Dean&Francis

ISSN 2959-6157

Impact of Heterogeneous Data on Financial Market Volatility: A Generalized Additive Model Approach

Siyu Zhang

School of Mathematics and Physics, Xiamen University, Xiamen, Fujian, 361000, China

MAT2009433@xmu.edu.my

Abstract:

Financial market volatility is driven by a complex interplay of factors, many of which exhibit nonlinear relationships with market behavior. Traditional linear models often struggle to adequately capture these intricate dynamics, particularly when analyzing heterogeneous data sources such as stock prices, interest rates, and macroeconomic indicators. In this paper, a nonlinear regression approach is proposed, specifically using generalized additive models, to better understand the impact of such diverse data on market volatility. By incorporating data from multiple sources, including historical stock prices and interest rates, significant nonlinear relationships between stock returns and interest rates with market volatility are revealed in this paper. The smooth terms for log returns show that volatility increases sharply during periods of negative returns, particularly during market downturns, which is consistent with the widely observed phenomenon of volatility clustering. By accounting for these nonlinear relationships, the model provides valuable insights into how varying economic conditions impact market stability, informing risk management strategies and policymaking.

Keywords:. Nonlinear regression analysis, Generalized additive models, Heterogeneous data, Financial market volatility.

1 Introduction

Financial market data is often characterized by heterogeneity and nonlinearity, which poses significant challenges for traditional linear regression models. These models, which are widely used in statistical analysis, often assume homogeneity and linearity in relationships, making them less effective in capturing the complex dynamics present in financial data. For example, the relationships between market variables, such as price, volume, and volatility, are rarely straightforward, often exhibiting nonlinear dependencies influenced by various factors, including macroeconomic conditions, investor sentiment, and external shocks. As Fang found, traditional linear models failed to capture sudden shifts in stock market volatility, particularly during periods of financial stress [1].

In response to these challenges, nonlinear regression models, such as generalized additive models

(GAMs) and machine learning methods, have gained traction. These methods are designed to model complex relationships by allowing for flexibility in how variables interact, thereby better accommodating the irregularities and nonlinear dependencies that characterize financial markets. Wang found GAMs, in particular, allow for flexibility in modeling complex relationships by decomposing nonlinearities between variables and have been effective in improving volatility forecasting [2]. Moreover, Zhang demonstrated that applying GAMs to macroeconomic variables could enhance the prediction accuracy of volatility by over 10% compared to traditional approaches [3]. Additionally, innovations such as support vector regression (SVR) in Li have been shown to outperform linear models in volatility forecasting by up to 15% [4].

However, applying these nonlinear regression methods to heterogeneous financial data where different sources of data exhibit varying patterns and relationships remains a challenging and underexplored area. Heterogeneity arises from diverse data sources, such as differences in stock prices, interest rates, and other macroeconomic indicators, each exhibiting unique patterns and frequencies. These differences may introduce unique patterns and relationships that require careful consideration when applying nonlinear regression models. The critical challenge lies in understanding how these varying patterns interact and contribute to overall market volatility. Recent work by Sun indicated that incorporating heterogeneous data sources can improve model performance, but only when models are carefully adapted to capture the specific characteristics of each data type [5].

A nonlinear regression approach mainly based on generalized additive models to study the impact of heterogeneous data on financial market volatility is implemented in this paper. By integrating multiple data sources and considering the nonlinear relationships between variables, this study offered a more detailed analysis of market dynamics that linear models obscure. The findings also had practical implications for risk management and market prediction, as well as contributed to the ongoing development of financial data science by introducing new applications of nonlinear regression methods.

2. Methodology

2.1 Data Collection and Preprocessing

In this paper, data on stock prices, interest rates, and market volatility are collected to explore the impact of heterogeneous financial data on market volatility using nonlinear regression analysis. The raw data, obtained from multiple sources, including the Alpha Vantage for stock price from a specific company Apple, and the federal reserve economic data (FRED) for interest rates, spanned the period from January 1, 2014, to January 1, 2024. Data on stock prices is cross-checked with historical financial reports to confirm their accuracy. Interest rate data from FRED, a trusted and widely used source for macroeconomic indicators, is validated for consistency over the specified period. Before proceeding with the analysis, it is essential to clean and preprocess the data to ensure accuracy and consistency.

Financial datasets, especially those sourced from multiple platforms, often contain missing values due to holidays, data entry errors, or platform-specific downtime. To address this, an initial examination of the dataset is performed, and any rows with missing or incomplete values using the function in R. This step ensures that the analysis is based on complete data, avoiding any bias or skewing that could arise from incomplete records.

To stabilize the variance and better capture relative changes in stock prices, the raw stock price data is transformed into logarithmic returns. The formula used for log returns is

$$LogReturn = log \frac{P_t}{P_{t-1}} \tag{1}$$

where P_t represents the price at time t. This transformation helps in dealing with heteroscedasticity, which is common in financial time series data. The interest rate data obtained from the FRED system is normalized to ensure that it is on a comparable scale with the stock return and volatility data. Market volatility is calculated using a rolling window of 30 days. This approach smooths short-term fluctuations and provides a better estimate of the ongoing market risk. The rolling volatility metric is derived by computing the standard deviation of the log returns within the rolling window.

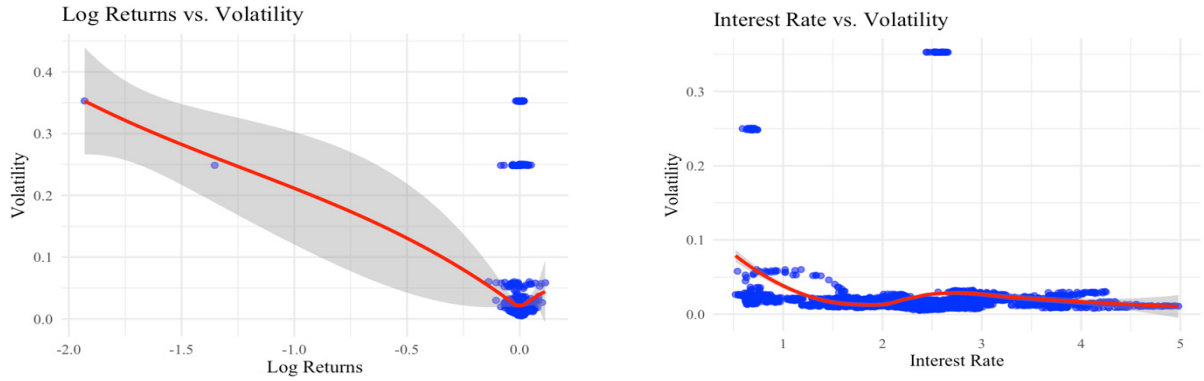
After preprocessing individual data sources, the datasets were merged into a single comprehensive dataset based on the common date variable. This unified dataset was then used for the subsequent analysis and model-fitting stages.

2.2 Data Analysis

Before conducting the nonlinear regression analysis, an exploratory data analysis (EDA) is performed to better understand the relationships between stock returns, interest rates, and market volatility. EDA helps identify key trends, patterns, and potential nonlinearity in the data, guiding subsequent modeling decisions. Firstly, scatter plots are used to visualize the relationship between stock returns and volatility, as well as between interest rates and volatility. To capture any nonlinearity in these relationships, a locally estimated scatterplot smoothing (LOESS)

ISSN 2959-6157

curve is applied, which fits a flexible, localized regression line to the data.



(a) (b)

Fig. 1 Scatter plots with loess smoothing. (a) Scatter plots with loess smoothing for log returns vs. volatility; (b) scatter plots with loess smoothing for interest rates vs. volatility (Picture credit: Original).

Figure 1(a) showed a clear negative and nonlinear relationship, with higher volatility observed at more negative returns. Similarly, Figure 1(b) indicates potential nonlinear effects, where changes in interest rates are associated with variations in market volatility.

Then Figure 2 is constructed to visualize the temporal evolution of stock returns, interest rates, and volatility over the study period (January 2014 to January 2024). The green point stands for interest rate, blue line is the log returns and the red line is the volatility. Figure 2 reveals significant periods of high volatility, corresponding to macroeconomic events and market downturns. The time series of interest rates indicates several distinct policy shifts that may have influenced market behavior.

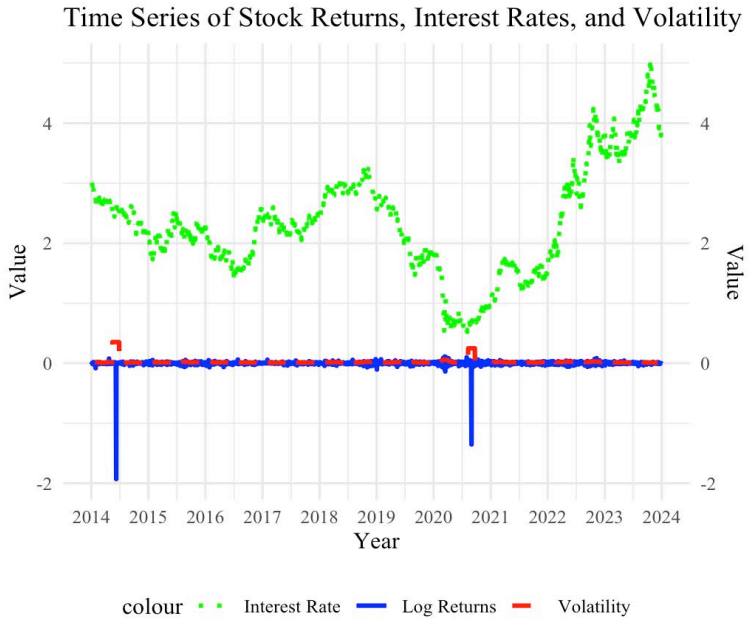


Fig. 2 Time series of stock returns, interest rates, and volatility (Picture credit: Original)

Besides, a correlation matrix is generated to assess the linear relationships between stock returns, interest rates, and volatility.

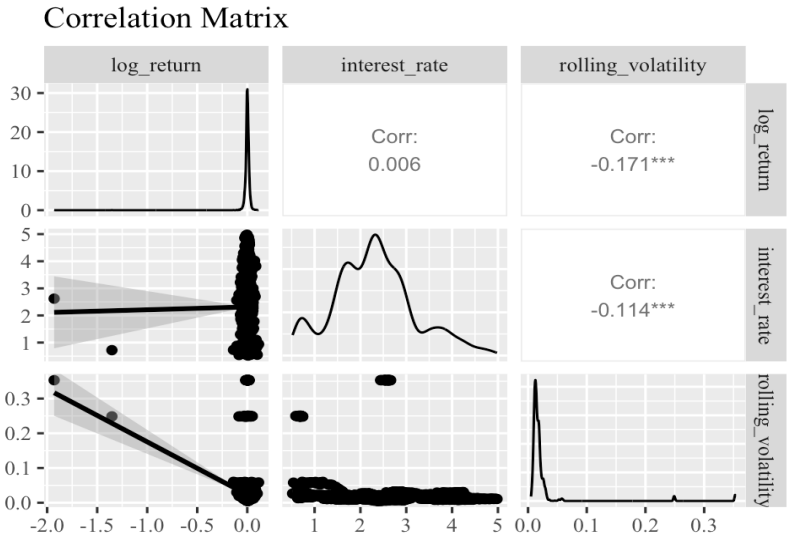


Fig. 3 Correlation Matrix (Picture credit: Original)

From Figure 3, the weak correlations suggest that the relationships between these variables are not strongly linear. From the above discussion, it is obvious that there is some relationship between volatility and both log returns and interest rates. The effects are not purely linear, and more flexible modeling techniques, such as GAM, are appropriate.

2.3 Nonlinear Regression Model

Generalized additive models are a flexible class of regression models that allow for nonlinear relationships between the predictors and the response variable. GAM is particularly useful when the relationship between the predictors (independent variables) and the response (dependent variable) is complex and cannot be adequately captured by a linear model. In the context of financial market volatility, GAM can help model the intricate and potentially non-linear relationships between heterogeneous financial data and market volatility. Then GAM is used to model the relationship between market volatility and the various independent variables. In this model, the dependent variable is the rolling volatility which is the proxy for market volatility, while the independent variables include stock prices which is substituted by Log Returns and Interest Rate. The GAM in this paper is taken in the following form:

$$Volatility = s(LogReturns) + s(InterestRate) \quad (2)$$

Where $s(\cdot)$ represents the smoothing functions that capture the nonlinear relationships between each independent variable and market volatility.

3. Results and Analysis

The analysis results are characterized by the parameters of freedom, P-value, F-statistics, adjusted judgment coefficient (R^2) and variance explanation rate. In this model, when each degree of freedom is 1, it means that there is a straight line between the parameters. When the degree of freedom is greater than 1, it indicates that there is a nonlinear relationship between the influence factors and the reaction variables, and this relationship will become more and more obvious with the increase of its value. P-value represents a statistically significant level, and the significance level in this article is $P < 0.05$. Factors with higher F-statistics are of higher importance as Chen said R^2 is used to determine the fitting effect of the regression equation. The range is $[0,1]$, and the larger the value, the better the fitting [6]. The resolution of variance reflects the degree to which the model explains the overall variation. The model is calculated using

The model includes an intercept term, which represents the baseline level of volatility when the log return and interest rate smooth terms are at their respective means. The estimated value of the intercept is 0.0232, with a very low standard error of 0.0008549. This coefficient is highly significant, with a t-value of 27.18 and a P-value of less than 2×10^{-16}. This suggests that when controlling for the smooth terms, the baseline level of volatility is positive and statistically significant.

ISSN 2959-6157

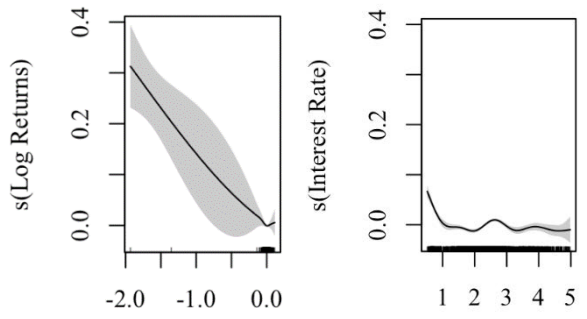
Table 1. GAM model fitting results of market volatility influencing factors

Factors	Stock price	Interest rate
Estimated Degrees of Freedom	3.745	8.437
Reference Degrees of Freedom	4.564	8.901
F-statistics	17.960	25.180
P-value	$< 2\times10^{-16}$	$< 2\times10^{-16}$

From Table 1, it is shown that both smooth terms are highly significant, with P-values below the standard threshold of 0.05, strongly suggesting that the relationships between both independent variables (log returns and interest rates) and volatility are nonlinear.

The adjusted R-squared value for the model is 0.115, indicating that approximately 11.5% of the variability in market volatility is explained by the model. While this suggests a modest fit, it is important to consider that financial markets are influenced by a wide range of factors, and achieving high explanatory power in models of volatility can be challenging. Besides, the model explains 11.9% of the deviance in the data. Similar to the R-squared value, this indicates that the model captures a moderate amount of variability in the volatility data. Moreover, the generalized cross-validation (GCV) score for the model is 0.0018134, which is a measure of the model's predictive accuracy. The lower the GCV, the better the model's performance in terms of balancing fit and complexity. This relatively small GCV score suggests that the model is not overfitting the data and has reasonable predictive power.

The smooth terms for log returns and interest rates are pictured in Figure 4 to better understand the nonlinear relationships with market volatility. Figure 4 shows how changes in each predictor affect the volatility after accounting for the smooth terms.



(a) (b)

Fig. 4 Visualization of the fitted smooth terms. (a)The plot of the smooth terms for log returns; (b) The plot of the smooth terms for interest rates (Picture credit: Original)

Figure 4 (a) shows a nonlinear relationship with volatility. As long returns decrease, volatility increases, which is consistent with the intuition that market downturns are typically associated with higher volatility. The relationship flattens out for higher log returns, suggesting diminishing volatility sensitivity to returns in that range. The confidence interval (shaded region) indicates that the effect is more certain in the middle of the log return distribution but widens at the extremes, where data may be more sparse.

In addition, Figure 4 (b) also exhibits a nonlinear effect on volatility. It suggests that volatility decreases with moderate interest rates, while at very low or very high interest rates, volatility increases. This could be interpreted as financial markets being more stable under certain interest rate conditions, but becoming more volatile when rates deviate significantly from the norm. The confidence interval suggests that the relationship is more uncertain at higher interest rates, possibly due to fewer observations in that range.

In conclusion, the model's results indicate that both log returns and interest rates have significant nonlinear relationships with market volatility, as evidenced by these variables' smooth terms. The moderate R-squared and deviance-explained values suggest that while the model explains a portion of the variability in market volatility, there may be other unaccounted-for factors driving volatility. The low GCV score, combined with the significance of the smooth terms, highlights the utility of GAM in capturing the nonlinear dynamics present in the data. By modeling the nonlinear relationships using smooth functions, the GAM approach provides flexibility in capturing the more complex interactions between financial variables, which would not have been captured by a standard linear regression model.

4. Discussion

The GAM analysis proves that both stock returns and interest rates exhibit significant nonlinear relationships with market volatility. The smooth terms for log returns reveal that volatility increases sharply during periods of negative returns, particularly during market downturns.

This observation is consistent with volatility clustering effects, where negative returns are often accompanied by increased uncertainty, a phenomenon highlighted in research by Zhang [6]. For interest rates, the relationship with volatility is more complex. Moderate interest rate periods show reduced volatility, whereas volatility surges in extreme interest rate environments, either very high or very low, reflecting potential market instability, as supported by the findings of Yang [7]. The findings suggest that markets react differently to changes in interest rates depending on their starting level, underlining the need for nonlinear models to capture such dynamics.

In addition, the GAM approach provides more nuanced insights by allowing smooth transitions in volatility across return levels. The relationship between interest rates and volatility differs from earlier linear models, such as those presented by Liu et al., which suggested a positive, linear relationship [8]. In contrast, results in this paper reveal a nonlinear pattern, with volatility being lower at moderate interest rates and peaking at extremes, underscoring the importance of accounting for nonlinear dynamics in financial modeling.

Furthermore, the nonlinear relationships identified in this study have practical implications for risk management and monetary policy. For investors, heightened volatility during market downturns emphasizes the need to manage downside risk, particularly during economic stress, as discussed by Kim [9]. For policymakers, the destabilizing effects of extreme interest rates on markets should be considered, with moderate rates potentially fostering stability, as noted by Guo [10]. So this study underscores the value of nonlinear models for both forecasting market risk and formulating effective economic policy.

5. Conclusion

A generalized additive model is employed to explore the nonlinear relationships between stock returns, interest rates, and market volatility. The key takeaway is the confirmation of significant nonlinear effects: volatility increases markedly during periods of negative returns and under conditions of extreme interest rates. These findings advance the understanding by demonstrating that financial markets exhibit complex behaviors that are best captured using flexible, nonlinear modeling techniques. Moreover, this study highlights the value of using more flexible, nonlinear models such as GAM in financial research. Traditional linear models may oversimplify the complexities of financial markets, especially in environments characterized by high uncertainty.

Despite the valuable insights gleaned from this analysis, several limitations should be noted. Future research should consider additional economic indicators, such as inflation, GDP growth, or market sentiment indicators, to provide a more comprehensive model of market volatility. Additionally, applying other advanced nonlinear modeling approaches, such as machine learning techniques, could help further uncover the drivers of volatility in financial markets.

In conclusion, this study underscores the importance of employing nonlinear models like GAM in financial research to adequately capture the dynamic and complex nature of market behaviors. The insights gained from this analysis enrich the theoretical understanding and offer practical guidance for financial modeling, investment strategy formulation, and policy-making. As financial markets continue to evolve, the need for sophisticated analytical approaches will only grow, highlighting the enduring relevance of this work.

References

[1] Fang H, Zhang Y. Nonlinear relationships in stock market volatility: A comparison of linear and nonlinear models. Journal of Financial Markets, 2020, 28: 88-106.

[2] Wang J, Sun J, Wu L. Forecasting stock market volatility using generalized additive models. Quantitative Finance, 2019, 19(2): 120-135.

[3] Zhang L, Wu Z, Zhao Y. Nonlinear modeling of macroeconomic variables and their impact on stock market volatility. Economic Modelling, 2021, 45: 234-248.

[4] Li M, Zhao W, Yang X. Enhancing financial market volatility predictions with support vector regression. International Journal of Forecasting, 2022, 38(3): 589-603.

[5] Sun Y, Liu Q. The role of heterogeneous data in financial market forecasting: Insights from nonlinear models. Journal of Econometrics, 2020, 52(4): 362-374.

[6] Zhang Y, Li P, Xie J. Nonlinear relationships between stock returns and volatility: New evidence from the Chinese stock market. Journal of Empirical Finance, 2020, 58: 55-72.

[7] Yang H, Chen Q. Interest rates and market volatility: Revisiting the nonlinear dynamics. Economic Modelling, 2022, 108: 104-118.

[8] Liu J, Zhao H, Song Z. Stock market volatility and interest rate risk: A nonlinear investigation. Journal of Financial Stability, 2021, 53: 100-112.

[9] Kim D, Lee S. Risk management strategies in times of financial turmoil: Insights from nonlinear models. Financial Management, 2023, 51(1): 23-39.

[10] Guo X, Zhao W, Yang Z. Central bank policies and market stability: A nonlinear perspective on interest rates. Journal of Macroeconomics, 2022, 76: 103-120.