International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Realized Volatility Prediction in Stock Market

Tejaswini Gadham[1], Aakash Bolla[2] , Shay Adithya Kumar[3]

[1,2]Student Scholar
Department of Computer Science Engineering
Swami Vivekananda Institute of Technology
Tejaswinigadham6@gmail.com

[3]Student Scholar
Department of Electrical and Electronics Engineering
Hyderabad Institute of Technology and Management
adiithyashah@gmail.com

---***---

Abstract - In our study, we examine various MIDAS (Mixed Data Sampling) regression models as a predictive tool for stock market volatility. These models vary in the regressors used, including squared return, absolute return, realized volatility, realized power, and return range.By analyzing equity return data, we find that the ability to see daily, consisting of 5-minute absolute returns, emerges as the most influential predictor of future volatility This is measured by an increase in quadratic variation. Notably, this approach outperforms models that rely solely on observed changes, which are based on prior increases in quadratic change. Surprisingly, our findings suggest that using high-frequency (5-minute) data directly does not improve the prediction of volatility.In summary, our study reveals the effectiveness of daily realized power, using 5-minute absolute returns, as a good indicator of future volatility in the stock market. We show that this approach outperforms models that rely solely on observed changes. Surprisingly there is no bet using high-frequency data.

Key Words: MIDAS, Squared return, Absolute return

1.INTRODUCTION

The analysis focuses on future scenario changes measured as an increase in quadratic change over a one-week to one-month period Researchers use MIDAS regression to examine the predictive power of different daily returns, a squared return, . absolute return, realized volatility, realized power included (intra-daily -By using the same number of parameters and maximum lags, including sum of absolute returns), daily range, the MIDAS system facilitates if regressors are compared directly. The findings show that daily realized power consistently outperforms other indicators of daily volatility, including squared absolute return, realized volatility, and daily range, at the Dow Jones Index and individual stock returns across six series. The superiority of the observed power is reflected not only in the in-sample goodness-of-use measure but also in the out-of-sample prediction Daily range also appears to be a valuable predictor, more than squared with absolutely every day useful. Importantly, the study departs from the traditional auto-regression model-building approach common in the ARCH literature. Comparisons with auto-regressive volatility models show that the MIDAS regression produces superior predictive performance for both in-sample and out-of-sample volatility predictions In addition, the MIDAS regressions allow a simple weighting function, which is determined by two parameters estimated from the data. The researchers found that daily follow-up delays of more than about 50 days did not contribute significantly to any prediction of relapse. Surprisingly, the study shows that using high-frequency data directly in volatility forecasting is not optimal compared to using daily regression

2. HARDWARE AND SOFTWARE SETUP

The brief introduction of different modules used in this project is discussed below:

2.1 HARDWARE REQUIREMENTS:

a. System : Pentium IV 2.4 GHz.
b. Hard Disk : 40 GB.
c. Floppy Drive : 1.44 Mb.
d. Monitor : 15 VGA Colour.
e. Mouse : Logitech.
f. Ram : 512 MB.

2.2 SOFTWARE REQUIREMENTS:

a. Operating system : Windows 7 Professional.
b. Coding Language : python

3. IMPLEMENTATION

3.1 MODULES:

To implement this project, we designed the following module.

a. **Upload Drug Data Set:** This module provides a button to upload a drug data set. Once the data set is uploaded, the system processes it and produces the desired output.

b. **Preprocessing Data set:** This module allows the user to divide the data set into training and testing subsets. It ensures that the data set is properly partitioned for model training and analysis.
c. **Train regression without optimization:** This module trains a regression model on the uploaded data set without any optimization method. The model is trained with predefined parameters and settings.
d. **Polynomial Optimized Linear Regression:** This module trains a regression model with both linear polynomial optimization methods. It uses advanced algorithms and optimization techniques to improve the performance of the regression model. Outputs from this module include appropriate assessment standards and model performance indicators.
e. **Pre & Post Optimization SSE Graph:** This module generates a graph of the cumulative squared error (SSE) for the regression model before and after optimization. The graph provides a visual representation of how optimization techniques affect the accuracy and efficiency of the model.

Each module is designed to perform specific tasks and provide the desired information. Users can interact with the system by clicking the corresponding buttons for each module. The purpose of this project is to optimize analysis and regression models using the data provided.

4. ARCHITECTURE

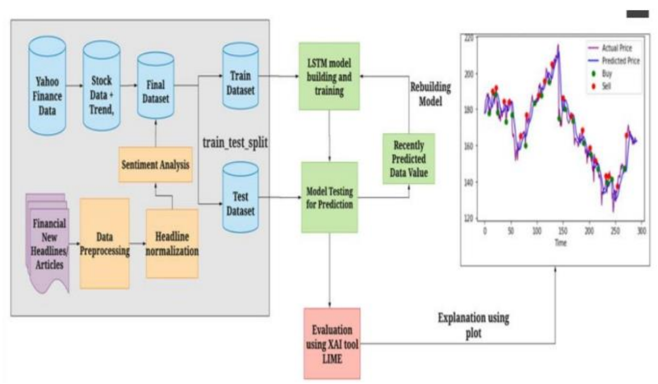

Fig-4: Architecture

4.1 LSTM (LONG SHORT-TERM MEMORY) ALGORITHM :

LSTM, short for short-term short-term memory, is a type of recurrent neural network (RNN) that excels at long-term dependent capture, and has become increasingly popular in various industries for its control capabilities whole sequences of data are processed rather than individual so data points such as images. This makes it ideally suited for applications such as speech recognition and machine translation. One of the main advantages of LSTM is its unique structure which includes feedback links. These connections enable the network to efficiently process all sequential data, allowing long-term dependencies to be detected and modeled. This is in contrast to traditional RNNs, which tend to struggle with frequent degradation, limiting their remote capture capabilities.LSTM achieves its sophisticated functionality through the use of specialized memory cells, which can choose to store or forget information at any given time step.

There are three main types of networks: input gateway, forget gateway, and output gateway. These gates control the flow of information into the network, determining how to combine new information, what information to forget, and how to calculate the final result.

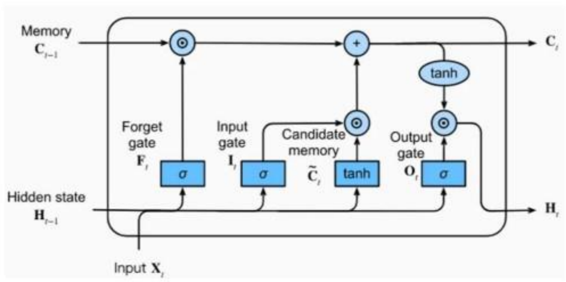

Fig-4.1: LSTM Architecture

5. SYSTEM TEST

The primary purpose of testing is to identify the defect or deficiencies in a work item. This involves systematically testing a software system or product to identify potential bugs or vulnerabilities. Testing provides a means of evaluating the performance and operation of components, sub-assemblies, assemblies, or the final product itself. The goal is to ensure that the software system meets specified requirements and user expectations and does not experience unacceptable failures. Testing includes the process of rigorously and deliberately applying a software system to test cases and scenarios. This includes creating test cases, entering specific data, and analyzing the output to see if it matches expected behavior. The goal is to identify any discrepancies or inconsistencies that may indicate errors or deviations from the desired functionality.

The entire testing system consists of tests for specific purposes. Each test is designed to meet specific testing requirements. Some common tests are:

a. **Functional Testing:** This test focuses on verifying the functional aspects of the software system to ensure that it works as intended based on specified requirements.
b. **Performance testing:** Performance testing examines system behavior under various operating conditions to evaluate performance, scalability, and resource utilization.
c. **Security testing:** Security testing aims to identify weaknesses and vulnerabilities in the system's security mechanisms, and to ensure that critical data is adequately protected.
d. **Usability Testing:** Usability testing examines the user friendliness and convenience of the software system, and evaluates the ease of use and overall user experience,
e. **Regression testing:** Regression testing involves retesting a previously tested functionality to ensure that there are no changes or updates in the software that would introduce new bugs or affect existing functionality.
f. **Integration testing:** Integration testing monitors communication and compatibility between parts or modules of a software system to ensure smooth communication and seamless integration.

g. **Acceptance testing:** Acceptance testing is conducted to verify that the software system meets the acceptance criteria defined by the end users or stakeholders.

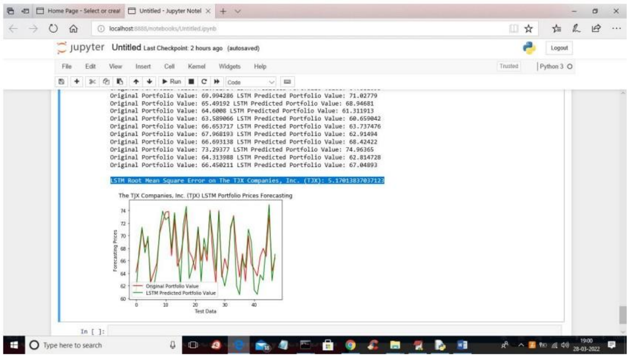

Fig-5: Output

In above screen with TJX data set, we got MSE value as 5.17 and both lines are having little difference. So, by using single LSTM model we can forecast asset values of any stock Company.

6. CONCLUSION

In our analysis of the predictability of regression volatility using MIDAS regression, we found several notable findings. Our analytical approach allows us to compare forecast models that use different parameters of variability, frequency, and lag length. Although the main focus of our paper is volatility forecasting, the MIDAS model itself is versatile and can be applied to a wide range of analyzes with data sampled at different frequencies. We present several interesting observations on the forecasting of weekly to monthly observed volatility in stock markets. First, we find that consistently observed power appears to be a better predictor as opposed to consistently observed change. Furthermore, we find that daily and daily absolute returns outperform their squared return counterparts in capturing changes in volatile future returns. This finding challenges the heavy emphasis in the literature on class return as the primary measure. Furthermore, we confirm that daily range exhibits a unique ability to predict future volatility and is second only to observed strength. This is consistent with Gallant et al. (1999), Alizadeh et al. (2002), and Engle and Gallo (2003), who used different methods and data sources. Finally, we show that incorporating high-frequency data directly does not necessarily improve volatility forecasts.

ACKNOWLEDGEMENT

We would like to thank all the authors of different research papers referred to during the writing of this paper. It was very knowledge-gaining and helpful for further research to be done in the future.

REFERENCES

1. A¨ıt-Sahalia, Y., Mykland, P. A. and Zhang, L. (2003), “How often to sample a continuoustime process in the presence of market microstructure noise”, Working Paper w9611, NBER.
2. Alizadeh, S., M. Brandt and F. X. Diebold, (2002), ”Range-based estimation of stochastic volatility models”, Journal of Finance, 57, 1047-1091.
3. Andersen, T. and T. Bollerslev (1997), “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long Run in high-frequency Returns”, Journal of Finance, 52, 975-1005.
4. Andersen, T. and T. Bollerslev (1998), “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts”, International Economic Review, 39, 885-905.
5. Andersen, T., T. Bollerslev, F. X. Diebold and P. Labys (2001), “The Distribution of Exchange Rate Volatility”, Journal of American Statistical Association, 96, 42-55.
6. Andersen, T.G., Bollerslev, T. and F.X. Diebold, (2002), “ Parametric and Nonparametric Volatility Measurement,” in L.P. Hansen and Y. A¨ıt-Sahalia (eds.), Handbook of Financial Econometrics . Amsterdam: North-Holland, forthcoming.
7. Andersen, T., T. Bollerslev, F. X. Diebold and P. Labys (2003), “Modeling and Forecasting Realized Volatility”, Econometrica, 71, 529-626.
8. Andersen, T.G., Bollerslev, T. and F.X. Diebold, (2003), “Some Like it Smooth, and Some Like it Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility,” Manuscript, Duke, Northwestern and University of Pennsylvania.
9. Andreou, E. and E. Ghysels (2002), “Rolling sample volatility estimators: some new theoretical, simulation and empirical results” Journal of Business and Economic Statistics, 20, 3, 363-376. 49
10. Baillie, R.T., T. Bollerslev and H. O. Mikkelsen (1996), “Fractionally integrated generalized autoregressive conditional heteroskedasticity” Journal of Econometrics, 74, 3-30.
11. Bandi, F.M. and J.R. Russell,(2003), “Microstructure noise, realized volatility, and optimal sampling”, Working paper, Graduate School of Business, The University of Chicago.
12. Barndorff-Nielsen, O. and N. Shephard (2001), “Non-Gaussian Ornstein–Uhlenbeckbased models and some of their uses in financial economics (with discussion), Journal of the Royal Statistical Society, Series B, 63, 167-241.
13. Barndorff-Nielsen, O. and N. Shephard (2002a), “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”, Journal of the Royal Statistical Society, Series B, 64, 253-280.
14. Barndorff-Nielsen, O. and N. Shephard (2002b) “Estimating quadratic variation using realised variance” Journal of Applied Econometrics, 17, 457-477.
15. Barndorff-Nielsen, O. and N. Shephard (2003a) “How accurate is the asymptotic approximation to the distribution of realised volatility?” in D.W.K. Andrews, J. Powell, P. Ruud and J. Stock (ed.), Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg, Cambridge University Press, (forthcoming).