Long Memory Realized Volatility Modeling Executive Summary

Uri Miron
(Advisor: Ron DiRusso, Hyman Beck & Co.)

Volatility modeling has been an extremely active area of research in quantitative finance over the last quarter century. The last decade has seen an explosion of two related trends in volatility modeling, the use of high-frequency data for ex-post estimation of *realized volatility* (RV), and the introduction of a class of *long memory* models, inspired by the slow decay of RV autocorrelations.

In this work, we explore the long memory Heterogenous Autoregressive Model of Realized Volatility (HAR-RV) of Corsi (2004), and investigate whether incorporating exogenous factors, such as macroeconomic indicators and implied volatility, improves its forecasting power. HAR-RV models realized volatility as an autoregressive process driven by past realized volatilities over different aggregation periods, corresponding to the holding periods typical of different types of investors.

The use of long memory realized volatility models requires robust and accurate estimation of ex-post realized volatility. Since high frequency trade (and quote) prices contain significant market microstructure frictions, these frictions must be properly accounted for to estimate the "true" returns volatility. We utilize the recently developed Two Scales Realized Volatility (TSRV) estimator of Ait-Sahalia et. al (2006) in this study. TSRV explicitly accounts for the strong microstructure bias in high frequency data, thus allowing the use of higher frequency data than other RV estimators. It further uses subsampling for increased efficiency. For this study, estimators were calculated using prices sampled at a five second frequency, whereas with the "traditional" RV estimator, data is typically sampled at 5-30 minute frequencies in order to avoid excessive microstructure effects.

HAR-RV forecasts were calculated for two exchange traded funds tracking broad U.S. equity indices, SPY (S&P 500) and QQQ (Nasdaq 100), using TAQ trade data from March 1999 through July 2008. All forecasts are out of sample, and recalculated daily via linear regression on a rolling window of 750 days. Mincer-Zarnowitz regressions of forecasted on actual realized volatilities reveal that the predictive power of the model is quite high, which is also apparent from a graphical inspection of the corresponding plots. Forecasting periods of 1, 5, 10, 15 and 20 days were examined, with the best performance typically observed at the 5 or 10 day level. The results are robust to different sub-periods within the sample, and it appears that the forecasts adjust quite quickly to changes from low to high volatility periods.

Out of sample testing of forecasts incorporating exogenous factors along with HAR-RV predictions show reduced forecasting power for all factors. In sample regressions, on the other hand, result in some factors with weightings that are non-zero with statistically significant confidence intervals. This would seem to suggest that these factors have relevance, but the information they represent is already well expressed in the high-frequency prices.

In practical terms, the success of the HAR-RV model makes improving on it a formidable challenge, and the high number and low frequency of potential exogenous factors makes data mining a concern (though hardly one that is unique to volatility modeling).