

Forecasting Realized Volatility Using Subsample Averaging*

Huiyu Huang¹, Tae-Hwy Lee²

¹GMO Emerging Markets Equity, Berkeley, USA

²Department of Economics, University of California, Riverside, USA

Email: talee@ucr.edu

Received March 1, 2013; revised April 1, 2013; accepted April 8, 2013

Copyright © 2013 Huiyu Huang, Tae-Hwy Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

When the observed price process is the true underlying price process plus microstructure noise, it is known that realized volatility (RV) estimates will be overwhelmed by the noise when the sampling frequency approaches infinity. Therefore, it may be optimal to sample less frequently, and averaging the less frequently sampled subsamples can improve estimation for quadratic variation. In this paper, we extend this idea to forecasting daily realized volatility. While subsample averaging has been proposed and used in estimating RV, this paper is the first that uses subsample averaging for forecasting RV. The subsample averaging method we examine incorporates the high frequency data in different levels of systematic sampling. It first pools the high frequency data into several subsamples, then generates forecasts from each subsample, and then combines these forecasts. We find that in daily S&P 500 return realized volatility forecasts, subsample averaging generates better forecasts than those using only one subsample.

Keywords: Subsample Averaging; Forecast Combination; High-Frequency Data; Realized Volatility; ARFIMA Model; HAR Model

1. Introduction

The rich dynamics in ultra-high-frequency financial data may be captured to improve estimation of quadratic variation (integrated variance) or forecasting volatility. There is a considerable amount of literature addressing the estimation issue, whereas little work has been done on forecasting. This paper contributes to the forecasting issue in the high-frequency data literature by examining whether it pays to incorporate the intraday data and, more importantly, how to incorporate the high-frequency information to achieve better performance for forecasting daily return volatility. Further, we seek to link the estimation and forecasting aspects by adopting and tailoring methods proposed for estimation of quadratic variation to forecasting.

The existing literature on high-frequency data, especially that devoted to the realized volatility issue, can be grouped roughly into two areas: estimating quadratic variation using realized volatility (RV) and forecasting volatility using RV. References [1,2] establish that RV,

defined as the sum of squared intraday returns of small intervals, is an asymptotically unbiased estimator of the unobserved quadratic variation as the interval length approaches zero.

However, in the presence of market microstructure noise, this helpful property of RV is contaminated. More recent works investigating this issue include [3-8]. When the observed price process is the true underlying price process plus microstructure noise, it is shown that RV will be overwhelmed by the noise and explodes when the sampling frequency approaches infinity. Therefore, it may be optimal to sample less frequently than is the case in the absence of noise. References [6,8] establish through a subsampling scheme improved estimators for quadratic variation. The original subsampling idea can be traced back to [9], where for the first time an unbiased data-driven estimator of volatility and a subsample averaging volatility estimator are proposed. The bias-adjusted estimator of [6] based on the subsample averaging method is able to eventually push the estimation bias to zero. Reference [8] shows that subsampling is highly advantageous for RV estimators based on discontinuous kernels.

Besides the use of high-frequency information in vola-

*The opinions expressed in this paper are those of the authors and do not necessarily reflect those of Grantham, Mayo, Van Otterloo and Company LLC.

tility estimation, volatility forecasting using high-frequency information has been addressed as well. Reference [10] represents an approach to forecasting volatility using RV. The model they propose is a fractional integrated AR model, ARFI(p, d), for logarithmic RV obtained from foreign exchange rates data of 30-minute frequency, and demonstrates the superior predictive power of their model. Noticeably, the sampling frequency used in [1] is, however, 5 minutes when they analyze the distributional properties of RV.

Under the presence of microstructure noise, it may be optimal to sample less frequently than is the case in the absence of noise. As shown by [6,8,9], subsampling schemes improve estimators for quadratic variation. When it comes to forecasting RV, if the highest frequency returns are not necessarily the optimal frequency, the subsampling of the available highest frequency data will lead to several systematically sampled subsamples. In this paper we attempt to answer two questions: (1) can we use these subsamples to improve the out-of-sample forecasting? and (2) how do we incorporate the subsample information when producing forecasts of daily realized volatility? We find that as long as the available highest-frequency data are not at the optimal frequency, we can improve prediction of RV out-of-sample through simple averaging of forecasts produced from subsamples.

The paper is organized as follows: Section 2 describes the data and subsamples. Section 3 discusses two models for forecasting daily realized volatility by subsample averaging. Section 4 presents their out-of-sample relative performances. Section 5 concludes.

2. Data and Subsampling

The data we use consists of S&P 500 index values at 5-minute intervals recorded between 9:30 a.m. and 4:00 p.m. (total 390 minutes a day) from June 9, 1997 to May 30, 2003, a total of 1501 days and 117,078 observations¹. In cleaning the data, those periods of market closings are treated as no variation in index values, thus there exist 78 ticks each trading day. We can construct data-driven volatility, for instance, realized volatility, from this 5-minute high-frequency data.

Let Δ be the interval in minutes over which the returns are computed. The Δ -minute return is the log-difference of two consecutive index values over Δ -minutes, multiplied by 100

$$r_{t-1+i\Delta, t-1+(i-1)\Delta} \equiv p_{t-1+i\Delta} - p_{t-1+(i-1)\Delta}, \quad (1)$$

where p_t denotes the logarithm of the S&P 500 index value. Even if the interval Δ can be as small as 5 min-

utes given the data available to us, we choose larger intervals such as $\Delta = 15, 30, 60$ (minutes), in order to construct subsamples. When $\Delta = 15$ we can construct three sets (subsamples) of Δ -minute return series. When $\Delta = 30$ we can construct six sets (subsamples) of Δ -minute return series. When $\Delta = 60$ we can construct 12 sets (subsamples) of 60-minute return series by sampling every 12th observation of 5-minute returns.

The RV is calculated by the sum of squared Δ -minute returns within a day. Define realized volatility of day t as

$$RV_t^{(\Delta)} = \sum_{i=1}^m r_{t-1+i\Delta, t-1+(i-1)\Delta}^2, \quad (2)$$

where

$$m = 390/\Delta \quad (3)$$

is the number of Δ -minute returns per day.

To construct a subsample, we consider an interval $\Delta > 5$. The subsampled return data has the longer time interval Δ than 5 minutes, within which

$$K = \Delta/5 \quad (4)$$

number of subsamples of the Δ -minute returns are observed. We consider $\Delta = 15, 30, 60$, each producing a different number $K = 3, 6, 12$ of subsamples.

For example, the intraday returns for $\Delta = 15$ produce daily realized volatility $RV_t^{(15)}$, computed from (2) using $m (= 26)$ 15-minute returns. As there are three 5-minute intervals within 15 minutes, we have $K = 15/5 = 3$ subsamples, producing three subsample RV forecasts for a day.

3. Volatility Forecast Models

Based on the subsample averaging methodology in [6] and [9], we demonstrate the benefit of subsample averaging in out-of-sample forecasting of daily volatility. We aim to check if the predictive ability of an RV forecasting model can be improved by averaging forecasts generated using the same model but from different subsamples.

In comparing the predictive ability of the subsample averages with the benchmark forecast without subsample averaging, the key issues are: (1) the volatility proxy to compare forecasts with; (2) the loss function for forecast evaluation; and (3) the forecasting model. For the daily volatility proxy we use the daily RV computed from 5-minute returns. For the loss function for forecast evaluation we use the mean squared forecast errors (MSFE). Regarding the forecasting model, we consider two models, ARFI model of [10] and HAR model of [11].

3.1. ARFI Model

To generate a forecast for tomorrow's (one day ahead)

¹We are grateful to George Jiang who generously shared this high-frequency intraday data with us. The data are extracted from the contemporaneous index levels recorded with the quotes of SPX options from the CBOE.

conditional realized volatility, we estimate the fractionally integrated autoregressive model, ARFI(p,d) for

$$X_t \equiv \ln(RV_t^{(\Delta)}) \tag{5}$$

for each of $\Delta = 15, 30, 60$. Specifically, the model can be written as follows:

$$\phi_p(L)(1-L)^d(X_t - \mu) = \epsilon_t \tag{6}$$

where $\phi_p(L)$ is the p th order AR lag polynomial. We fix $p = 1$ and $\phi_p(L) = 1 - \phi L$. For each Δ , we estimate μ, ϕ , and d , the long-memory parameter by using the method of [10,12].

The presence of microstructure noise prevents us from sampling too frequently (with Δ too small) when calculating RV. See [3]. If $RV_t^{(k)(\Delta)}$ (with $\Delta > 5$) series are less noisy and more accurately estimated than $RV_t^{(5)}$, where $k = 1, \dots, K$ are indices of subsamples, then using these likely leads to a better forecast $\widehat{RV}_{T+1}^{(k)(\Delta)}$ than $\widehat{RV}_{T+1}^{(5)}$. However, abandoning the finer information seems not very sensible. Thus, a possible way to balance this is the subsample averaging approach, which uses all K subsamples, meaning no intra-day information is tossed out, and then taking the average over $k = 1, \dots, K$ to produce a final forecast. At the same time, the contaminating effect of microstructure noise in the stage of computing RV is moderated through this subsampling procedure (using $\Delta > 5$ instead of $\Delta = 5$), and hence it is possible to achieve better performance in forecasting RV.

The subsample averaging method is as follows. First, for a given Δ , we estimate the ARFI(p,d) model for each of K subsamples generated by systematically sampling the original 5-minute index data every K steps:

$$\begin{aligned} \phi_p^{(k)}(L)(1-L)^{d(k)}(X_t^{(k)} - \mu^{(k)}) &= \epsilon_t^{(k)}, \\ k &= 1, \dots, K \end{aligned} \tag{7}$$

where $X_t^{(k)}$ denotes k th daily logarithm of RV in day t computed using the k th subsample. Following [10] we use ARFI(p,d) model on the natural logarithm transformation of RV, $X_t^{(k)} \equiv \ln(RV_t^{(k)(\Delta)})$, and then exponentiated back to RV, $RV_t^{(k)(\Delta)} = \exp(X_t^{(k)})$. Running the model on each subsample we obtain K number of RV forecasts $\widehat{RV}_{T+1}^{(k)(\Delta)}$, $k = 1, \dots, K$. Next, we compute their simple average

$$\overline{RV}_{T+1}^{(\Delta)} = \frac{1}{K} \sum_{k=1}^K \widehat{RV}_{T+1}^{(k)(\Delta)} \tag{8}$$

to obtain the combined forecast (denoted Subsample-Mean in **Table 1**). Similarly, ‘‘Subsample-Median’’ is to take the median of $\widehat{RV}_{T+1}^{(k)(\Delta)}$, $k = 1, \dots, K$.

We compare Subsample-Mean forecast $\overline{RV}_{T+1}^{(\Delta)}$ and Subsample-Median forecast with the benchmark forecast (denoted Benchmark in **Table 1**). The benchmark is computed using only one subsample $\widehat{RV}_{T+1}^{(k)(\Delta)}$ with $k = K$ the last subsample and abandoning the other $K - 1$ subsample information.

3.2. HAR Model

The heterogeneous autoregressive (HAR) model of the realized volatility in [11] is inspired by the heterogeneous market hypothesis and is able to reproduce memory persistence (although not formally a long memory model) as well as many other major features of financial data. The square-root of daily RV

$$Y_t \equiv \sqrt{RV_t^{(\Delta)}}, \tag{9}$$

is assumed to have an AR-type process on past RVs over different intervals of aggregation (daily, weekly, and monthly):

$$Y_{t+1} = c + \beta^D Y_t + \beta^W Y_t^W + \beta^M Y_t^M + e_{t+1}, \tag{10}$$

where

$$Y_t^W = \frac{1}{5}(Y_{t-1} + Y_{t-2} + \dots + Y_{t-5}) \tag{11}$$

is the weekly aggregated RV, and

$$Y_t^M = \frac{1}{20}(Y_{t-1} + Y_{t-2} + \dots + Y_{t-20}) \tag{12}$$

is the monthly aggregation.

As illustrated in Section 3.1 above, we consider $\Delta = 15, 30, 60$. The subsample averaging method for the HAR model works in exactly the same way as the ARFI model. First, we estimate the HAR model for each of K subsamples generated by systematically sampling the original 5-minute index data every K steps. We take K subsamples of the Δ -minute returns data when computing $Y_{t+1}^{(k)}$ ($k = 1, \dots, K$), apply the HAR model to each subsample to generate the subsample forecast $Y_{T+1}^{(k)}$ at day T for the next day $T + 1$, and then use the simple average or median to obtain the subsample averaging forecasts. The only difference here as compared to the ARFI case is the model used to produce forecasts. Note that here we use the square root of RV, as in [11], thus MSFE values of Y will appear in different scale than those of ARFI(p,d), when we show out-of-sample results in the next section.

4. Results

Evaluating the volatility forecast involves the selection of a volatility proxy given that the true underlying volatility is latent and subject to a researcher’s own definition.

Recent papers addressing this include [13,14]. Generally, the suggested volatility proxy is the realized volatility in some particular form, coupled with an MSFE evaluation criterion.

Our empirical work involves the comparison of forecasting performances of subsample averaging (mean and median) of the forecasts using all K Δ -minute return subsamples with the benchmark model using only one Δ -minute return subsample. We consider three cases (A, B, and C) of subsampling for $\Delta = 15, 30, 60$ in **Tables 1** and **2**.

In **Table 1** where the ARFI(p, d) model is used to forecast RV, we find that subsample averaging (mean and median) improves upon the benchmark in all three cases for Subsample-Mean and in two cases for Subsample-Median, quite substantially for Case A. In **Table 2** where the HAR model is used, Subsample-Mean and Subsample-Median are substantially better than the benchmark for all three subsampling cases.

5. Conclusion

We propose a forecasting methodology that uses subsample averaging to forecast daily realized volatility.

Table 1. Using ARFI model for RV.

Case	Δ	m	K	Benchmark	Subsample-Mean	Subsample-Median
A	15	26	3	1.3119	1.0581	1.0561
B	30	13	6	1.1854	1.1610	1.1774
C	60	6.5	12	1.2828	1.2392	1.2859

Notes: This table compares the performance of the ARFI model for the daily realized volatility (RV) forecast without and with subsampling average. We use the ARFI(p, d) model proposed by [10] where $p = 1$ and d is estimated by [12] method dynamically for each case. The out-of-sample size is $P = 500$ Days, and the in-sample size is $R = 1000$ Days. We report the out-of-sample MSFE with three different subsampling cases. Δ -minute returns have m observations per day. Subsample-Mean or Subsample-Median is taken over the K subsample forecasts. The benchmark uses only the last subsample forecast. Bolded term indicates the better-than-benchmark case.

Table 2. Using HAR model for Sqrt(RV).

Case	Δ	m	K	Benchmark	Subsample-Mean	Subsample-Median
A	15	26	3	0.1512	0.1043	0.0971
B	30	13	6	0.1370	0.0752	0.0738
C	60	6.5	12	0.1151	0.0670	0.0659

Notes: This table compares the performance of the HAR model for the square root of the daily realized volatility (RV) forecast without and with subsampling average. We use the HAR model proposed by [11], which is applied to the square root of RV and estimated dynamically for each case. The out-of-sample size is $P = 500$ Days, and the in-sample size is $R = 1000$ Days. We report the out-of-sample MSFE with three different subsampling cases. Δ -minute returns have m observations per day. Subsample-Mean or Subsample-Median is taken over the K subsample forecasts. The benchmark uses only the last subsample forecast. Bolded term indicates the better-than-benchmark case.

While subsample averaging has been proposed and used in estimating RV, this paper is the first to use subsample averaging for forecasting RV. In this paper, we show that subsample averaging, which was originally suggested to overcome the bias in estimating quadratic variation under the presence of market microstructure noise, can also help forecast RV out-of-sample. In an application of S&P 500 index daily volatility forecasting, using two classical forecasting models for RV, we find that the subsample averaging forecast generally and substantially improves upon forecasts using only one subsample without averaging over all subsamples. We expect that subsample averaging method can enhance even more the forecast ability of RV when much higher-frequency data (15-seconds, for instance) are available.

REFERENCES

- [1] T. G. Andersen, T. Bollerslev, F. X. Diebold and P. La-bys, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, Vol. 96, No. 453, 2001, pp. 42-55. <http://dx.doi.org/10.1198/016214501750332965>
- [2] O. E. Barndorff-Nielsen and N. Shephard, "Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society Series B*, Vol. 64, No. 2, 2002, pp. 253-280.
- [3] Y. Aït-Sahalia, P. A. Mykland and L. Zhang, "How Often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, Vol. 18, No. 2, 2005, pp. 351-416. <http://dx.doi.org/10.1093/rfs/hhi016>
- [4] F. M. Bandi and J. R. Russell, "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, Vol. 75, No. 2, 2008, pp. 339-369. <http://dx.doi.org/10.1111/j.1467-937X.2008.00474.x>
- [5] P. R. Hansen and A. Lunde, "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, Vol. 24, No. 2, 2006, pp. 127-218. <http://dx.doi.org/10.1198/073500106000000071>
- [6] L. Zhang, P. A. Mykland and Y. Aït-Sahalia, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, Vol. 100, No. 472, 2005, pp. 1394-1411. <http://dx.doi.org/10.1198/016214505000000169>
- [7] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, "Designing Realised Kernels to Measure the Expost Variation of Equity Prices in the Presence of Noise," *Econometrica*, Vol. 76, No. 6, 2008, pp. 1481-1536. <http://dx.doi.org/10.3982/ECTA6495>
- [8] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, "Subsampling Realised Kernels," *Journal of Econometrics*, Vol. 160, No. 1, 2011, pp. 204-219. <http://dx.doi.org/10.1016/j.jeconom.2010.03.031>
- [9] B. Zhou, "High-Frequency Data and Volatility in For-

- eign-Exchange Rates,” *Journal of Business and Economic Statistics*, Vol. 14, No. 1, 1996, pp. 45-52.
- [10] T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys, “Modeling and Forecasting Realized Volatility,” *Econometrica*, Vol. 71, No. 2, 2003, pp. 579-625.
<http://dx.doi.org/10.1111/1468-0262.00418>
- [11] F. Corsi, “A Simple Approximate Long Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, Vol. 7, No. 2, 2009, pp. 174-196.
<http://dx.doi.org/10.1093/jjfinec/nbp001>
- [12] J. Geweke and S. Porter-Hudak, “The Estimation and Application of Long Memory Time Series Models,” *Journal of Time Series Analysis*, Vol. 4, No. 4, 1983, pp. 221-238.
<http://dx.doi.org/10.1111/j.1467-9892.1983.tb00371.x>
- [13] A. J. Patton, “Volatility Forecast Comparison Using Imperfect Volatility Proxies,” *Journal of Econometrics*, Vol. 160, No. 1, 2011, pp. 246-256.
<http://dx.doi.org/10.1016/j.jeconom.2010.03.034>
- [14] P. R. Hansen and A. Lunde, “Consistent Ranking of Volatility Models,” *Journal of Econometrics*, Vol. 131, No. 1, 2006, pp. 97-121.
<http://dx.doi.org/10.1016/j.jeconom.2005.01.005>