A CASE STUD FOR STOCK MARKET PREDICTION REALIZED VOLITILITY PREDICTION IN STOCK MARKET

J. Pujitha[1], Dr.I.Satyanarayana[2], Madhani Swathi[3], Karnati Rithika Reddy[4], Nalla Mukesh Reddy[5], Marikanti Vamshi Krishna[6], Lenkala Kishore[7]

[1]Assistant Professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

[2]Professor, Department of MECH & PRINCIPAL, Sri Indu Institute of Engineering & Technology, Hyderabad

[3,4,5,6,7] IV[th] Btech Student, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

ABSTRACT

Stock market is an important and active part of nowadays financial markets. Stocktime series volatility analysis is regarded as one of the most challenging time series forecasting due to the hard-to-predict volatility observed in worldwide stock markets. In this paper we argue that the stock market state is dynamic and invisible but it will be influenced by some visible stock market information. Existing research on financial time series analysis and stock market volatility prediction can be classified into two categories:in depth study of one market factor on the stock market volatility prediction or prediction by combining historical price fluctuations with either trading volume or news. In this paper we present a service-oriented multi-kernel based learning framework (MKL) for stock volatility analysis. Our MKL service framework promotes a two-tier learning architecture. In the top tier, we develop a suite of data preparation and data transformation techniques to provide a source-specific modeling, which transforms and normalizes a source specific input dataset into the MKL ready data representation. Then we apply data alignment techniques to prepare the datasets from multiple information sources based on the classification model we choose for cross-source correlation analysis. In the next tier, we develop model integration methods to perform three analytic tasks: (i)building one sub-kernel per source, (ii) learning and tuning the weights for sub-kernels through weight adjustment methods and (iii) performing multi-kernel based cross- correlation analysis of market volatility. To validate the effectiveness of our service oriented MKL approach, we performed experiments on HKEx 2001 stock market datasets with three important market information sources: historical prices, trading

volumes and stock related news articles. Our experiments show that 1) multi-kernel learning method has a higher degree of accuracy and a lower degree of false prediction, compared to existing single kernel methods; and 2) integrating both news and trading volume data with historical stock price information can significantly improve the effectiveness of stock market volatility prediction, compared to many existing prediction methods.

1.INTRODUCTION:

The conditional volatility literature, starting with Engle's (1982) ARCH-class of models, has been successful at capturing the dynamics of return variance using simple parametric models. A measure of that success is the widespread use of such models in all areas of finance by academics and practitioners alike. And while most researchers would agree that it is important to have a good prediction model of conditional volatility, the question of what model to use is still unsettled**.**

When it comes to forecasting volatility, there are many existing models in addition to the benchmark ARCH/GARCH models of Engle (1982) and Bollerslev (1986) which cast future variance as a polynomial of past squared returns, i.e., $\hat{\sigma}^2_{t+1|t} \equiv A(L)r^2_t$. One alternative is to look for variables, other than squared returns, that relate to future volatility. Ding et al. (1993) and several others show that low- frequency components of volatility might be better captured by absolute returns instead of squared returns. Also, Alizadeh et al. (2002) and Gallant et al. (1999) find daily ranges (high-low price ranges) to be good predictors of volatility. Another rapidly growing research area focuses on data-driven models of realized volatility computed from intra-daily returns sampled at very short intervals such as 5 minutes (Andersen and Bollerslev (1998)).1 All these models suggest a variety of possible ways to forecast volatility. Hence, it seems natural to ask whether some of the suggested predictors are clearly dominated by others and whether there are real benefits from using high-frequency data.2 These questions have proven difficult to answer because the models considered are so different in terms of regressors, frequencies, parameterizations, and return histories, that is it difficult to directlycompare them.

We use Mixed Data Sampling (henceforth MIDAS) regression models introduced in Ghysels, Santa-Clara and Valkanov (2002a,b) to provide answers to these questions.

MIDAS regressions allow us to run parsimoniously parameterized regressions of data observed at different frequencies. There are several advantages of using mixed data sampling regressions. They allow us to study, in a unified framework, the forecasting performance of a large class of volatility models which involve:

(i) data sampled at different frequencies;

(ii) various past data window lengths; and

(iii) different regressors. The specification of the regressions

combine recent developments regarding estimation of volatility and a not so recent literature on distributed lag models.3 We focus on predicting future conditional variance, measured as increments in quadratic variation (or its log transformation) from one week to one month horizons, because these are the horizonsthat are most widely used for option pricing, portfolio management, and hedging applications.The MIDAS regressions can also be used to model asymmetries and the joint forecasting power of the regressors. In fact, Engle and Gallo (2003) use the multiplicative error model (MEM) of Engle (2002) and find improvements in forecasting volatility from the joint use of absolute returns, daily ranges, and realized volatilities using S&P 500 index returns data. Interestingly enough, their results agree with ours, despite the different data set and different method, as they argue that range-based measures in particular provide a very good forecast of future volatility.

2. METHOD OF IMPLEMENTATION

Predicting the realized volatility of the stock market is a complex task, and there are several methods that can be used to implement such a project. Here are some stepsthat can be taken to implement a realized volatility prediction project:

1. **Data collection:** The first step in implementing a realized volatility prediction project is to collect historical data on the stock market. This can be done by accessing publicly available data sources such as Yahoo Finance or Google Finance. The data collected should include the prices of the stocks, the volume of trades, and any other relevant variables that may affect stock prices.

2. **Data preprocessing:** Once the data has been collected, it should be preprocessed to ensure that it is clean and usable for analysis. This may include removing missing or duplicate data points, standardizing the data, and transforming the data into a format that can be easily analyzed.

3. **Feature engineering:** In order to predict realized volatility, it is importantto identify relevant features that may affect stock prices. This may include technical indicators such as moving averages, relative strength index (RSI), or momentum indicators. Other factors such as news events, economic indicators, or company-specific events may also be important features to consider.

4. **Model selection:** There are several models that can be used for realized volatility prediction, including autoregressive integrated moving average (MAE).

5. **Model deployment:** Once the model has been trained and evaluated, it can be deployed to make real-time predictions of realized volatility. This may involve integrating the model into an existing trading platform or developing a custom application for making predictions.
Overall, implementing a realized volatility prediction project involves a combination of data collection, preprocessing, feature engineering, model selection, training, and evaluation. Each of these steps requires careful consideration and attention to detail in order to achieve accurate and reliable predictions of realized volatility

(ARIMA) models, general autoregressive conditional heteroscedasticity (GARCH) models, or machine learning models such as neural networks or support vector machines (SVMs). The choice of model will depend on the specific characteristics of the data and the performance metrics of interest.

6. **Model training and evaluation:** Once a model has been selected, it should be trained using historical data and evaluated using a test dataset. Performance metrics such as mean squared error (MSE), mean absolute error

3.CONCLUSION:

We study the predictability of return volatility with MIDAS regressions. Our approach allows us to compare forecasting models with different measures of volatility, frequencies, and lag lengths. While the main focus of this paper is volatility forecasting, it is clear that the MIDAS framework is general in nature and can find a good use in any empirical investigation that involves data sampled at different frequencies. Simplicity, robustness, and parsimony are three of its main attributes.

We report several intriguing findings regarding the predictability of weekly to monthly realized volatility in equity markets. First, we find that daily realized power outperforms daily realized volatility and that daily and intra-daily absolute returns outperform respectively daily and intra-daily squared returns. This set of results suggests that absolute returns are very successful at capturing fluctuations in future return volatility, despite the predominant emphasis in the literature on squared returns. Also, we find that daily ranges are extremely good forecasters of future volatility and are only second to realized power. This last finding is consistent with results in Gallant et al. (1999), Alizadeh et al. (2002) and Engle and Gallo (2003), among others, who use different methods and different data. Finally, we show that the direct use of high-frequency data does not necessarily lead to better volatility forecasts.

Ghysels (2004) also show that realized power has far less measurement noise in comparison with realized volatility and bi-power. They also discuss other appealing features of realized power, such as higher persistence and predictabiity in comparison with bi-power and realized volatility.

In a recent paper, Andersen,

Our paper leaves one unanswered issue: Why is realized power such a good predictor of future volatility? While there still isn't a satisfactory answer to this question, several recent papers have made considerable progress. First, there is now an elegant asymptotic theory that was developed in a set of papers by Barndorff- Nielsen and Shephard (2003b, 2004) showing how realized power, and a measure called bi-power, exclude the jump component of increments in quadratic variation. Building on the Barndorff-Nielsen and Shephard analysis, Forsberg and Ghysels (2004) study the theoretical properties of realized power in the context of continuous time stochastic volatility models. They confirm the intuition that realized power is persistent process more closely related to the long run component of volatility, which explains its success as a regressor in the context of MIDAS regressions.

Bollerslev and Diebold (2003)

We have kept the mixed data sampling regressions as simple as possible in the interest of clarity and conciseness. However, there are a host of issues, such as asymmetries.

4.FUTURE ENHANCEMENT

There are several possible future enhancements that can be considered for a realized volatility prediction project in the stock market. Here are some ideas:

Incorporating additional data sources:

One potential enhancement is to incorporate additional data sources into the analysis, such as sentiment analysis of news articles, social media sentiment, or macroeconomic indicators. These data sources could provide additional insights into market trends and help to improve the accuracy of the prediction

Feature engineering:

Another potential enhancement is to improve the feature engineering process, which involves selecting and transforming the relevant features for the model. This could involve exploring new features or combinations of features, or using advanced feature selection techniques to identify the most important features.

Model selection and hyperparameter tuning:

Another potential enhancement is to explore alternative modeling techniques or algorithms and perform more extensive hyperparameter tuning to optimize the model's performance. This could involve using ensemble models or deep learning techniques to improve the accuracy of the prediction.

Real-time prediction:

Another potential enhancement is to improve the speed and efficiency of the prediction process to enable real-time prediction of realized volatility. This could involve implementing more efficient algorithms, using cloud-based computing resources, or optimizing the code for parallel processing.

Integration with trading platforms:

Another potential enhancement is to integrate the model with existing trading platforms or APIs, allowing traders to use the predictions to inform their trading decisions. This could involve developing custom trading algorithms or implementing a real-time trading dashboard that displays the prEdictions and other relevant market data.

5. REFERENCES

1. A¨ıt-Sahalia, Y., Mykland, P. A. and Zhang, L. (2003), "How often to sample a continuous-time 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 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.
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. Baillie, R.T., T. Bollerslev and H. O. Mikkelsen (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity" Journal of Econometrics, 74, 3-30.