM to predict returns of the S&P 500 index using closing price data. Their study utilized a binary matrix to handle potential biases in the data and adopted the "rmsprop" optimizer for efficient training, demonstrating the effectiveness of LSTM neural networks in capturing financial time series dynamics [7]. Gao et al. (2017) extended this framework by leveraging a richer set of market data, including opening price, closing price, highest price, lowest price, adjusted price, and transaction volume as input variables for the LSTM model. Their results, evaluated through various performance metrics, highlighted significant improvements over traditional prediction methods, emphasizing the advantage of incorporating diverse market information into LSTM models [8].

Han and FU (2023) propose an advanced LSTM model, Bi-Directional Long Short-Term Memory (Bi-LSTM) model, for stock price prediction, addressing the challenges of financial market volatility and sequential dependencies. Unlike traditional LSTM, Bi-LSTM processes data in both forward and backward directions, capturing comprehensive short- and long-term patterns [9].

These studies collectively demonstrate that LSTM, whether applied to closing price data, enriched market datasets, or enhanced with optimization techniques, is a reliable and effective tool for stock return prediction. Inspired by these findings, we employ LSTM to predict the average return for the next few days, establishing a central baseline for the first day's stock price. This predicted center serves as a foundation for identifying reasonable future stock values and detecting anomalies.

### 2.1.2 GARCH for Volatility Prediction

In the LSTM-GARCH combined approach, the standard deviation of returns is used as the benchmark for defining a reasonable volatility range. Estimating the future standard deviation of returns based on historical data requires selecting an appropriate model and determining the length of the historical dataset, often referred to as the lookback window, for one-step-ahead predictions. In this project, we adopted the GARCH(1,1) model and utilized a 252-day lookback window. The following sections provide the rationale behind these choices.

The GARCH model (Generalized Autoregressive Conditional Heteroskedasticity) is an extension of the ARCH model, originally introduced by Robert Engle. Unlike ARCH, which models the conditional variance of returns based solely on past squared returns, GARCH incorporates both past squared returns and past variances, making it more comprehensive. As a generalization of ARCH, GARCH provides greater flexibility in modeling time-varying volatility [3].

The key reason for using GARCH models over simple historical standard deviation is their ability to account for volatility clustering, a phenomenon where periods of high volatility are followed by high volatility, and low volatility by low volatility. This feature is particularly useful in financial modeling because markets often experience phases where volatility behaves non-constantly. Historical data, on the other hand, often assumes constant volatility, which can miss these dynamics [5]. In the context of anomaly detection, this is especially relevant: when market volatility is high, the range of acceptable stock fluctuations should widen accordingly, and only larger deviations should be considered anomalies. Therefore, considering the time-varying and clustering nature of volatility in our modeling is crucial, which is why we selected GARCH for this task.

In this approach, we employed the GARCH(1,1) model for all stocks without specifically determining the order for each individual stock. GARCH(1,1) is a widely used model for forecasting volatility. Hansen and Lunde (2005) demonstrated that no other advanced models provide significantly better forecasting performance than GARCH(1,1) [10]. Similarly, Javed and Mantalos (2013) highlighted that extensive investigations by researchers and academicians have consistently shown that the performance of GARCH(1,1) is satisfactory [11]. These findings, combined with the model's simplicity and intuitive interpretability, support its use in this study for estimating volatility efficiently across a large dataset.

Regarding the lookback window, we selected 252 days, equivalent to one trading year, based on both theoretical and empirical support from the literature. Bollerslev (1986) highlighted the importance of sufficient data length for reliable parameter estimation in GARCH models, with subsequent studies indicating that 250–300 data points are generally adequate for capturing key financial time series dynamics. This makes 252 trading days a practical choice, as it provides a balance between incorporating enough historical data for stability and maintaining sensitivity to recent market changes [3].

Additionally, Alexander (2008) emphasized the significance of aligning the lookback window with market cycles [2], given the annual patterns commonly observed in financial markets. Using a one-year window ensures that the analysis captures these cycles while remaining computationally manageable. Tsay (2010) further supported 252 days as a standard in financial modeling, making it an effective and widely adopted benchmark for volatility estimation [13].

## 2.2 Isolation Forest in Anomaly Detection

Isolation Forest (iForest), introduced by Liu et al. (2008), is a specialized algorithm for anomaly detection that fundamentally diverges from traditional methods by isolating anomalies through recursive partitioning rather than modeling normal behavior to identify deviations. The algorithm is based on the observation that anomalies, being rare and distinct, require fewer partitions to isolate compared to regular data points. This property forms the basis of its computational efficiency and scalability, particularly advantageous for large and high-dimensional datasets. By constructing an ensemble of isolation trees (iTrees) using random feature splits, iForest ensures robustness and adaptability across diverse datasets. Moreover, its independence from data distributional assumptions makes it highly effective in noisy and dynamic environments such as financial markets [12].

The foundational work by Liu et al. (2008) highlighted iForest's capability to isolate anomalies directly by leveraging their unique characteristics rather than relying on predefined notions of normalcy. This approach is especially relevant to financial datasets, where anomalies such as abrupt price changes or market shocks often signify significant deviations from typical patterns [12].

Subsequent research has further validated iForest's efficacy and relevance. Abubakar (2023) emphasized the algorithm's computational efficiency and robustness compared to density-based methods, noting its effectiveness in handling large-scale financial datasets. These features are critical for financial applications, where real-time anomaly detection can significantly influence decision-making [1]. Similarly, Togbe et al. (2020) demonstrated iForest's applicability to streaming data, highlighting its utility for continuous monitoring and anomaly detection in dynamic environments, particularly aligned with the needs of real-time stock market analysis.

Xu et al. (2023) provided additional insights into iForest's performance across diverse high-dimensional datasets, demonstrating its resilience to noise and adaptability to dynamic data streams. These attributes are especially pertinent to financial markets, where stock price data exhibit complex, multi-dimensional characteristics. Xu et al.'s findings reinforce iForest's suitability for identifying stock price anomalies within volatile and evolving market conditions [14].

Collectively, these studies establish a strong foundation for employing iForest in anomaly detection. Its scalability, computational efficiency, and robustness are particularly well-suited for financial contexts characterized by high-dimensional data and dynamic behavior. These insights support the application of iForest in detecting cross-sectional anomalies in stock markets, affirming its role as a critical tool for financial anomaly detection.

## 2.3 CAPM

The Capital Asset Pricing Model (CAPM) is a widely used framework for understanding the relationship between an asset's risk and its expected return. It assumes that a stock's return can be explained by its sensitivity to overall market movements (systematic risk) and factors specific to the stock (idiosyncratic risk). To operationalize this, researchers regress individual stock returns against a market index, isolating the portion of returns attributable to market dynamics.

This regression process is crucial as it allows for the decomposition of a stock's behavior into market-driven and stock-specific components. The resulting residuals, or error terms, represent the portion of a stock's return unexplained by the market. These residuals often highlight idiosyncratic factors such as unique risks, inefficiencies, or mispricing. [6] underscores that these residuals frequently correlate with anomalies, including deviations caused by non-normal distributions of returns, which CAPM inherently overlooks. As such, they provide valuable insights into inefficiencies and structural characteristics of individual stocks that the CAPM framework cannot fully capture.

Further, [4] demonstrates that CAPM residuals often reflect dynamic market anomalies such as the size effect, momentum, or book-to-market effect. These anomalies suggest systematic deviations from CAPM's predicted relationships, indicating that residuals carry information about mispricings or unaccounted risks. By examining these residuals, researchers can uncover hidden patterns and refine the understanding of asset pricing, making CAPM error terms a critical tool for detecting anomalies and advancing financial analysis.

## 2.4 Gaps and Opportunities

Regarding the LSTM-GARCH combined method: While both LSTM and GARCH are well-studied individually, combining the two for anomaly detection represents our innovation. Previous research on anomaly detection based solely on individual stock returns has primarily relied on time-series statistics, typically using the historical mean as the center and the historical standard deviation as the benchmark for reasonable variations. In contrast, our approach leverages the more predictive capabilities of LSTM and GARCH, accounting for trends in returns and the volatility clustering phenomenon. This enables the identification of anomalies with greater accuracy and robustness.

For the Cross-Section Isolation Forest method: While Isolation Forest is known for its efficiency and scalability in anomaly detection, it is traditionally applied to identify anomalies within a single stock's time series. However, our approach shifts this perspective by analyzing multiple stocks simultaneously on the same day. Instead of isolating unusual points within a stock's historical data, we detect which stocks stand out relative to their peers in the cross-section. By partitioning stocks using features like momentum, liquidity, and volatility, our method uncovers market-wide anomalies that traditional approaches may overlook. This provides a more comprehensive view of stock behavior, supporting daily stock screening and cross-sectional strategy development.

# 3 Data

The underlying data for this research are sourced from the Bloomberg Terminal. For the purpose of this study, we have selected historical stock data for the Russell 2000 index, focusing on different levels of granularity to meet the needs of each model. The Russell 2000 index comprises the smaller two-thirds of the Russell 3000 index, specifically excluding the largest 1000 companies by market capitalization.

This choice highlights the Russell 2000's suitability for anomaly detection research. Unlike large-cap stocks, which are more liquid and closely followed by analysts, smaller-cap stocks have lower liquidity, limited coverage, and higher idiosyncratic risks. These factors lead to greater inefficiencies and more unusual trading patterns, providing a richer dataset for anomaly detection. By focusing on the Russell 2000, this study aims to capture these inefficiencies and gain insights into less visible segments of the market.

Since the composition of the Russell 2000 index changes over time, we selected the constituent list as of November 1, 2024, for consistency and convenience in this study. The selected list includes 1963 stocks, representing the companies that were part of the index on that date. This static selection ensures that our analysis remains focused and avoids complications arising from changes in index membership over the research period. The dataset includes dates, daily adjusted closing prices, high, low, and volume for each stock over the period from January 1, 2014, to November 18, 2024. However, since the time frame is fixed, not all stocks included in the Russell 2000 index as of November 1, 2024, were listed before January 1, 2014. This results in missing data for some stocks during the earlier part of the period.

To illustrate this, the cumulative number of stocks included in the dataset over time is shown in Figure 1. Starting with 1157 stocks in 2014, the count steadily increases each year as more companies are added to the index, reaching the full 1963 constituents by 2024. This gradual growth reflects the dynamic nature of the Russell 2000, as new companies enter the market and older ones exit.

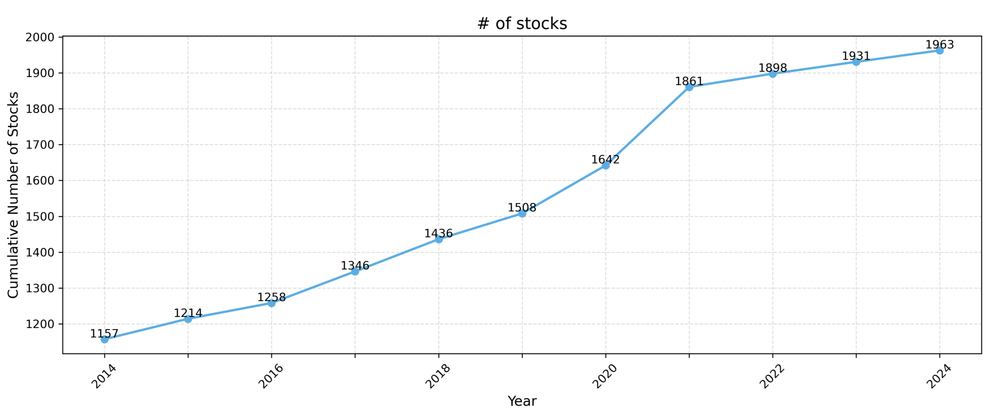



Figure 1: Cumulative Number of Stocks Present in Data by Year (Starting from 2014)

Figure 2 shows the adjusted closing prices and trading volumes on the first trading day of November each year, with extreme outliers excluded (bottom 98% for adjusted close and bottom 95% for volume). The median adjusted close price stays relatively stable, ranging from 17.15 to 24.94, while the spread widens in later years, reflecting increased price dispersion among stocks. Median trading volume shows greater fluctuation, rising from approximately $1.26 \times 10^5$ in 2014 to $3.39 \times 10^5$ in 2024. The larger range and higher peaks in trading volume suggest increased market activity and variability in stock liquidity over time.

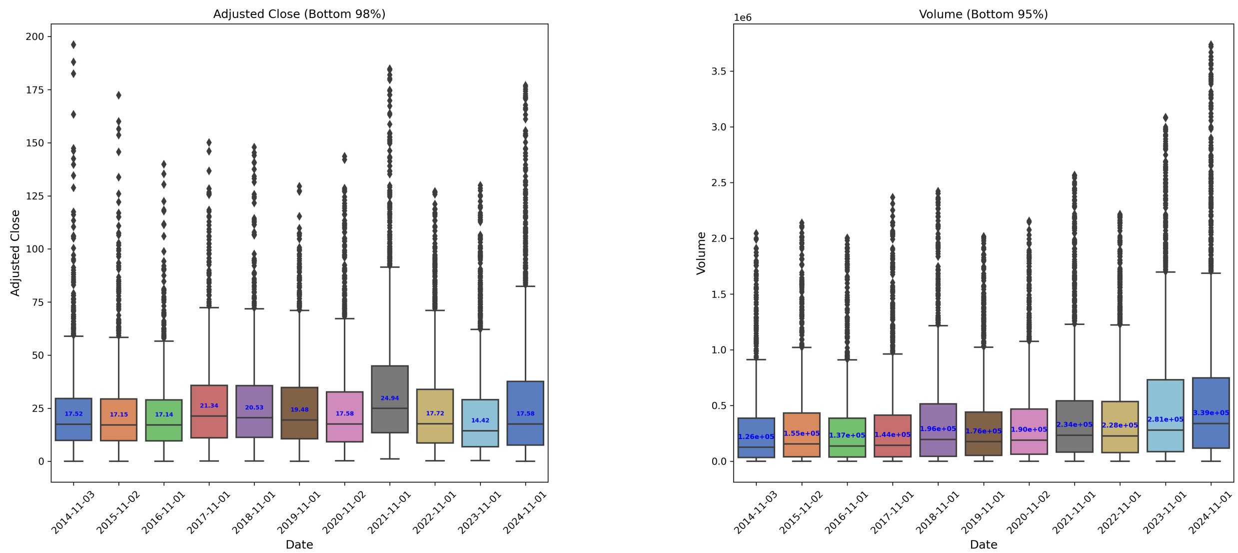



Figure 2: Distribution of Stock Adjusted Close/Volume (First Trading Day of November Each Year)

# 4 Model 1: Hybrid LSTM-GARCH

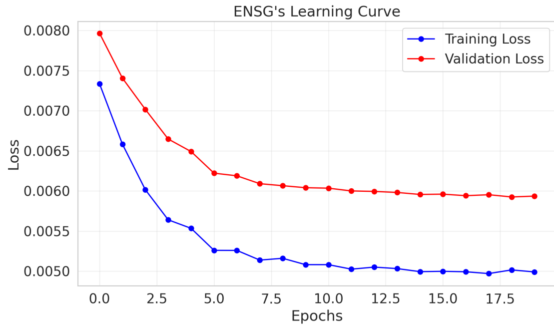
## 4.1 Methodology

### 4.1.1 LSTM for Return Prediction

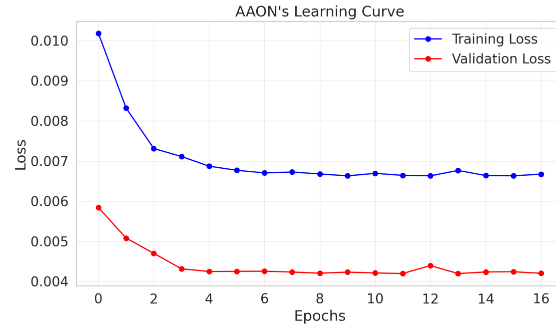
LSTM networks have proven to be an effective tool for forecasting stock returns, leveraging their ability to model long-term dependencies and complex temporal patterns in time series data. In this research, we aim to use LSTM to predict the return center for future periods, defined as the average of the future returns over a given period. Using the predicted return center, a GARCH model is applied to estimate a rational range of returns, incorporating volatility dynamics. Returns that fall outside this range are classified as anomalies, enabling robust detection of unusual market behaviors.

To forecast financial returns, we implemented a LSTM, leveraging its capability to capture temporal dependencies in sequential data. Initially, we computed daily returns from adjusted closing prices and derived the expected returns as a 5-day rolling mean, shifted by one period to align with future predictions. The dataset was normalized using the StandardScaler to ensure consistent feature scaling, then split into training and testing sets with a ratio of 80:20. A sliding window approach was employed to generate input sequences of 14 time steps for the LSTM model. The architecture consisted of an LSTM layer with 64 units to model temporal patterns, followed by a Dropout layer with a 10% rate to mitigate overfitting, and a Dense layer for single-value prediction. The model was trained with the Mean Absolute Error (MAE) loss function and the Adam optimizer, incorporating an EarlyStopping mechanism to halt training when validation loss ceased improving for three consecutive epochs.

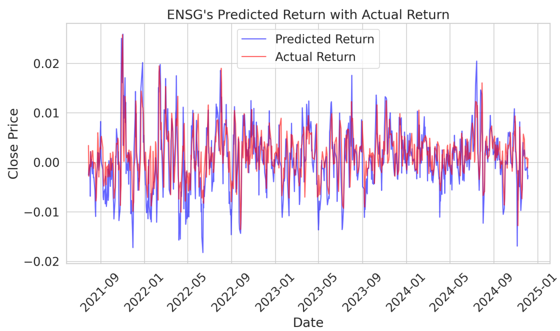
Implementing the approach above, the performance of learning curve presented in Figure 3a 3b shows the model's training and validation loss over epochs. For two equity, the training and validation loss decreases steadily with each epoch, indicating that the model is effectively learning from the training data and improving its predictions over time. Importantly, the gap between training and validation loss remains consistent showing that the model performs well on both training and unseen data without significant overfitting or underfitting.



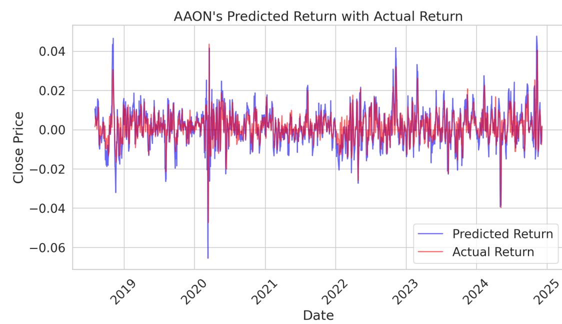

(a) ENSG's Learning Curve of LSTM Model




(b) AAON's Learning Curve of LSTM Model




(c) ENSG's Predicted Return with Actual Return




(d) AAON's Predicted Return with Actual Return

Figure 3: Learning curves and predicted vs. actual returns for ENSG and AAON.

By comparing the predicted and actual returns of two stocks in Figure 3c 3d, one with high return volatility and the other with low return volatility, we observed that the predicted average returns over the next five days closely align with the actual averages. Especially for AAON in Figure 3b, most of the time when the stock returns experience significant spikes, the predictions align closely with the actual returns. This demonstrates the strong predictive power of the LSTM model.

### 4.1.2 LSTM + GARCH

We implemented the LSTM model to predict the average return for the next 5 trading days based on the past 14 days of returns. For volatility, we employed a GARCH(1,1) model to make predictions. For each day's volatility prediction, we used the past year's (252 days) data to estimate the model coefficients. For each subsequent day, we generated predicted returns and predicted volatility, from which we established a benchmark range defined as the predicted return plus or minus X times (to be determined by users in the interactive interface) the standard deviation of the predicted volatility. If the actual returns fall outside this range, we classify them as anomalies detection.

## 4.2 Results

In this method, the user first defines a number of standard deviations as a threshold. If the actual stock returns fall outside the range of LSTM-predicted returns $\pm$ this number of standard deviations (as estimated by the GARCH model), the return is classified as an anomaly. Therefore, the choice of the number of standard deviations used as the threshold plays a significant role in anomaly detection. To gain insight into a reasonable range for this threshold, we first analyze and summarize the impact of different thresholds on the number of anomalies identified.

We predicted the LSTM-based returns and GARCH-based volatility for stocks in the Russell 2000 index from January 3, 2023, to November 8, 2024. Stocks with missing data were excluded from the analysis (since a GARCH model requires a lookback window of 252 days, missing data implies that the stock was not listed at the beginning of 2022). After filtering, a total of 1,847 stocks were included in the approach.

First, we analyzed the number of days each stock was identified as an anomaly during this period under different thresholds. The Table 1 below provides the statistical description of this analysis.

Table 1: Number of Days Identified as Anomalies Per Stock for Different Thresholds of Standard Deviations (2023.01-2024.11)

| Metric | **1.0** | **1.5** | **2.0** | **2.5** | **3.0** |
|---|---|---|---|---|---|
| Count | 1847 | 1847 | 1847 | 1847 | 1847 |
| Mean | 100.55 | 44.25 | 20.19 | 10.25 | 5.97 |
| Std | 36.86 | 17.54 | 8.58 | 4.64 | 3.05 |
| Min | 5 | 1 | 0 | 0 | 0 |
| 25% | 74 | 31.5 | 14 | 7 | 4 |
| 50% | 110 | 47 | 21 | 10 | 6 |
| 75% | 130 | 58 | 26 | 13 | 8 |
| Max | 164 | 85 | 45 | 26 | 16 |

During the selected period from 2023 to November 2024 (a total of 473 trading days), when the threshold is set to 1 standard deviation, the average number of days identified as anomalies per stock is approximately 101. This number significantly drops to around 44 days when the threshold is increased to 1.5 standard deviations. However, the ratio of 44/473 is still relatively high for anomalies.

When the threshold is further raised to 2 and 2.5 standard deviations, the number of days identified as anomalies per stock decreases to approximately 20 and 10 days, respectively, which is more reasonable. At a threshold of 3 standard deviations, the average number of anomaly days per stock drops to around 6 days, which is too few. This could result in smaller selected intervals that contain no anomalies at all.

Therefore, for users, a suitable range for the threshold would be between 2 and 2.5 standard deviations.

In addition, we plotted the time series of the number of stocks identified as anomalies each day under different thresholds, as shown in Figure 4 below. It can be observed that on specific days, significantly more stocks are identified as anomalies compared to others. In the selected period (January 2023 to November 2024), the date with the highest number of stocks identified as anomalies is November 6, 2024, the first trading day after the US presidential election results were announced. If the threshold is set to 2 standard deviations, 850 of 1,847 stocks were identified as anomalies on that day.

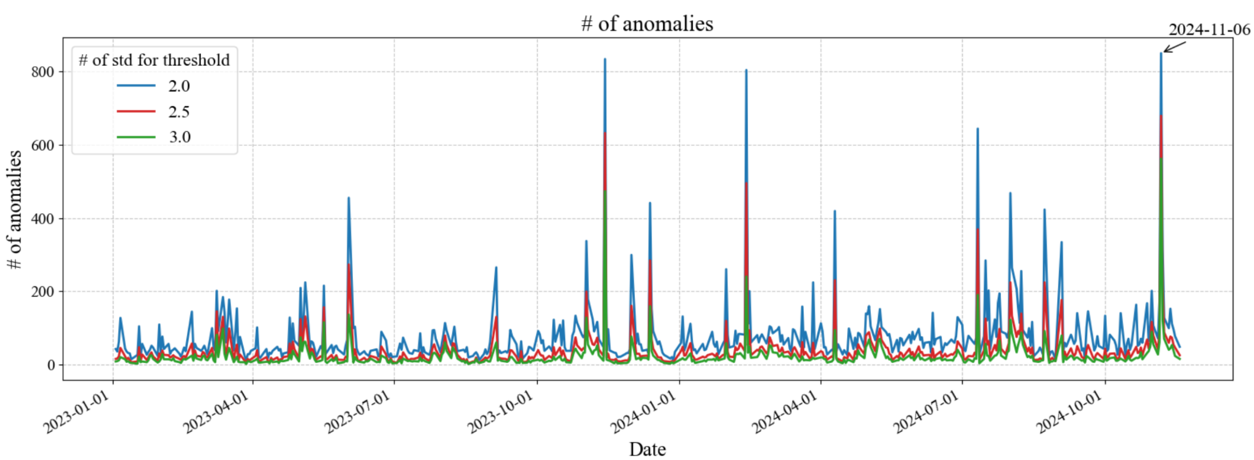



Figure 4: Time Series Plot of Anomaly Counts

The LSTM-GARCH model only considers the historical price/returns data of each stock and does not take macroeconomic factors into account. Therefore, during macroeconomic events that impact the stock market, many stocks can simultaneously be identified as anomalies. This method is well-suited for applications that involve reviewing the historical performance of individual stocks, as it can identify anomalies caused by both macroeconomic events and stock-specific events. However, when users are only interested in price movements caused by stock-specific heterogeneity, this method may not be ideal.

### 4.3 Observation

In the following, we take AAOI as an example to examine the characteristics of the anomalies identified by the LSTM-GARCH model.

First, when we map the anomalies onto returns in Figure 5, we observe that the anomalous returns consistently appear at turning points in the return series. This means that if the return on day t is identified as an anomaly, the absolute return on day t+1 is unlikely to exceed that of day t. In other words, we can predict that the next day's return will likely have a smaller magnitude. This reflects the characteristic of mean reversion in stock prices: a stock may experience a sharp rise or fall on a specific day due to special circumstances, but such large price movements typically do not persist. Additionally, the plot highlights the volatility clustering effect of the GARCH model: the volatility of returns around September 2023 is similar to that in August 2024. However, there are no anomalies in September 2023, whereas August 2024 has a concentration of anomalies. This is because the stock was more volatile in the summer of 2023 compared to June 2024, leading to GARCH-predicted volatility being higher. Consequently, returns needed to reach greater absolute values to be classified as anomalies.

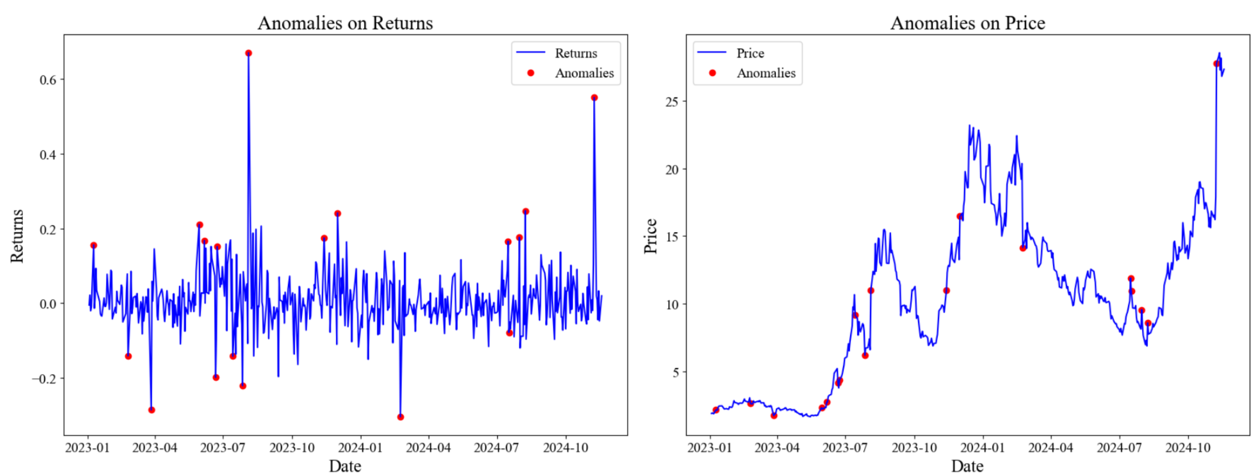



Figure 5: AAOI 's Time Series Returns with Anomalies

Next, when we examine the price chart in Figure 5, we see that only some anomalies correspond to turning points in price, unlike in the returns chart. This discrepancy arises because the model only considers return information. When a stock experiences a sharp increase on day t and is identified as an anomaly, the stock may continue to rise on day t+1 but with a smaller magnitude. In such cases, anomalies capture the turning points in returns but not necessarily in price.

## 5 Model 2: Cross-Section Isolation Forest

### 5.1 Methodology

Our methodology utilizes market data from Russell 2000 components over the period from January 1, 2014, to November 18, 2024, to construct features and train an Isolation Forest model for detecting daily anomalous stocks.

Initially, we derived 56 features from High, Low, Open, Close, and Volume data, then refined them through single-factor efficiency backtests and correlation evaluations, selecting 22 features for model training.

Using a 30-day lookback window of feature values from Russell 2000 components, the Isolation Forest model was trained to generate anomaly scores for each stock. The anomaly threshold is set at the 5th percentile of the anomaly scores, with a small adjustment of 0.01 subtracted to avoid fixing the number of detected anomalies. If a stock's predicted anomaly score falls below the threshold, its current status is classified as an anomaly. In general, there are around 3% to 4% of anomalous stock each day.

To evaluate the model's performance, we analyzed future stock performance using two metrics: the future price change rate and the CAPM residual error, which is the regression residual between individual stock prices and the Russell 2000 index price. By comparing the variance of these metrics between anomalous and non-anomalous stocks, we found that significantly higher variance in the anomalous group indicates the model's ability to effectively identify stocks with abnormal behavior.

### 5.1.1 Isolation Forest

Isolation Forest is an anomaly detection algorithm that leverages an ensemble of binary decision trees, known as Isolation Trees, to identify anomalies in a dataset. Unlike density-based or distance-based approaches, Isolation Forest operates on the principle that anomalies are "few and different." Since anomalies tend to be more isolated from the rest of the data, they can be separated with fewer splits when traversing decision trees. This makes the algorithm efficient and effective, particularly for high-dimensional datasets.

The core idea behind Isolation Forest is to recursively partition the data by randomly selecting features and thresholds to create branching decision nodes. During training, a random sub-sample of the dataset is assigned to each tree, and each node in the tree splits the data until every data point is isolated or a maximum tree depth is reached. Once the ensemble of Isolation Trees is constructed, the algorithm evaluates new data points by scoring their "isolation depth"—the average path length required to isolate a point across all trees. Anomalies, being more isolated, typically have shorter path lengths and higher anomaly scores.

### 5.1.2 Feature Engineering

**Feature Construction** We first constructed 56 features that can be categorized into three groups: Momentum features, including RSI, Stochastic Oscillator, and CCI, which capture price trends and oscillations. Market sentiment and liquidity features comprise the Accumulation/Distribution Index (Acc/Dist) and Ease of Movement (EOM), providing insights into buying or selling pressure and ease of price movement. Return and risk features include daily returns and volume-based return variations, highlighting stock-specific performance and short-term risk factors. Together, these features provide a comprehensive framework for detecting anomalies across diverse market conditions.

The **Relative Strength Index (RSI)** measures price momentum by comparing average gains to average losses over specific periods. RSI is calculated as:

$$\mathrm{RSI} = 100 - \left(\frac{100}{1+RS}\right),$$

where

$$RS = \frac{\text{Average Gain over } n \text{ periods}}{\text{Average Loss over } n \text{ periods}}.$$

We calculated RSI across different timeframes, including 5, 10, and 15 days, to capture both short-term and long-term price trends.

The **Stochastic Oscillator** evaluates the position of a price relative to its recent high-low range, providing insights into overbought or oversold conditions. The Stochastic Oscillator is defined as:

$$\%K = \frac{\text{Current Close} - \text{Lowest Low}}{\text{Highest High} - \text{Lowest Low}} \times 100,$$

where:

$$\text{Lowest Low} = \min(\text{Low over last } n \text{ periods}),$$

$$\text{Highest High} = \max(\text{High over last } n \text{ periods}).$$

We computed this feature over multiple periods (5, 10, and 15 days) to ensure comprehensive coverage.

The **Ease of Movement (EOM)** indicator integrates price and volume data to measure how easily a security's price moves. EOM is defined as:

$$\text{EOM} = \frac{\left(\frac{\text{High}+\text{Low}}{2}\right) - \left(\frac{\text{Prior High}+\text{Prior Low}}{2}\right)}{\text{Volume} \times \frac{(\text{High}-\text{Low})}{\text{Scale}}}.$$

The Scale factor is set to $10^6$ to normalize the volume, as commonly used in EoM calculations, to prevent large volume values from disproportionately affecting the indicator and to ensure that the EoM remains within a range of single or double digits. Positive values indicate upward price movements with lower resistance, while negative values suggest the opposite. Calculations were performed across 5, 10, and 20-day windows to capture movement dynamics over different timeframes.

The **Commodity Channel Index (CCI)** measures the deviation of the typical price from its rolling average. It is calculated as:

$$\text{CCI} = \frac{\text{Typical Price} - \text{SMA of Typical Price}}{0.015 \times \text{Mean Deviation}},$$

where:

$$\text{Typical Price} = \frac{\text{High} + \text{Low} + \text{Close}}{3}.$$

This feature is effective in identifying trend reversals and extended movements. Calculations across rolling windows of 5, 10, and 20 days ensured the model could identify both short-term and longer-term anomalies in price behavior.

**Daily Returns**, representing the percentage change in price from one day to the next, were included to reflect short-term performance. Daily returns are calculated as:

$$\text{Daily Return} = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}} \times 100.$$

This feature helps detect anomalies in immediate price movements, which may signify outlier events or shifts in market sentiment.

The **Accumulation/Distribution Index (Acc/Dist)** evaluates money flow trends by combining price and volume data. It is calculated as:

$$\text{A/D} = \text{A/D}_{\text{prev}} + \left(\frac{\text{Close} - \text{Low} - (\text{High} - \text{Close})}{\text{High} - \text{Low}}\right) \times \text{Volume}$$

Positive trends indicate accumulation, while negative trends suggest distribution.

Finally, **Volatility Returns**, which measure the degree of variation in daily returns over a rolling window, were used as a proxy for market risk. Volatility is defined as:

$$\text{Volatility} = \sqrt{\frac{\sum_{i=1}^{n}(\text{Return}_i - \text{Mean Return})^2}{n}}.$$

Higher volatility often indicates greater uncertainty or opportunity, making this feature crucial for detecting periods of heightened risk or turbulence.

We further used **Linear Regression** To enhance the usefulness of these features for anomaly detection, we applied regression analysis to several of them, including RSI, Stochastic Oscillator, EOM, CCI, Acc/Dist, and Volatility Returns. The regression metrics included:

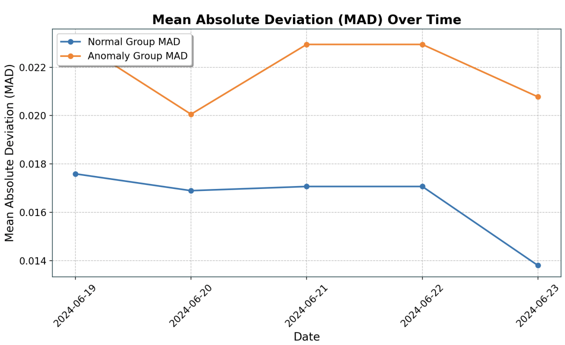
- **Slope:** Captures the rate of change, highlighting the direction and intensity of trends.
- **R-value:** Quantifies the strength of the trend, ensuring that observed changes are non-random.
- **P-value:** Assesses the statistical significance of the trends, filtering out noise.

In total, we built 56 features across three categories: Momentum features (RSI, Stochastic Oscillator, CCI) to capture price trends, Market sentiment and liquidity features (Acc/Dist, EOM) to reflect trading pressure and movement ease, and Return and risk features (daily returns, volume-based variations) to highlight performance deviations and risk. These features provide a robust framework for anomaly detection in diverse market conditions.

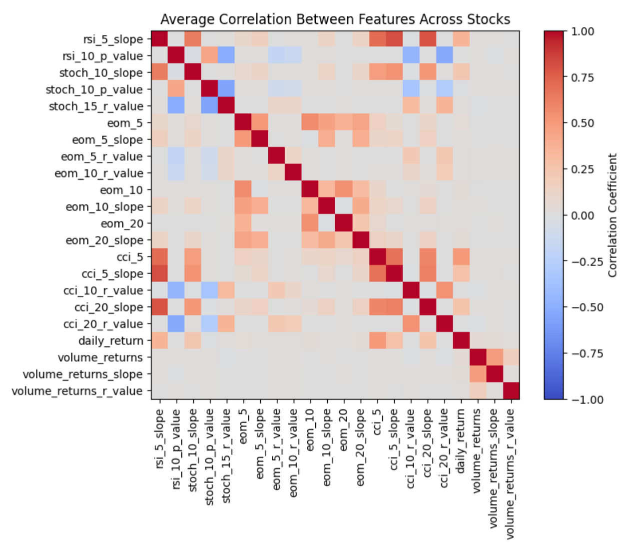
**Feature Selection** To optimize the performance of the anomaly detection model, we conducted a feature selection process to narrow the available features down to 22. This involved eliminating redundant and noisy features to ensure the model's robustness and interoperability.

To evaluate feature effectiveness, we tested each feature individually within the model and analyzed its anomaly detection results. We retained only those features that ensured the mean absolute deviation of the anomaly group consistently exceeded that of the normal group during a 5-day testing period(2024.06.19 - 2024.06.23). Figure 6a is an example of the feature eom_10_r_value which satisfies the above requirement. This process reduced the feature set to 28.

To further refine the feature selection, we computed the correlation matrix for the features within each stock and averaged them across all stocks to measure overall feature correlation in Figure 6b. As shown in the Figure 6b, most features are weakly correlated. After excluding features with absolute correlations greater than 0.8, we retained 22 features for anomaly detection.




(a) Mean Absolute Deviation over time




(b) Feature Correlation Metrics

Figure 6: Feature Selection in Cross-Section Isolation Forest Model

**Feature Transformation** We applied standardization to scale all features to a consistent range, ensuring uniformity and preventing scaling biases in feature representation. While the isolation forest algorithm is not sensitive to feature distances, standardization was necessary given the wide variation in feature ranges. For instance, the p-value is bounded between 0 and 1, while Ease of Movement (EOM) can range from -1,000 to 50,000. Without standardization, features like EOM could dominate the model's decision boundaries, skewing anomaly detection.

By transforming all features to have a mean of 0 and a standard deviation of 1, we ensured equal contribution to the model's decision-making process. This preprocessing step improved numerical stability and supported consistent interpretation, enhancing the overall robustness and reliability of the anomaly detection pipeline.

#### 5.1.3 Evaluation Approach

**Feature Importance in Anomaly Detection** To evaluate the contribution of individual features to anomaly detection, we utilized SHAP (SHapley Additive exPlanations). SHAP is a game-theoretic approach for interpreting machine learning models, offering insights into feature importance by attributing the model's predictions to individual features.

In unsupervised learning, where explicit labels are absent, SHAP provides a robust framework to quantify the marginal contribution of each feature to the model's output. By assigning Shapley values, SHAP ensures a fair and consistent calculation of each feature's impact on the discovery of patterns within the data, making it particularly valuable in the context of unsupervised learning.

For models like Isolation Forest, which detect anomalies by identifying patterns in the data rather than relying on labels, SHAP interprets anomaly scores instead of prediction labels. Anomaly scores measure how "anomalous" a data point is, and SHAP pinpoints the features that contribute most significantly to these scores. This makes SHAP an effective tool for measuring feature importance in unsupervised anomaly detection models.

The SHAP value for a feature $j$ is computed as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[ f(S \cup \{j\}) - f(S) \right],$$

where:

- $F$ is the set of all features,
- $S$ is a subset of features excluding $j$,
- $f(S)$ is the anomaly score predicted by the Isolation Forest when using the features in subset $S$,
- $f(S \cup \{j\})$ is the anomaly score when feature $j$ is added to subset $S$.

This formula ensures that SHAP values fairly allocate the contribution of the anomaly score among all features by evaluating their marginal impact across all possible feature subsets.

To determine overall feature importance, SHAP values are aggregated across all data points. For a feature $j$, the mean absolute SHAP value is calculated as:

$$\text{Feature Importance}_j = \frac{1}{N} \sum_{i=1}^{N} |\phi_j^{(i)}|,$$

where $\phi_j^{(i)}$ is the SHAP value for feature $j$ in the $i$-th data point, and $N$ is the total number of data points.

This aggregation ranks features based on their average contribution to the anomaly scores, highlighting those that are most influential in identifying anomalies. Features with high importance often reveal the key drivers of anomalous behavior in the dataset, providing actionable insights for further analysis and investigation.

By integrating SHAP with the Isolation Forest model, we bridge the gap between interpretability and unsupervised learning. This integration not only enhances the transparency of anomaly detection but also improves the practical utility of the system by identifying critical features that drive anomalous patterns.

**Deviation Measurement for Anomaly** To evaluate the effectiveness of the Isolation Forest model in identifying anomalies, we employed two measurements for deviations in future returns: residual errors and absolute future returns. These measurements provide a robust framework for understanding how well the detected anomalies correlate with significant deviations in actual market behavior.

The residual errors were derived from the Capital Asset Pricing Model (CAPM), which examines the relationship between stock returns in the Russell 2000 index and the market, represented by the Russell 2000 ETF (IWM). CAPM decomposes stock returns into systematic ($\beta R_m$) and unsystematic components ($\epsilon$), as follows:

$$R_i = \alpha + \beta R_m + \epsilon,$$

where $R_i$ is the stock return, $R_m$ is the market return, $\alpha$ is the stock-specific intercept, and $\epsilon$ captures idiosyncratic risk.

Using rolling windows of 240 and 30 days, we applied Ordinary Least Squares (OLS) regression to calculate the residuals ($\epsilon$) as:

$$\epsilon = R_i - (\alpha + \beta R_m).$$

Shorter windows capture dynamic, short-term stock behavior, while longer windows reveal more stable trends. Large residuals reflect stock-specific deviations driven by idiosyncratic events like earnings surprises or company-specific news, providing a critical measure of anomalies unrelated to market movements.

In addition to residual errors, we calculated absolute future returns to measure the magnitude of return deviations without considering directionality. Specifically, we evaluated the sum of absolute daily returns over the next five days as the horizon:

$$\text{Absolute Future Return} = \sum_{t=1}^{5} |R_{i,t}|,$$

where $R_{i,t}$ is the daily return of the stock on day $t$. This metric quantifies short-term deviations following anomalies, offering a direct measure of extreme price fluctuations.

**Deviation Metrics for Anomaly Detection** To quantify differences between the anomaly group (data points flagged as anomalies) and the normal group, we computed the following metrics:

- **Mean Absolute Deviation (MAD):** The average of absolute residual errors or absolute future returns within each group:

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}|,$$

  where $x_i$ is an individual residual error or absolute future return, and $\bar{x}$ is the group mean.

- **Standard Deviation (SD):** The variability of residual errors or absolute future returns within each group:

$$\text{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$$

By comparing the MAD and SD between the anomaly and normal groups, we assessed whether anomalies are associated with significantly higher deviations in future returns. Higher metrics in the anomaly group suggest a strong correlation between detected anomalies and large or unpredictable return deviations, validating the model's ability to identify meaningful market irregularities.

## 5.2 Results

In this method, we use a 30-day lookback window for the model by default, although users can adjust the number of days as desired in the application. The model analyzes 1,963 stocks to perform a cross-sectional analysis. Since we utilized the Russell 2000 index components as of November 18, 2024, some stocks were unavailable in earlier years. During testing, we detected anomalies in stocks from 2023-12-31 to 2024-10-25. At the start of the year, some stocks had missing data, and we excluded these stocks on a day-by-day basis. This adjustment resulted in a stock pool of approximately 1,920 stocks, which is close to the original stock pool size.

As shown in Figure 7, the MAD of the anomaly group remained higher than that of the normal group for 81.7% of time throughout the testing period, which is significantly higher than 50%. This consistent pattern underscores the model's effectiveness in identifying stocks with greater variability in their performance, which aligns with the characteristic behavior of anomalies in financial data.

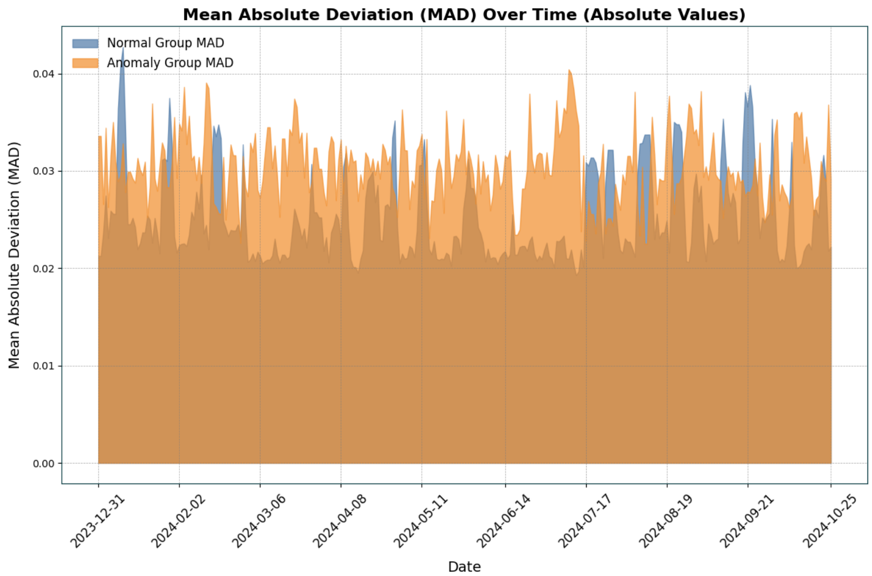



Figure 7: Mean Absolute Deviation of Normal and Anomaly Group

Additionally, the MAD of the anomaly group exhibits more pronounced fluctuations over time, reflecting the dynamic nature of anomalous stock behavior in response to varying market conditions. In contrast, the MAD of the normal group remains relatively stable, indicating that normal stocks exhibit more predictable and consistent performance. This divergence between the two groups reinforces the validity of the model's classifications and highlights its utility in isolating stocks that deviate significantly from typical market behavior.

The higher MAD values for anomalies also suggest a higher level of uncertainty or risk associated with these stocks, which could be crucial for risk management and decision-making processes in financial applications. This comparison of MAD values over time provides a comprehensive view of how the model differentiates between stable and unstable stock behaviors, adding further confidence to its application in anomaly detection.

## 5.3 Observation

In the following, we examine the anomalies identified by the Cross-Section Isolation Forest model using the same stock, AAOI, as analyzed in the LSTM-GARCH model.

When mapping the anomalies onto returns in Figure 8 during the period from October 2023 to November 2024, anomalies were observed only twice: on September 3, 2024, and September 16, 2024. Interestingly, these anomalies did not align with extreme return values on either day, indicating that this method does not focus solely on return extremes but captures more nuanced cross-sectional deviations. This suggests the model's ability to identify subtle irregularities that are not apparent in isolated return analysis.

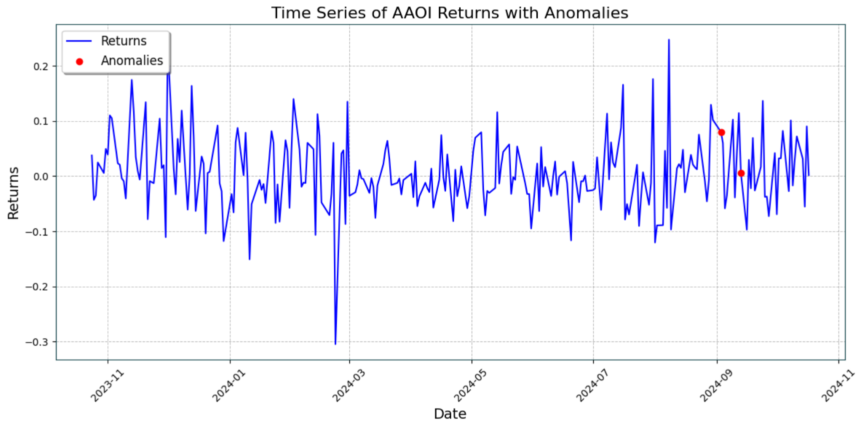
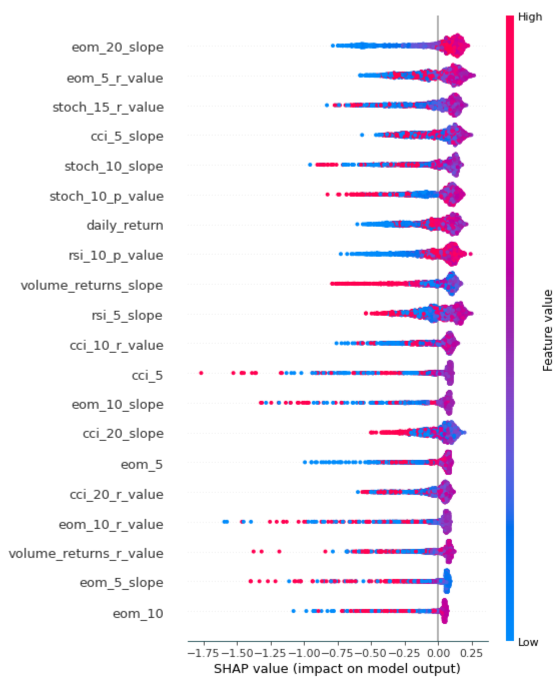
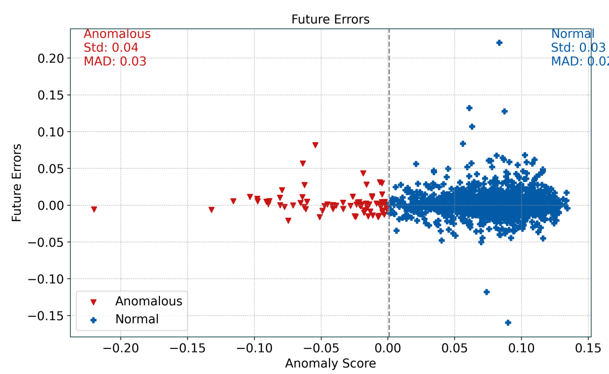



Figure 8: AAOI's Time Series Returns with Anomalies

To further understand the factors influencing the anomaly classification on September 3, 2024, we examined the SHAP summary plot in Figure 9a. The results show that the most impactful indicators contributing to the anomaly were eom_20_slope (Slope of Ease of Movement over a 20-day period) and eom_5_r_value (r-value of Ease of Movement correlation over 5 days). These features highlight the role of liquidity-related metrics in the anomaly detection process, suggesting that deviations in price movement relative to volume were significant on that day. Other indicators, such as momentum and volatility measures, also contributed but with less impact, demonstrating the model's ability to integrate multiple features for anomaly detection.




(a) Feature Importance and Contribution to Anomalies




(b) Future Errors Between Anomalous and Normal Stocks

Figure 9: Analysis of Anomalies Detected on September 3, 2024

Finally, we analyzed the future errors for September 3, 2024, as shown in Figure 9b, to evaluate the performance of the anomaly classification. The anomaly scores for anomalous points were distributed more widely compared to the tightly clustered scores of normal points. The standard deviation (SD) and median absolute deviation (MAD) for anomalous errors were 0.04 and 0.03, respectively, compared to 0.03 and 0.02 for normal errors. This indicates that normal stocks exhibit more stable performance, while anomalies are more volatile, aligning with the expectation that anomalies represent unusual market behavior.

Although the normal group includes a few extreme errors that are more dispersed than any point in the anomaly group, the majority of normal points remain tightly clustered around zero. In contrast, nearly all points in the anomaly group show consistent deviations away from zero, further highlighting the distinction between the two groups. This reinforces the model's ability to capture abnormal patterns that differ significantly from the stable behavior typically observed in normal stocks.

## 6 Interface Design and Functionality

We have created an interactive interface for two methods, developed using Dash, and here we will explain the operation of the demo. Figure 10 shows the interface for the LSTM-GARCH approach.

Users can input the stock ticker they want to analyze, specify the time range for anomaly detection, and customize the threshold for standard deviations used to define anomalies. By clicking the "Generate the Graph" button, the interface displays a graph of the stock price or returns (based on the selected tab) over the chosen time range, highlighting anomalies detected under the given threshold. Users can adjust the threshold and click the "Add New Markers" button to overlay anomalies identified with the new threshold onto the existing graph. This allows for a convenient comparison of anomalies detected under different thresholds.

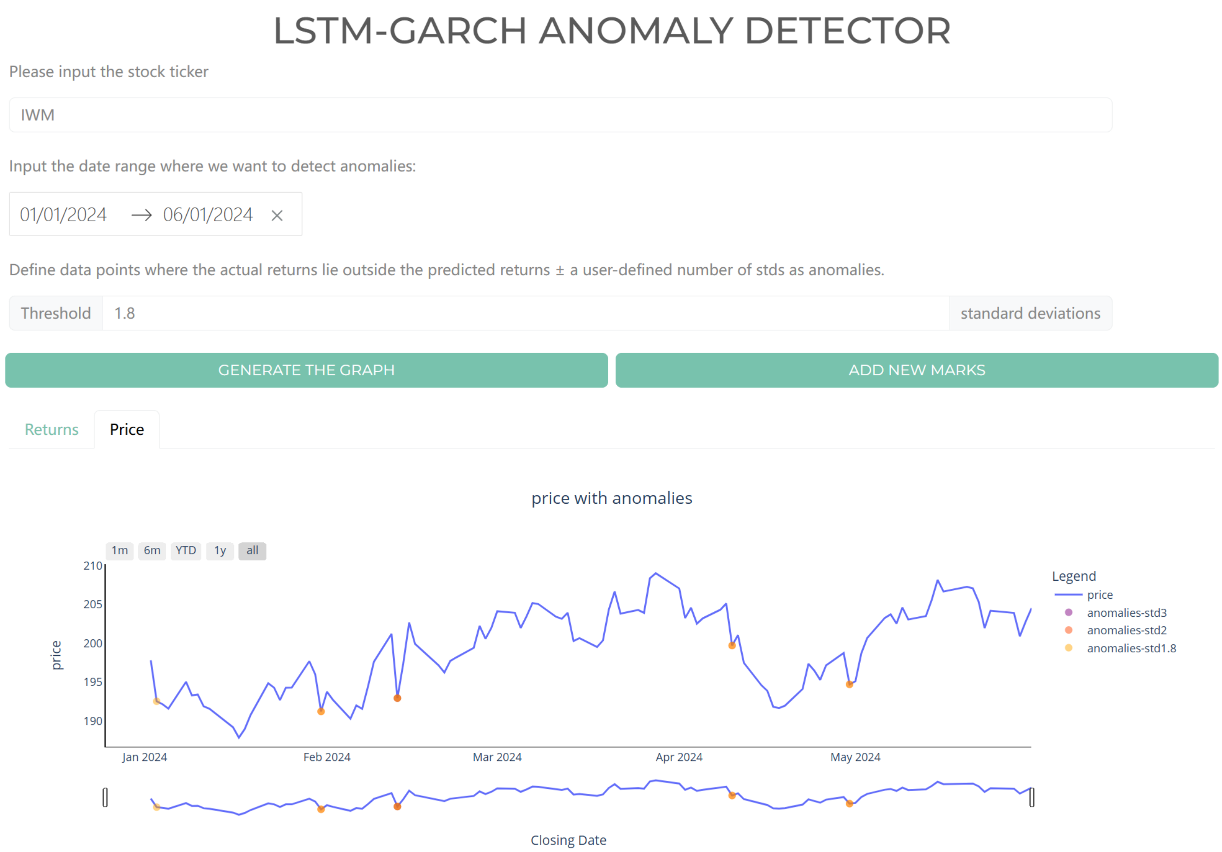



Figure 10: Interface for Model 1

Since the LSTM-GARCH model requires re-fitting the GARCH component daily for one-step predictions, real-time calculations can be slow for longer time ranges. To address this, we precomputed daily volatility and return predictions and integrated them into the backend of the interface. This way, instead of recalculating predictions during user interactions, the interface directly retrieves the precomputed results, ensuring faster response times. This setup provides users with a seamless and efficient experience for exploring anomaly detection dynamically.

Figure 11 demonstrates the interface for the second approach, Isolation Forest. This interface is designed for cross-sectional analysis, providing insights into anomaly detection for a specific day within the Russell 2000 stocks. Users can input the desired date in the "Target Date" field, specify the length of the training set, and click "Run!" to view the results. The interface outputs an anomaly score and label for each stock on the selected date, and users can download the data table in CSV format by clicking the "Download" button.

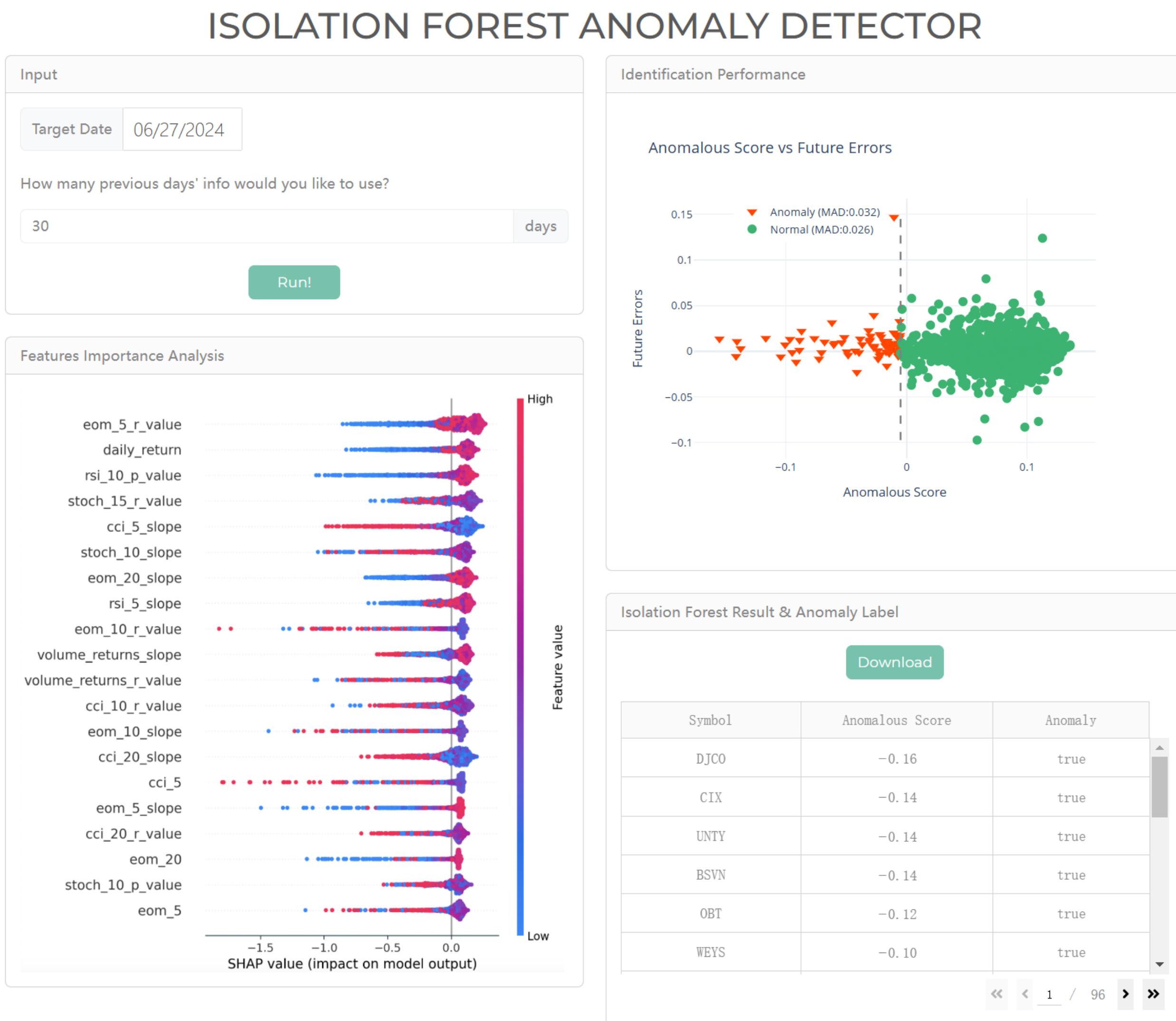



Figure 11: Interface for Model 2

In addition, the interface includes a feature importance analysis card, which highlights the key factors contributing to the identification of anomalies. The "Identification Performance" card further visualizes the financial predictive effectiveness of anomaly detection. It illustrates that stocks identified as anomalies tend to exhibit unusual behavior in subsequent days, reflected in larger absolute CAPM error terms. A scatter plot and the mean absolute deviation (MAD) of CAPM error terms are displayed to compare normal data with anomalies, providing a comprehensive overview of the detection results.

# 7 Conclusion

In our project, we implemented two distinct machine learning approaches—LSTM-GARCH and Isolation Forest—to detect stock return anomalies, using data from the Russell 2000 index. The LSTM-GARCH model combined the power of LSTM networks for time-series forecasting with the GARCH model for volatility estimation. This hybrid framework enabled us to identify anomalies in individual stock returns by comparing the actual returns to predicted values, considering both temporal dependencies and volatility clustering. The results showed that the LSTM-GARCH model was effective in detecting temporal anomalies, particularly those driven by sudden price movements or volatility changes that deviated significantly from the predicted behavior.

On the other hand, the Isolation Forest model, applied for cross-sectional anomaly detection, focused on identifying stocks that deviated significantly from their peers on a given day. By utilizing features such as momentum, liquidity, and volatility, the Isolation Forest model captured relative anomalies across multiple stocks in a manner that considered their interaction within the broader market context. This approach proved effective in detecting anomalies that were not easily identifiable using traditional uni-variate methods. To validate the detected anomalies, we employed the CAPM to analyze the residuals from stock returns, which quantify deviations unexplained by market movements. The results from the cross-sectional Isolation Forest analysis confirmed the model's effectiveness, showing that the anomalous group exhibited greater dispersion compared to the normal group, which highlights the model's potential as a robust tool for identifying anomaly points in individual stocks.

Future work could focus on integrating external macroeconomic factors, such as interest rates or geopolitical events, into both models to improve anomaly detection accuracy. Additionally, incorporating real-time data feeds could allow for more timely identification of anomalies, making the models more applicable for active trading or risk management applications. Further advancements in the Isolation Forest algorithm, such as the implementation of deep learning-based approaches, could help address limitations related to high-dimensional or non-linear data, leading to more effective anomaly detection in complex datasets.

# References


Abubakar, F. (2023). Anomaly detection in stock prices using isolation forest. *Journal of Financial Data Analysis*, *10*(2), 123–135.

Alexander, C. (2008). *Market risk analysis, quantitative methods in finance*. John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Chen, Z. (2002). Residual analysis in the intertemporal capm framework: Anomalies and dynamic factors. *Financial Economics Review*, *37*, 123–150.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative finance*, *1*(2), 223.

Desmoulins-Lebeault, F. (2002). Empirical tests of the capm: Distributional properties of residuals and risk factors. *Journal of Financial Studies*, *45*, 789–810.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, *270*(2), 654–669.

Gao, P., Zhang, R., & Yang, X. (2020). The application of stock index price prediction with neural network. *Mathematical and Computational Applications*, *25*(3), 53.

Han, C., & Fu, X. (2023). Challenge and opportunity: Deep learning-based stock price prediction by using bi-directional lstm model. *Frontiers in Business, Economics and Management*, *8*(2), 51–54.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873–889.

Javed, F., & Mantalos, P. (2013). Garch-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, *42*(8), 1917–1933.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422.

Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

Xu, H., Pang, G., Wang, Y., & Wang, Y. (2023). Deep isolation forest for anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*.