Modeling and Forecasting Asset Return Volatility by Support Vector Machine and Heterogeneous Autoregressive Model.

By

Md. Ashraful Islam Khan

Abstract

Volatility, the standard deviation of the continuously compounded returns of a financial instrument over a specific time horizon. According to Andersen, Bollerslev, and Diebold, (2003), "*Volatility is central to asset pricing, asset allocation and risk management*". It is simply a measure of the degree of price movement in a stock, futures contract or any other market. More broadly, it refers to the degree of (typically short-term) unpredictable change over time of that financial instrument.

Volatility is both the boon and bane of all traders, you can't live with it and you can't really trade without it. If the markets simply rose steadily upward every day, investing would be a much easier without risk. But in reality, the market is volatile, with ups and downs that affect our investments. Understanding volatility is, therefore, important to develop an investment strategy that manages the inherent risk associated with investing. Actually volatility is often used to quantify the risk of the instrument over a specific time horizon. Most of the financial researchers are mainly concerned with modeling and forecasting volatility in asset returns to quantify the risk of financial instruments over a particular time period so that the risk manager and practitioners can realize whether their portfolio will decline in the future and they may want to cell it before it becomes too volatile. Thus, volatility plays the key roles in the theory and applications of asset pricing, optimal portfolio allocation, and risk management.

It is now widely accepted that expected returns, volatility, and broader financial risk measures all vary over time. Researches on time varying volatility using the time series models have been active ever since Angle (1982) introduced the ARCH model. Corsi et al. (2001) and Corsi (2009) proposed the Heterogeneous Autoregressive Realized Volatility (HAR-RV) model as an alternative to the ARFIMA model, based on the HARCH (Heterogeneous ARCH) model of Müller et al. (1997b) and this model has quickly become popular for modeling the dynamics of RV and other related volatility measures due to its ease estimation and extendability of the baseline model. Another recent development in the RV literature is the approach due to Barndorff-Nielsen and Shephard (2004, 2006), Andersen et al. (2003, 2007) of decomposing the RV into the contribution of continuous sample path variation and that of jumps. Extending the theory of quadratic variation of semimartingales, Barndorff-Nielsen et al. (2006) provided an asymptotic statistical foundation for this decomposition procedure under very general conditions.

It is well observed that if the underlying assumptions hold properly then the HAR-RV models with Ordinary Least Squares (OLS) technique perform best. But in many real situations those assumptions may not hold and that may mislead the whole inference procedure. Tukey (1960) points out that a tacit hope in ignoring deviations from the ideal model was that they would not matter; that statistical procedures which were optimal under the strict model would still be approximately optimal under the approximate model. Unfortunately, it turned out that this hope is often drastically wrong; even mild deviations often have much larger effects than were anticipated by most statisticians. So a check in model adequacy can often be essential in analyzing data and using the model.

This study introduces a nonlinear hybrid model, Support Vector Heterogeneous Autoregressive Realized Volatility (SVM-HAR-RV) model, and aims to compare the forecasting performance of this model with the classical HAR-RV model in the field of financial study. The hybrid models are now frequently used in the financial literature. In Chapter one, I have reviewed the application of hybrid models in financial study.

The Support Vector Machine (SVM) is an efficient nonlinear semi-parametric approach, introduced by Vepnik (1995), that guarantees to obtain globally optimal solution (see, e.g., Cristianini and Shawe-Taylor, 2000), which solves the problems of multiple local optima in which the neural network usually get trapped into. Neural Network (NN) is also another powerful tool for prediction problems due to its best ability to estimate any function arbitrary with no priori assumption on data property (see, e.g., Haykin 1999). This study is designed to apply the Support Vector Machine approach to different Heterogeneous Autoregressive Realized Volatility (HAR-RV) models, named SVM-HAR-RV models, to empirically forecast the daily realized volatility of the Nikkei 225 index.

In this study, I first estimate the Realized Volatility (RV) and its different measures of Nikkei 225 Stock Index. To mitigate the market microstructure noise, this study considers the optimal intraday sample. The average daily optimal sampling frequency is observed 7.828 ($\approx$8), i,e $\frac{1}{\Delta} = 8 \Rightarrow \Delta = 33.75$. Therefore, 30-minunite returns are considered as optimally sampled returns here along with 5-minute, 10-minute and 15-minute intraday returns.

From the summary statistics we observe that the unconditional distribution of the daily return series is negatively skewed but highly significantly nonnormal with high positive kurtosis. The LB statistics also indicate that the series is significantly serially correlated. We also observe that the log transformation of RV's brings down the sample skewness and kurtosis values for all series but still significantly nonnormal. The value of skewness and kurtosis are close to zero (0.030) and three (3.719) for 5-minute intraday returns while (0.059 and 3.574) for 10-minute, (0.117, 3.615) for 15-minute and (0.087, 3.635) for optimally sampled intraday returns. All the transformed series remain highly significantly non-normally distributed and highly significantly serially correlated. It is

observed from Table 2.1 that the average jumps and MSNR-Jumps are (1.471 and 1.504) for 5-minute, (1.3791 and 1.427) for 10-minute, (1.434 and 1.448) for 15-minute and (1.600 and 1.508) for optimally sampled intraday returns with positive minimum values respectively that implies more than one jump occurred in every single days while from Table 2.2, we observe the minimum value zero for the shrinkage and microstructure noise robust estimator of significance jumps for all four series.

The forecasting performance of the HAR-RV and SVM-HAR-RV models are compared in Chapter three. Where I first fit the Heterogenious Autoregressive Realized Volatility (HAR-RV) class models by Ordinary Least Squares (OLS) technique and then by SVM. As mentioned earlier, HAR-RV models fitted by SVM method will be known as SVM-HAR-RV models. A comparative study is made in the context of daily volatility forecasting among HAR-RV and SVM-HAR-RV class of Models. It is observed that the SVM-HAR-RV models produce better forecasting ability compare to the classical HAR-RV models. The in sample forecasting performance of SVM-HAR-RV models is outstanding. The out-of-sample forecasting performance is also better than HAR-RV models.

Giot and Laurent (2004) and Clements et al. (2008). Giot and Laurent (2004) compare an ARCH-type model and a model using realized volatility in terms of forecasts of Value-at-Risk. They show better performance of Skewed student APARCH model. Clements et al. (2008) narrow their study to focus exclusively on models based on realized volatility and show comparatively better performance of HAR model for quantile forecasts. Following Giot and Laurent (2004) and Clements et al. (2008), I extend the study to modeling and forecasting one day ahead Value-at-Risk (VaR). Since volatility forecast is the key input to VaR forecasting, therefore, I consider the volatility forecasts series of HAR-RV and SVM-HAR-RV models and compute the conditional realized volatility. An ARCH type model is then applied to these two classes of conditional realized volatility series. An AR (2) model is fitted for the conditional mean (the lag is computed using minimum AIK value) and five different models for the conditional variance. Both classes of models pass

the Kupiec's Likelihood Ratio and Dynamic Quantile tests. From chapter four we observe that the GJR-GARCH and EGARCH specification with Skewed Student innovation distribution produce better forecast. It is observed that the SVM-HAR-RV-ARCH type models produce better VaR forecast compared to HAR-RV-ARCH type models. Better performance observed when 10-minute and 15-minute intraday returns are used to estimate daily realized volatility.

Engle et al. (1990) and Bandi et al. (2008). Engle et al. (1990) compare the performance of four models, i.e., on the basis of an artificial ("hypothetical" in their terminology) option market for one-day options and $1 shares of NYSE priced from the moving average of squared daily returns, ordinary least squares: the variance rate is estimated from the standard error on an AR(1) on the daily rate, ARMA(1,1) on squared errors and GARCH(1,1). Following Engle et al. (1990), Bandi et al. (2008) combine 14 different methods of realized volatility estimation as a measure of actual volatility and compare their performance in the context of profits from option pricing and trading when realistic market microstructure contaminations play a role for one-day options and $1 shares of S&P500 index. Following Engle et al. (1990) and Bandi et al. (2008) I finally extend the study to option pricing and trading context to compare the performance of SVM-HAR-RV models with the classical HAR-RV models in the context of profits from option pricing and trading. Using forecasts from these two classes of models for realized volatility with different intraday returns, agents price short-term options on Nikkei 225 index before trading each other at average prices. A comparative study between models is made to examine their ability in option pricing and trading context on the basis of higher Sharpe rations and average profits. It is observed from chapter five that the SVM-HAR-RV model with Polynomial kernel function produces better performances followed by SVM-HAR-RV model with Laplacian kernel function.

Examine the whole study, it can be concluded that the SVM-HAR-RV models have higher ability to capture the volatility dynamics. It can also be concluded that the hybrid models could be better choice to handle the financial time series.