A Hybrid Framework of GARCH and Classification Models for Stock Market Volatility

KENTA NANAUMI s1290203 Supervised by Prof. Takahiro Tsuchiya

Abstract

In this study, I propose a hybrid approach that combines the generalized Autoregressive Conditional Heteroskedasticity (GARCH) model with machine learning classification methods to improve volatility forecasting in financial markets. Using daily Nikkei 225 data from 2020 to 2024, I integrated GARCH-based volatility forecasts with a machine learning classification technique, and I developed a trading strategy to determine market positions.

The results showed that the hybrid strategy outperformed both the traditional GARCH-only approach and the market benchmark under stable market conditions and was statistically significant. However, during periods of sudden changes in volatility or strong market uptrends, the effectiveness of this strategy declines, as evidenced by its poor performance in 2023-2024. This study reveals that the hybrid model is particularly effective in stable market environments where volatility exhibits autoregressive patterns, but requires further refinement to cope with abrupt changes in market volatility and strong directional trends. These results contribute to our understanding of how machine learning can enhance traditional financial models and underscore the importance of developing strategies in response to market conditions.

1 Introduction

In recent analyses of financial data, volatility is an essential indicator for quantifying the risk of asset price fluctuations. Many financial practices and theoretical studies, such as portfolio construction and risk management, rely on the accuracy of future volatility estimation. For this reason, various volatility models have been proposed. Among them, the GARCH model proposed by Bollerslev (1986) [1] is widely used as a standard tool to describe conditional variance in time series. The simple mechanism of updating the conditional variance at each point in time using past variances and error terms has been greatly developed in the field of econometrics, and various extensions of the GARCH model have been proposed (such as EGARCH, TGARCH, and FIGARCH [2]) to model complex phenomena observed in practice such as asymmetry, long-term memory, and jump components.

However, in stock price time series, sudden changes in volatility due to exogenous factors (policy announcements, geopolitical risks, pandemics, etc.) are observed, and these exogenous shocks are often not fully captured by the autoregressive fluctuation patterns assumed in the GARCH model. In addition, the GARCH model generally assumes a normal or Student-t distribution for the probability distribution, but actual financial data often has fat tails and excess kurtosis. The result is that the GARCH model fits very well in a particular period, but does not work at all in another period.

2 Objectives

This study aims to improve the prediction performance and address the traditional GARCH model problem by combining GARCH model-based forecasts with machine-learning classification methods. Specifically, I reframe the classification problem as "whether the actual volatility of the next day exceeds the model-based predicted volatility" and formulate trading strategies. The use of machine learning enables nonparametric, data-driven estimation that does not assume a certain probability distribution, making it easier to capture complex patterns such as nonlinearities that may be missed by conventional GARCH models.

In addition, by comparing the hybrid strategy that combines the GARCH model and machine learning with the traditional GARCH model-only strategy and market benchmark performance, respectively, by year and market environment, the conditions under which the hybrid strategy works effectively will be clarified.

3 Theoretical Frameworks

In the following, I detail the definition and parameter estimation method of the GARCH model, the mathematical basis of the classification models used, and the evaluation metrics used to evaluate the model.

3.1 GARCH model

3.1.1 Definition of GARCH model

The generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was proposed by Bollerslev [1]. This model assumes that the conditional variance depends on the past conditional variance and the past error term. Today, the GARCH models are widely used to capture dynamic changes in volatility in financial time series data. The GARCH(p, q) model is expressed by the following equations:

$$
\begin{aligned}
r_t &= \mu + \epsilon_t \\
\epsilon_t &= \sigma_t z_t
\end{aligned}
$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{1}$$

In equation 1, the error between the return r_t and the expected rate of return μ at a point in time is denoted by ϵ_t, and z_t is the independent and identically distributed (i.i.d.) noise, often assuming a standard normal distribution or t-distribution. ω is a constant term, reflecting the average variability of the entire market, and α_i is the coefficient of the error term, whose value varies with the magnitude of the past error ϵ_{t-i}, which indicates the impact of the past error ϵ_{t-i} on the current conditional variance σ_t^2. Additionally, β_i represents the effect of past conditional variance on the current conditional variance.

3.1.2 GARCH(1, 1) model

The GARCH model used in this study is the GARCH(1, 1) model in its most basic form. From equation 1, the GARCH(1, 1) model can be expressed as follows:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{2}$$

3.1.3 Parameter Estimation Method

The maximum likelihood estimation method is used to estimate the parameters of the GARCH model [6]. the log-likelihood function of the GARCH(1, 1) model is expressed by the following equation:

$$
\begin{aligned}
L = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N} \log(\omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) \\
-\frac{1}{2}\sum_{t=1}^{N} \frac{\epsilon_t^2}{\omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2}
\end{aligned}
\tag{3}
$$

The maximum likelihood estimator is obtained by maximizing this log-likelihood function.

3.1.4 Logarithmic Return

This study uses logarithmic returns. It defines as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{4}$$

where, P_t is the asset price at time t, and $P_{t-1}$is the price at the previous time step.

3.2 Classification models

In this study in order to consider volatility as a binary classification, i.e., whether or not it exceeds a certain threshold, the following three classification models are used:

3.2.1 Decision Tree

A decision tree is a tree-structured model that hierarchically classifies data by partitioning the feature space [5]. Each internal node has feature-based partitioning criteria, and leaf nodes predict class labels. Each partition maximizes the reduction of impurity.

3.2.2 Random Forest

Random Forests is a method of making predictions as an ensemble by constructing a large number of decision trees by a method of randomly sampling some data from the original data, called bagging [3]. Each tree is trained on a different subsample and subset of features. Each decision tree is constructed to minimize the impurity index independently, and the prediction is made by majority voting as a whole.

3.2.3 Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting) is an advanced gradient boosting algorithm that builds an ensemble of decision trees sequentially [4]. Each new tree is added to correct the errors of the previous models, enhancing predictive performance. XGBoost incorporates various optimizations for improved speed and accuracy.

3.3 Definition of classification model evaluation indicators

This section describes the metrics used to compare the performance of classification models.

3.3.1 AUC-ROC (Area Under the Curve Receiver Operating Characteristic Curve)

The ROC curve plots Recall against FPR for different threshold values. The AUC-ROC measures the area under the curve. It ranges between 0-1, with higher values indicating better performance. When the target is imbalanced data, the performance of the model may be overestimated.

3.3.2 AUC-PR (Area Under the Precision-Recall Curve)

The PR curve plots Precision against Recall for different threshold values. The AUC-PR measures the area under the PR curve and is particularly effective in evaluating models on imbalanced datasets.

3.4 Definition of Trading model evaluation indicator

This section describes the metrics used to compare the performance of the trading strategy.

3.4.1 Paired t-test

The paired t-test is a method of testing whether the difference between the means of two variables of interest is zero. The t-test produces the following two key statistics:

t-value: Represents the ratio of the observed difference between the means of the two variables to the variability of the data. Under the null hypothesis, it indicates how many standard deviations away from the hypothesized difference the observed difference is. A higher t-value indicates a greater difference between the two variables.

p-value: Represents the probability of observing a difference equal to or more extreme than the measured difference, assuming the null hypothesis is true. a sufficiently small p-value indicates that the observed difference is unlikely to have occurred by chance.

- **Two-tailed test:** Tests whether the means are different in either direction
- **One-tailed test:** Tests whether one mean is specifically greater than the other

3.4.2 Sharp Ratio

The Sharpe ratio represents profitability per unit of risk. In other words, the higher the Sharpe ratio, the more efficient the investment strategy is at reducing risk and the greater the expected return. It is defined as:

$$Sharpe\ ratio = \frac{R_p - R_f}{\sigma_p} \tag{5}$$

where R_p is the average return of the portfolio, R_f is the return on risk-free assets, and σ_p is the standard deviation of the portfolio return.

4 Dataset

In this study, daily historical data for the Nikkei 225 from 2013 to 2024 is used. This data will be used to compute daily logarithmic returns for training the GARCH model and feature values for training the classification model.

5 Methodology

This section describes the methodology employed to investigate the effectiveness of the hybrid model, in the following order: data preparation, machine learning model, transaction model, the evaluation method.

5.1 Volatility Measurement

The following methods are used to measure volatility.

5.1.1 Predicted volatility

The GARCH(1,1) model is used to forecast volatility. Specifically, the GARCH model is trained using data from the past five years to forecast the next day's volatility. Each time the date changes, the model is refitted to reflect the latest market information. The values obtained here are used as thresholds for the classification problem.

5.1.2 Actual volatility

For actual volatility, the absolute value of the logarithmic return is used.

5.2 Preprocessing

5.2.1 Data Splitting

5.2.1.1 GARCH model The GARCH model is fitted using the past five years of data, and the data period is updated by sliding the data period with each date change.

5.2.1.2 Classification models Split the dataset into training and test data as follows:

- **Training period**: 2 years of data for each training period (e.g. 2018-2019, 2019-2020, etc.).
- **Test period**: one year immediately following each training period (e.g., 2020, 2021, etc.).

5.2.2 Feature Values

The feature values are categorized into three groups:

1. **Trend Indicators**
   - **SMA (Simple Moving Average)**: Simple average of prices over any given period. Indicates the overall direction of the trend.
   - **EMA (Exponential Moving Average)**: Average calculated with more weight given to the recent price.
   - **MACD(Moving Average Convergence Divergence)**: Measures trend strength and direction.
2. **Volatility Indicators**
   - **Threshold**: Predicted volatility based on the GARCH Model.
   - **ATR (Average True Range)**: Reflects market volatility levels over a specific period.
   - **STD**: Standard deviation of prices for any given period.
3. **Event-Based Indicators**
   - **japan_event, usa_event**: Highlight significant geopolitical or policy-related events.

All features, except the Threshold, are shifted one day forward to prevent data leakage.

5.3 Hybrid Model

The hybrid strategy combines the predictions of the GARCH model with machine learning classification. This strategy is expected to improve the accuracy of transaction decisions by complementing the limitation of the conventional GARCH model, which has significantly lower prediction accuracy for data without an autoregressive structure, with estimation that does not rely on the establishment distribution using machine learning methods.

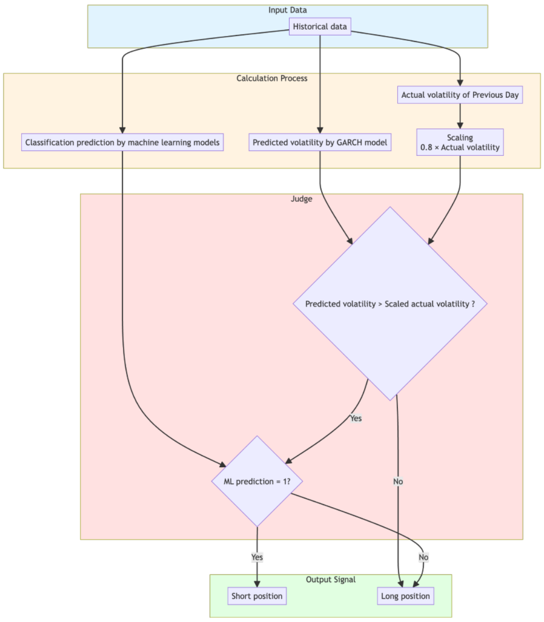

Figure 1: Hybrid strategy Logic flow

5.3.1 Logic

This section describes the logic flow of the hybrid strategy shown in Figure 1.

1. **Input data**
   In this model, historical market data is used as input. This data includes logarithmic return, actual volatility, and so on. This data is used for fitting the GARCH model and for calculating the classification model feature values.

2. **Calculation Process**
   Predicted volatility: Using the past 5-year logarithmic returns, the GARCH(1,1) model is refitted daily to predict the next day's volatility. This value is set as the "threshold" used in the following Judge and classification problem.

   Scaled previous day's actual volatility: Scales the previous day's actual volatility by a factor of 0.8. This scaled value is used in comparison to the predicted volatility.

   Classification Prediction
   Train machine learning models with feature values (in 5.2.2). The model classifies whether the next day's actual volatility exceeds the threshold predicted by the GARCH model.

3. **Judge**
   Predicted Volatility > Scaled Actual Volatility ?
   Determine whether the predicted volatility by the GARCH model exceeds the scaled previous day's actual volatility. If it is exceeded, proceed to the next decision; if not, the signal is decided to be long position.

   ML Prediction = 1 ?
   Determine whether the machine learning model predicts that the next day's volatility will exceed the threshold. If it predicts that it will be exceeded, the signal is determined to be a short position; otherwise, it is determined to be a long position.

The hybrid model can adapt to diverse market conditions because the GARCH model, which follows the autoregressive structure of the GARCH model and is powerful in stable market environments, and the classification model, which can respond to nonlinear relationships and external sexual shocks, complement each other. In addition, by taking the logical product of their outputs, the risk of trading on the wrong signal is reduced, as neither will take a short position unless both output the same signal.

5.3.2 Naked Strategy

To demonstrate the effectiveness of the hybrid model, a comparison is made with the naked strategy, in which decisions are made using only the predictions of the GARCH model and the previous day's measured volatility, without the intervention of machine learning.

5.4 Evaluation

The evaluation process consists of two complementary perspectives: the performance of machine learning models in classification tasks and the profitability of trading strategies derived from these predictions.

5.4.1 Machine Learning Models Evaluation

Yearly Evaluation: Compares the model's forecasting performance for each year and identifies the model's strengths and weaknesses for a given year.

5.4.2 Trading Model Evaluation

Market Condition-Based Evaluation: Analyze model performance based on the presence or absence of trends and market conditions. Market conditions are assessed using the slope of the regression line of the benchmark (Nikkei 225 buy-and-hold strategy). Whether the market was volatile during the year is determined by the kurtosis of returns.

Comparison of Trading Strategies: To validate the machine learning methodology, the following methods are conducted.

Statistical Test: A paired t-test is performed to compare daily returns between strategies:

- **Hybrid vs Naked Strategy:** Tests whether the hybrid strategy significantly outperforms the naked strategy

- **Hybrid vs Benchmark:** Tests whether the hybrid strategy significantly outperforms the market benchmark

For each comparison, both two-sided and one-sided tests are conducted, where:

- The null hypothesis assumes no difference in mean returns between strategies
- The alternative hypothesis for the one-sided test assumes the hybrid strategy yields higher returns

Here, the one-tailed test is to test whether the hybrid strategy is significantly superior, so if the hybrid strategy is inferior, a p-value of 1 is output.

Risk-Adjusted Performance: Evaluate the risk-adjusted return of the strategy using a comparison of Sharpe ratios.

6 Result and Discussion

6.1 GARCH Predicted Volatility vs Actual Volatility

The comparison between the predicted volatility from the GARCH model and the actual volatility reveals important insights into the model's performance over the study period from 2020 to 2024. Figure 2 illustrates the daily predicted volatility (blue line) and actual volatility (orange line), while Table 1 summarizes the yearly correlation coefficients between the two measures.

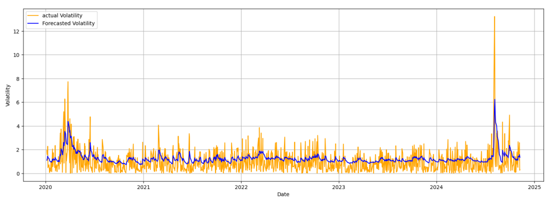

Figure 2: GARCH Volatility vs Actual Volatility

6.1.1 Correlation Analysis

Table 1 shows the yearly correlation coefficients between the predicted and actual volatilities. The correlation coefficients range from 0.50 to 0.66, indicating a moderate positive correlation throughout the years. Notably, the years 2020 and 2021 show relatively higher correlations (0.66 and 0.60, respectively).

6.2 Yearly Market Roughness

The roughness of the market each year is evaluated using the kurtosis of daily log returns. A higher kurtosis value indicates more extreme fluctuations in the market, which signal heightened volatility or heavy-tailed distributions. Table 2 summarizes the yearly kurtosis values.

Table 1: Yearly Correlation Coefficients between GARCH Predicted and Actual Volatility

Year	Correlation
2020	0.662533
2021	0.604375
2022	0.505344
2023	0.540804
2024	0.600220

Table 2: Yearly Market Roughness Based on Kurtosis

Year	Kurtosis
2020	7.497744
2021	1.083386
2022	0.701504
2023	-0.272263
2024	33.425157

As shown in table 2, 2024 exhibits extremely high kurtosis values of 33.43, indicating extreme market conditions likely driven by specific economic or geopolitical events. In addition, 2020 also showed a high kurtosis value (7.49), which aligns with the onset of the COVID-19 pandemic that caused significant disruptions in global financial markets.

In contrast, years 2021 to 2023 display low kurtosis values, indicating relatively stable market conditions with fewer extreme fluctuations.

6.3 Market Conditions

The market conditions each year are evaluated using the slope of the regression line derived from the benchmark (Nikkei 225 buy-and-hold strategy). A positive slope indicates an uptrend market, whereas a negative slope signifies a downtrend market. Table 3 summarizes the slopes of the yearly regression line, which provide insight into the general trends of the market.

Table 3: Yearly Market Conditions Based on Regression Line Slopes

Year	Regression Line Slope
2020	891.5549
2021	-26.3273
2022	60.4416
2023	1139.5079
2024	149.3319

The regression line slopes provide a quantitative measure of yearly market trends:

Downtrend Market: A year with negative slopes 2021, indicates bearish market trends characterized by declining prices. The negative slope in 2021 suggests downward movement in the market, potentially driven by external shocks or economic downturns.

Uptrend Market: Years with positive slopes, such as 2020 (891.5549) and 2023 (1139.5079), reflect an uptrend market with rising prices. The strong positive slope in 2020 aligns with a market rebound following the initial COVID-19 pandemic shock.

Stable Market: Years with small slopes, such as 2022 (60.4416) and 2024 (149.3319), suggest relatively stable market conditions with minimal overall price movement.

6.4 Classification model selection

As shown in the table 4, XGBoost consistently outperforms the other models in both AUC-PR and AUC-ROC across all years, demonstrating its robustness in handling volatility prediction tasks. In contrast, Random Forest and Decision Tree show relatively lower and less stable performance.

Table 4: AUC-PR and AUC-ROC comparison

Test year	Model	AUC-PR	AUC-ROC
2020	Random Forest	0.7359	0.5290
	Decision Tree	0.6504	0.5716
	XGBoost	**0.8489**	**0.7095**
2021	Random Forest	0.7714	0.6630
	Decision Tree	0.5818	0.5266
	XGBoost	**0.8878**	**0.8099**
2022	Random Forest	0.8174	0.6698
	Decision Tree	0.7183	0.6556
	XGBoost	**0.8854**	**0.8013**
2023	Random Forest	0.7681	0.6711
	Decision Tree	0.6638	0.6105
	XGBoost	**0.8477**	**0.7826**
2024	Random Forest	0.7119	0.5947
	Decision Tree	0.5865	0.5724
	XGBoost	**0.8103**	**0.7263**

Based on the table 4, XGBoost was selected as the primary classification model for this study. It is appropriate because it can effectively handle complex patterns and unbalanced data sets.

6.5 Hybrid strategy vs Naked strategy

Table 5 summarizes the yearly statistical results, providing insights into the significance of the hybrid model's outperformance or underperformance compared to the naked strategy. These results demonstrate the effectiveness of the hybrid strategy relative to the naked strategy in most years.

Table 5: Comparison of Hybrid strategy and naked strategy

Year	T-Value	Two side P-Value	One side P-Value
2020	15.81	4.08×10^{-39}	2.04×10^{-39}
2021	24.94	4.61×10^{-69}	2.30×10^{-69}
2022	11.24	6.86×10^{-24}	3.43×10^{-24}
2023	33.87	2.04×10^{-94}	1.02×10^{-94}
2024	-10.90	3.66×10^{-22}	1.0000

Table 6: Comparison of Hybrid strategy and naked strategy Sharpe ratio

Year	Sharpe (naked)	Sharpe (Model)
2020	0.23	1.81
2021	0.59	2.69
2022	0.91	1.41
2023	-1.83	1.12
2024	0.98	-0.09

6.5.1 Years of Strong Model Outperformance

The high t-values from 2020 to 2023 indicate that the performance of the hybrid strategy is significantly better than the naked strategy. For example, the t-value for 2023 is 33.87 and the p-value is sufficiently low to suggest that the performance of the hybrid strategy is statistically significant. In addition, the Sharpe ratios for each year indicate that the strategy with machine learning is more profitable with less risk than the strategy with the GARCH model alone (Tab. 6).

6.5.2 A year of Model Underperformance

In 2024, with a negative t-value and a one-sided p-value of 1, the performance of the hybrid strategy is significantly inferior to that of the naked strategy; during the sharp volatility change in August, the naked strategy saw a sharp increase in assets in response to that change, while the hybrid strategy saw a significant decrease in assets. This means that the GARCH model predictions were correct, but the output of the classification model was incorrect and the correct position was not taken. This result, along with the weakness of the strategy that simply takes the output as a logical product, showed that it is essential to improve the classification accuracy of the classification model.

6.6 Benchmark vs Model

This section evaluates the performance of the hybrid trading strategy compared to the benchmark (Nikkei 225 buy-and-hold strategy).

6.6.1 Years of Model Outperformance

From 2020 to 2022, the t-values were high, and the hybrid strategy significantly outperformed the benchmark,

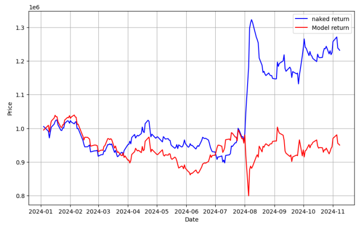

Figure 3: Performances of Both Strategies in 2024

Table 7: Comparison of Model-Based and Benchmark Trading Strategies

Year	T-Value	Two side P-Value	One side P-Value
2020	31.94	1.41×10^{-88}	7.03×10^{-89}
2021	21.20	2.78×10^{-57}	1.39×10^{-57}
2022	27.98	5.98×10^{-78}	2.99×10^{-78}
2023	-10.63	5.86×10^{-22}	1.0000
2024	-40.41	1.89×10^{-100}	1.0000

with a sufficiently low one-sided p-value indicating that the hybrid strategy's superiority was statistically significant. Comparing the Sharpe ratios in Tab. 8, the hybrid strategy outperformed the benchmark in all years, indicating that it is increasing assets with less risk. In 2021 and 2022, the slope of the regression line was relatively gentle, and the kurtosis was lower than in other years, suggesting that the market was relatively stable during these periods. On the other hand, while both the slope of the regression line and kurtosis were high in 2022, the volatility exhibited large fluctuations. However, as shown in Fig. 4, the pace of these fluctuations was moderate, and most of the period saw limited variation, indicating a relatively stable market. As a result, the GARCH model appeared to fit the market conditions better than in other years, leading to more favorable outcomes.

6.6.2 Years of Model Underperformance

In 2023, 2024, the t-value was negative and the one-sided p-value was sufficiently low to indicate that the

Figure 4: Comparison of Hybrid strategy with Benchmark in 2020

Table 8: Comparison of Hybrid strategy and Benchmark Sharpe ratio

Year	Sharpe (benchmark)	Sharpe (Model)
2020	0.77	1.81
2021	0.34	2.69
2022	-0.55	1.41
2023	1.71	1.12
2024	0.86	-0.09

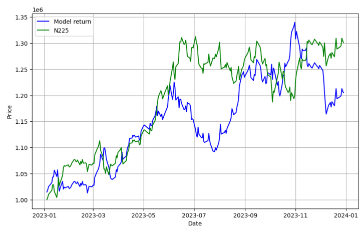

Figure 5: Comparison of Hybrid strategy with Benchmark in 2023

hybrid strategy was statistically significantly inferior.

2023 -Sharp upward trend-: As shown in Fig 5, 2023 was a market with a pronounced upward trend. since the GARCH model is based on the assumption that "current volatility fluctuates according to past autocorrelations," the structure of the GARCH model, which states that when a strong long-term trend occurs, the model "regresses to the mean," backfires, This can easily lead to lag with the actual rate of increase. In addition, since the market was relatively stable in 2021 and 2022, the classification model did not learn the continuous uptrend like 2023, and when the uptrend continued to some extent, it repeatedly outputs an incorrect judgment of "decline" based on past patterns, resulting in a decline in the accuracy of the model. Therefore, the performance of the hybrid model as a whole declined. It is thus essential to ensure the diversity of the data set.

2024 -Occurrence of a volatility spike-: Extremely high kurtosis was observed in 2024, as shown in table2, and a sharp volatility spike was observed within a short period. The classification model was trained on data showing relatively moderate fluctuations with no volatility spikes and could not adequately respond to the unknown spikes and drops. As noted in the comparison with the naked strategy(6.5.2), the GARCH model alone made correct predictions. However, because the hybrid model takes a logical product based on the forecasts of the two models, even if the GARCH forecast accurately indicates an increase in volatility if the classification model gives the wrong signal, the hybrid strategy as a whole will not be able to take advantage of the position. This indicates that simple model integration in the form of a logical product needs improvement.

The combination of the GARCH model and machine learning classification left superior performance in capturing market non-linearity and complex patterns compared to the parametric forecasting of the traditional GARCH model. The results show potential applications for volatility forecasting models.

7 Limitation and Future Work

The study revealed the following limitations of the hybrid model:

- Inadequate response to sudden changes in volatility.
- Inadequate response to strong upward trend markets.
- Decreased signal accuracy due to the simple integration of models.
- Not addressing Practical Constraints.
- Limited diversity in training data.

Add exogenous variables: Incorporating external factors such as geopolitical risk, monetary policy, and commodity prices into the classification model's features allows for early detection of volatility spikes.

Enhance data diversity: Expand the training data set to include more diverse market conditions, such as periods of sustained trends, periods of high volatility, and extreme events, to mitigate data imbalances. This will improve robustness.

Improve classification accuracy: Online learning methods that sequentially update models as new data comes in, and ensembles that use multiple models together to deal with rapid fluctuations in volatility. Improve model accuracy by adjusting appropriate hyperparameters.

Flexibility of model integration: Improve signal accuracy by implementing an integration method that goes beyond simple combining of GARCH and classification signals by considering not only logical products but also weighting and other rules.

Addressing Practical Constraints: Devise and simulate models that take into account transaction costs, slippage, liquidity, and other factors to apply to actual trading.

8 Conclusion

This study investigated the effectiveness of a hybrid strategy that combines the GARCH model with machine learning classification methods to forecast volatility in the Japanese stock market. This study aimed to complement the limitations of the traditional GARCH model by incorporating machine learning methods to capture complex market patterns and external shocks. The hybrid strategy showed superior performance compared to both the traditional GARCH model and market benchmarks under stable market conditions. The results indicate that the combination of GARCH forecasts and machine learning classification effectively captured market patterns during stable periods. On the other hand, the hybrid strategy showed limitations in volatility spikes and strong uptrends.

As a result, this study contributes to the field of financial forecasting by demonstrating that machine learning techniques can effectively complement traditional GARCH models. However, this study also reveals limitations of the hybrid strategy, especially in extreme market environments, and further refinements are needed to ensure robust performance in a variety of market environments.

Future research will focus on developing more adaptive models that can handle extreme market environments, and incorporating practical considerations such as transaction costs and market impact. In addition, exploring alternative models and methods of integration could address current limitations in responding to rapid market changes and strong trends.

References

[1] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of econometrics, vol. 31, no. 3, pp. 307–327, 1986.

[2] T. Bollerslev, "Glossary to arch (garch)," CREATES Research paper, vol. 49, 2008.

[3] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5–32, 2001.

[4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794, 2016.

[5] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, pp. 81–106, 1986.

[6] 湯浅辰丸, 鳥海不二夫, and 石井健一郎, "人工市場を用いたGARCH効果発生メカニズムの検証," 人工知能学会全国大会論文集, vol. JSAI2011, pp. 2H1OS181–2H1OS181, 2011.