Enhancing Financial Risk Forecasting: Machine Learning with and without Macroeconomic Variables Across Market Regimes

Candidate Number: 37279

2025-01-18

1 Introduction

Effective risk forecasting is essential for financial decision-making, helping institutions mitigate losses, optimize portfolios, and meet regulatory mandates. Traditional models like GARCH (Bollerslev, 1986) provide a solid foundation for volatility modeling but struggle with the complexities of dynamic financial markets due to their reliance on linearity and stationarity assumptions. This limitation has driven the adoption of advanced machine learning techniques such as Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), Gated Recurrent Units (GRU), and Gradient Boosting Machines (GBM) (Friedman, 2001), which excel in capturing non-linear relationships and high-dimensional data.

Building on existing research works and our previous study which demonstrates the effectiveness of Rnadom Forest and XGBoost in forecasting market volatility and assessing financial risk using historical data, this paper integrates the use of sophisticated models such as LSTM, GRU and GBM, and measures the impact of macroeconomic variables in financial risk forecasting. These models are particularly well-suited to handle temporal dependencies and long-term patterns, addressing the limitations identified in previous approaches. The incorporation of macroeconomic variables such as CPI, GDP and unemployment rate, alongside advanced feature engineering (e.g., log returns, rolling volatility, and RSI), allows for a more nuanced risk assessment across diverse market regimes(Bear, Bull and Crisis).

The study also employs threshold-based and Markov-switching models to dynamically classify market regimes, further refining predictive accuracy and risk evaluation. By bridging traditional methods like GARCH with modern machine learning approaches, as advocated in works such as Gu, Kelly, and Xiu (Gu et al., 2020), this research highlights how LSTM and similar models can enhance risk forecasting and provide actionable insights for portfolio management, trading strategies, and regulatory stress testing.

2 Data

The empirical analysis in this project relies on a dataset obtained from reputable sources. Stock-related data, including daily prices, returns, and trading volumes, were sourced from the Wharton Research Data Services (WRDS), a widely-used financial data repository. Macroeconomic variables such as the Consumer Price Index (CPI), unemployment rate, and Gross Domestic Product (GDP) were retrieved from the Federal Reserve Economic Data (FRED) database. These datasets span from January 1, 2000, to December 23, 2023, providing a robust basis for analysis.

2.1 Data

The analysis includes three stocks from distinct sectors: Apple (AAPL) in technology, Citigroup (C) in financial services, and Ford (F) in automotive. These selections ensure diverse sectoral coverage and enable an assessment of market dynamics across varying economic conditions. Following best practices from prior studies (Gu et al., 2020), the dataset underwent thorough preprocessing. Missing entries for prices, returns, and macroeconomic variables were removed, and duplicate records eliminated to maintain consistency. Logarithmic returns were computed to capture percentage price changes (McNeil et al., 2005).

Macroeconomic variables were normalized using a `StandardScaler` for comparability with stock-specific data. Feature engineering techniques, inspired by Bollerslev (Bollerslev, 1986) and Friedman (Friedman, 2001), were applied to create variables such as lagged returns, rolling volatility, moving averages, MACD, RSI, and bid-ask spreads. These features were designed to enhance the models' ability to capture complex market behaviors and trends.

2.2 Sample Characteristics and Descriptive Statistics

Table 1 summarizes the variables' distributions across the sample period, providing insights into central tendencies and variability critical for model development and performance evaluation.

Table 1: Descriptive Statistics of Key Variables

Metric	Count	Mean	Std. Dev.	Min	25th Percentile	Max
Price (PRC)	17,355	78.28	113.70	1.02	13.06	702.10
Trading Volume (VOL)	17,355	51.67M	110.12M	719,436	14.51M	1.89B
Returns (RET)	17,355	0.00064	0.0265	-0.3902	-0.0105	0.5782
Bid Price (BID)	17,355	78.26	113.69	1.01	13.06	702.11
Ask Price (ASK)	17,355	78.30	113.70	1.02	13.07	702.18
CPI (cpiret)	17,355	0.00636	0.00989	-0.0391	0.00066	0.0312
Unemployment Rate (UNRATE)	17,355	5.85	1.97	3.40	4.40	14.80
GDP	17,355	17,169.35	4,608.52	10,470.23	14,039.56	28,296.97

The summary statistics reveal notable patterns. For instance, the high standard deviation in trading volume indicates significant variability in market activity, while the large range in price highlights the diversity of assets included.

The CPI and unemployment rate exhibit relatively low variability, reflecting their macroeconomic nature and slower response to market fluctuations. Not only we are looking to capture high frequency market dynamics but also broader economic trends by applying machine learning techniques to these variables.

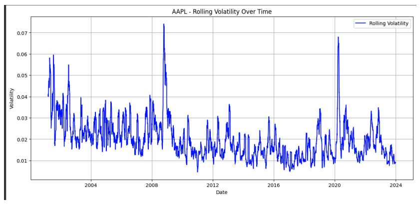
2.3 Key Events in the Time Series

The dataset captures significant events that have shaped financial markets. For example, during the 2007–2008 Global Financial Crisis, Citigroup experienced substantial price turbulence, reflecting the broader instability in financial services (Mandelbrot, 1963). In 2007, the launch of the iPhone marked a pivotal moment for Apple, as its stock exhibited sharp increases, transitioning the company into a leading technology innovator. Additionally, Ford's stock market demands have increased as a result of key technological advancement such as the development of its own electric vehicle in January 2022.

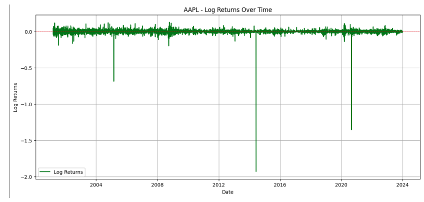
2.4 Data Visualization

These graphics provide insights into the fluctuations in Apple's stock, emphasizing the varying levels of market turbulence. Rolling volatility and returns are visualized to highlight periods of high and low market activity, as shown in Figure 1a and Figure 1b. Moving averages (20-day and 50-day) are plotted on Figure 1c alongside closing prices to identify trends and reversals, offering a clearer understanding of underlying market trends. Momentum indicators, such as the MACD and RSI, are also analyzed to evaluate overbought and oversold conditions in the market (Goodfellow et al., 2016). Figure 1d depicts the RSI for Apple over time, with thresholds for overbought and oversold levels marked to indicate critical trading conditions.

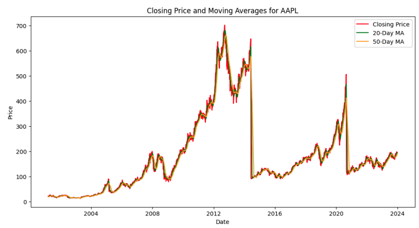
The dataset's diversity, both in sectoral representation and feature engineering, ensures a robust foundation for testing advanced forecasting models like LSTM, GRU, and GBM. By linking data characteristics to fundamental financial concepts such as returns, volatility, and market regimes, this study integrates empirical insights with theoretical underpinnings.



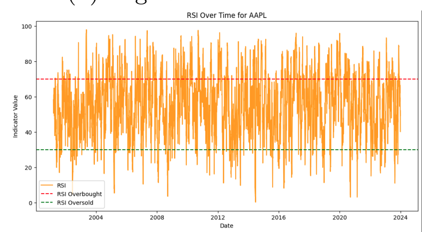
(a) Rolling Volatility Over Time



(b) Log Returns Over Time



(c) Moving Averages Over Time



(d) RSI Over Time for AAPL

Figure 1: Visualizations of AAPL's stock features.

3 Empirical Analysis

3.1 Methodology and Model Choice

The analysis begins with data preparation, including feature engineering (e.g., rolling volatility, log returns, MACD, and RSI) and normalization to enhance predictive accuracy. While traditional models like GARCH are foundational for volatility modeling, their reliance on linearity and stationarity assumptions limits their applicability in dynamic financial markets. To address these gaps, machine learning techniques that excel in capturing temporal dependencies and non-linear interactions are applied.

- **LSTM and GRU**: These recurrent neural networks model sequential dependencies effectively. LSTM is advantageous for long-term dependencies, while GRU provides a simpler and computationally efficient architecture (Hochreiter and Schmidhuber, 1997). Both models were implemented using the `Keras` library, optimized for regression tasks using the Adam optimizer and mean squared error (MSE) loss.

- **GBM**: As an ensemble learning method, GBM excels at modeling non-linear relationships in structured data (Friedman, 2001).
  The `GradientBoostingRegressor` from `scikit-learn` was employed with parameters such as a learning rate of 0.1 and a maximum depth of 3 to prevent overfitting.

3.2 Implementation and Parameterization

The computational environment utilized Python libraries, including `numpy`, `pandas`, and `matplotlib`, for data preprocessing, visualization, and analysis. Time-series splitting was implemented using `TimeSeriesSplit` to maintain temporal consistency between training and testing data (Pedregosa et al., 2011).

Key libraries used:

- `pandas` for data preprocessing.
- `scikit-learn` for GBM implementation and standardization.
- `Keras` for LSTM and GRU models.
- `matplotlib` and `seaborn` for visualizations.

3.3 Justification of Methodology

Applying different models with specific advantages allows us to compare their performances and decide the most suitable to study a specific type of stock. LSTM and GRU excel at modeling temporal dynamics, while GBM captures feature interactions effectively (Goodfellow et al., 2016; Friedman, 2001). Engineered features like rolling volatility and RSI further enhance the models' ability to detect financial patterns, ensuring robustness and adaptability (McNeil et al., 2005).To evaluate the performance of these models, metrics such as R^2 and Mean Squared Error (MSE) were calculated, providing a clear measure of their predictive accuracy and robustness. The impact of macroeconomic variables was assessed by initially running the models without them and then comparing the results to the performance of the models run with said variables. This straightforward yet effective approach highlights the significance of macroeconomic factors. While it would have been possible to measure the impact of each variable individually, for the sake of conciseness, the variables were tested collectively. By testing these advanced techniques and including macroeconomic variables, this analysis establishes a comprehensive framework for addressing the complexities of financial risk forecasting.

4 Results

This section compares model performances (R^2 and MSE) across market regimes (Bull, Bear, Crisis) with and without macroeconomic variables.

4.1 Model Performances with Macroeconomic Variables

The regime classification with all the dataset results shows significant variation in model performance across market conditions. The results for each regime are summarized as follows:

- **Bull Regime:** LSTM achieved an MSE of 194.86 with an R^2 of 0.93, while GRU performed better with an MSE of 108.38 and an R^2 of 0.97. GBM outperformed both recurrent neural network models with an MSE of 30.23 and a near-perfect R^2 of 1.00. These results indicate that GBM is highly effective in capturing non-linear relationships during bullish market conditions, aligning with its well-established strengths in handling structured data and feature interactions (Friedman, 2001).
- **Bear Regime:** In bearish conditions, LSTM recorded an MSE of 233.89 and an R^2 of 0.83, while GRU significantly improved on these metrics with an MSE of 91.73 and an R^2 of 0.97. GBM again performed best with an MSE of 42.11 and an R^2 of 0.99, underscoring its adaptability to volatile market environments. These findings resonate with the literature on machine learning's ability to outperform traditional models in high-volatility scenarios (Gu et al., 2020).
- **Crisis Regime:** During crisis periods, LSTM yielded an MSE of 28.27 and an R^2 of 0.99, while GRU slightly outperformed with an MSE of 21.39 and an R^2 of 0.99. GBM demonstrated exceptional performance, achieving an MSE of 4.38 and an R^2 of 1.00, indicating its robustness even in extreme market conditions. The ability of GBM to produce accurate and conservative risk estimates, such as VaR and ES, reinforces its utility in stress-testing applications (McNeil et al., 2005).

The exceptional performance of the GBM model, particularly in the bull regime with an R^2 of 1.00, highlights its potential for portfolio optimization and rebalancing strategies during periods of market growth. By leveraging GBM's ability to capture non-linear relationships and feature interactions, financial institutions can enhance their decision-making processes to maximize returns while managing risk exposure. Conversely, the robust results of LSTM and GRU models in crisis regimes, where R^2 values approached 0.99, suggest their applicability for real-time risk assessment and high-frequency trading in volatile market environments. These models can provide timely insights into market dynamics, enabling institutions to make proactive adjustments to mitigate potential losses. Furthermore, GBM's ability to produce accurate and conservative risk estimates, such as Value-at-Risk (VaR) and Expected Shortfall (ES), supports its use in stress-testing frameworks required under regulatory standards like Basel III. This adaptability across regimes demonstrates the importance of tailoring model selection to specific market conditions and objectives, ensuring comprehensive and resilient risk management strategies.

4.2 Model Performances Without Macroeconomic Variables

The regime classification without macroeconomic variables reveals the impact of excluding external economic indicators on model performance:

- **Bull Regime:** LSTM performed adequately with an MSE of 31.75 and an R^2 of 0.92, while GRU improved with an MSE of 22.97 and an R^2 of 0.94. GBM, however, delivered comparable performance with an MSE of 21.94 and an R^2 of 0.96, suggesting resilience in stable market conditions.
- **Bear Regime:** In bearish conditions, LSTM had an MSE of 14.42 and an R^2 of 0.79, whereas GRU outperformed with an MSE of 5.57 and an R^2 of 0.95. GBM's performance, with an MSE of 17.15 and an R^2 of 0.02, highlights its sensitivity to the absence of macroeconomic inputs, reaffirming the importance of external economic variables in boosting predictive accuracy (Hastie et al., 2009).
- **Crisis Regime:** During crises, LSTM showed respectable performance with an MSE of 20.06 and an R^2 of 0.91. GRU excelled with an MSE of 4.55 and an R^2 of 0.97. GBM achieved an MSE of 17.03 and an R^2 of 0.92, demonstrating consistency but highlighting its dependence on macroeconomic variables for optimal results.

The absence of macroeconomic variables significantly impacts model performance, particularly for GBM, which demonstrated sensitivity to missing external indicators. Despite this, GBM still showed resilience in stable conditions, suggesting its utility for applications where macroeconomic data is unavailable or less relevant, such as sector-specific portfolio optimization. GRU consistently outperformed in bearish and crisis regimes, with an R^2 of 0.95 and 0.97, respectively. This makes GRU particularly suitable for high-frequency trading and short-term risk management in volatile environments where real-time updates on market prices and trends are critical. Its ability to capture temporal dependencies even without macroeconomic inputs highlights its robustness in constrained-data scenarios. The findings underscore the importance of integrating external indicators where possible to enhance predictive accuracy, especially for models like GBM that rely on structured data interactions. However, the strong results of LSTM and GRU in volatile markets demonstrate that these models remain reliable for applications such as algorithmic trading and intraday risk assessment, even in the absence of macroeconomic data. These insights are valuable for financial institutions operating in niche markets or during periods when macroeconomic data is unavailable or delayed.

4.3 Comparison and Interpretation

The results indicate that GBM generally outperforms LSTM and GRU across all regimes when macroeconomic variables are included, achieving the highest predictive accuracy and lowest error metrics. This aligns with previous studies highlighting GBM's superior capacity to model complex interactions and feature importance (Friedman, 2001). The inclusion of macroeconomic variables enhances model performance, especially for GBM, which relies on structured inputs to capture intricate dependencies.

Conversely, LSTM and GRU maintain robust performance even without macroeconomic indicators, leveraging their ability to model temporal dependencies. This observation aligns with the foundational work on recurrent neural networks, which emphasizes their adaptability to sequential data (Hochreiter and Schmidhuber, 1997).

The performance differences across regimes reveal the models' adaptability to varying market conditions. In bullish and bearish markets, GBM excels due to its ability to handle non-linear dynamics effectively. During crises, LSTM and GRU demonstrate strong predictive capabilities, benefiting from their inherent sequential structure and ability to capture temporal dynamics in volatile conditions (Goodfellow et al., 2016).

These results underscore the importance of model selection based on market conditions and data availability. GBM is well-suited for scenarios requiring interpretability and risk estimation, while LSTM and GRU are effective for time-series forecasting, especially when external indicators are unavailable. This analysis provides a robust framework for applying machine learning models to financial risk forecasting and highlights the value of incorporating macroeconomic variables to enhance accuracy and robustness.

5 Analysis

The research question, data, empirical analysis, and results from this study provide valuable insights into broader financial applications, particularly in trading strategies, investment fund risk management, and regulatory compliance. By employing advanced machine learning models such as LSTM, GRU, and GBM, this research establishes a robust framework for forecasting market risks and analyzing financial volatility across diverse market regimes (Gu et al., 2020).

The results confirm that GBM excels in stable market conditions, offering portfolio managers a reliable tool for rebalancing strategies during growth periods. Its superior performance in forecasting risk metrics like Value-at-Risk (VaR) and Expected Shortfall (ES) positions it as a key instrument for designing trading strategies that account for potential downside risks (McNeil et al., 2005). Furthermore, its robustness during crisis conditions ensures effective risk mitigation, highlighting its versatility in diverse market scenarios (Friedman, 2001).

Conversely, the sequential modeling capabilities of LSTM and GRU make them particularly suitable for high-frequency and intraday trading applications, capturing short-term patterns and anomalies in financial time-series data (Hochreiter and Schmidhuber, 1997; Goodfellow et al., 2016). These results also highlight the importance of incorporating macroeconomic variables into forecasting models, which is particularly relevant for trading strategies focused on global economic trends.

For investment funds, this study provides practical tools for portfolio risk management and adherence to risk mandates. The strong performance of GBM across all market regimes demonstrates its utility for monitoring and mitigating risks in multi-asset portfolios (Hastie et al., 2009). By integrating engineered features such as rolling volatility, moving averages, and RSI, fund managers can gain deeper insights into both asset-specific and systemic risks. Additionally, the regime classification employed in this study emphasizes the importance of identifying market conditions—bullish, bearish, or crisis—to strategically adjust risk exposure. For example, GRU's strong performance in bearish regimes highlights its suitability for managing downside risks during volatile market periods (Gu et al., 2020).

The results also have significant implications for financial regulations, particularly in stress testing and capital adequacy assessments. The methodologies used in this study, especially the modeling of extreme market conditions as seen in the crisis regime analysis, can enhance the accuracy of stress-testing scenarios required under frameworks such as Basel III (McNeil et al., 2005). The inclusion of macroeconomic variables aligns with regulatory requirements to incorporate external economic factors in risk assessments, providing a more comprehensive view of systemic risks and financial stability.

In a broader context, this research highlights the growing importance of machine learning models in financial risk management and their ability to complement traditional approaches. The multi-model methodology demonstrates the need for flexible and adaptive techniques to address varying market conditions. While GBM excels in stable and growing markets, LSTM and GRU show strength in handling volatile and unpredictable environments.

Overall, this study bridges the gap between advanced quantitative techniques and their practical applications, demonstrating the transformative potential of machine learning in modern financial markets.

6 Conclusion

This project set out to explore the use of advanced machine learning models—LSTM, GRU, and GBM—for forecasting financial risk and market volatility across diverse market regimes. Using a comprehensive dataset that included stock-specific variables and macroeconomic indicators, the study demonstrated the models' strengths and limitations in capturing non-linear relationships and adapting to different market conditions. Several key lessons were learned through this research.

First, the inclusion of macroeconomic variables significantly improved model performance, particularly for GBM, which relies on structured inputs to capture complex interactions . Second, the choice of model depends heavily on the market regime and the specific application. GBM excelled in stable and growing markets, achieving the highest predictive accuracy, while LSTM and GRU proved effective in volatile and crisis conditions due to their ability to model temporal dependencies . The results emphasize the importance of regime classification for risk forecasting. Identifying market conditions—bullish, bearish, or crisis—enabled more tailored predictions, highlighting the practical applications of these models in trading strategies, portfolio risk management, and regulatory compliance.

Furthermore, the study underscored the value of macroeconomic features, such as rolling volatility, MACD, and RSI, in enhancing predictive accuracy. These features, rooted in financial risk management best practices , provide a nuanced understanding of asset-specific and systemic risks. This project also highlighted the growing role of machine learning in financial risk management. While traditional models, such as GARCH , remain foundational, the integration of machine learning techniques offers improved adaptability and performance in dynamic financial markets. The lessons learned from this study provide a robust framework for future applications, including stress testing , high-frequency trading, and global portfolio optimization.

In conclusion, this research demonstrates the practical utility of combining advanced models and engineered features for financial forecasting. By addressing complex financial patterns and leveraging diverse datasets, this project contributes to the growing body of knowledge on the intersection of machine learning and financial risk management. Future research could extend this work by incorporating alternative data sources, such as sentiment analysis and news data, and exploring cross-asset correlations to provide deeper insights into systemic risks and global market dynamics.

References

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

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4):394–419.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.