

Forecasting Realized Volatility Using Subsample Averaging*

Huiyu Huang[†]
Grantham, Mayo, Van Otterloo
and Company LLC

Tae-Hwy Lee[‡]
University of California, Riverside

February 2013

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 the subsample-averaging has been proposed and used in estimating RV, this paper is the first that uses the 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, that generates forecasts from each subsample, and then combine these forecasts. We find that, in daily S&P 500 return RV forecasts, subsample-averaging generates better forecasts than those using only one subsample without averaging over all subsamples.

Key Words: Subsample averaging. Forecast combination. High-frequency data. Realized volatility. ARFIMA model. HAR model.

JEL Classification Codes: C53, C58, G17

*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.

[†]GMO Emerging Markets, 2150 Shattuck Ave, Suite 900, Berkeley, CA 94704, U.S.A. Phone: +1 (510) 809-4594. Email: evelyn.huang@gmo.com.

[‡]Corresponding Author. Department of Economics, University of California, Riverside, CA 92521-0427, U.S.A. Phone: +1 (951) 827-1509. Fax: +1 (951) 827-5685. E-mail: taelee@ucr.edu.

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 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.

Andersen, Bollerslev, Diebold, and Labys (ABDL 2001) and Barndorff-Nielsen and Shephard (2002) establish that realized volatility (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, such nice property of RV is contaminated. Recent works investigating this issue include Aït-Sahalia, Mykland and Zhang (AMZ 2005), Bandi and Russell (2008), Hansen and Lunde (2006a), Zhang, Mykland and Aït-Sahalia (ZMA 2005), and Barndorff-Nielsen, Hansen, Lunde and Shephard (BNHLS 2008, 2011). 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 the case in the absence of noise. ZMA (2005) and BNHLS (2011) establish through a subsampling scheme improved estimators for quadratic variation. The original subsampling idea can be traced back to Zhou (1996), where an unbiased data-driven estimator of volatility and a subsample-averaging volatility estimator are proposed. The bias-adjusted estimator of ZMA (2005) based on the subsample-averaging method is able to eventually push the estimation bias to zero. BNHLS (2011) show that subsampling is highly advantageous for RV estimators based on discontinuous kernels.

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 address the following questions. Can we use these

subsamples to improve forecasting RV? How do we incorporate these subsamples? We find that as long as the available highest frequency data is not at the optimal frequency, we can improve prediction of RV out-of-sample, through simple averaging of forecasts produced from sub-samples.

The paper is organized as follows. Section 2 describes the data and subsampling. 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 Subsampling

The data we use consists of S&P 500 index values at 5-minute intervals recorded in 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 1,501 days and 117,078 observations.¹ In cleaning the data, those periods of market closings are treated as no variation in index values, thus there exists 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 minutes 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 returns series. When $\Delta = 60$ we can construct 12 sets (subsamples) of 60-minute returns series by sampling every 12th observations 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

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

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.

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

$$K = \Delta/5 \tag{4}$$

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 Zhou (1996) and ZMA (2005), 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: (i) the volatility proxy to compare forecasts with, (ii) the loss function for forecast evaluation, and (iii) the forecasting model. For the daily volatility proxy we use the daily RV computed from 5-minute returns. About 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 ABDL (2003) and HAR model of Corsi (2009).

3.1 ARFI Model

To generate a forecast for tomorrow’s (one day ahead) conditional realized volatility, we estimate the fractionally integrated autoregressive model, ARFI(p, d) for

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

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 Geweke and Porter-Hudak (GPH 1984) as in ABDL (2003).

The presence of microstructure noise prevents us from sampling too frequently (with Δ too small) when calculating RV. See AMZ (2005). 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 therefore 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:

$$\phi_p^{(k)}(L)(1 - L)^{d(k)}(X_t^{(k)} - \mu^{(k)}) = \epsilon_t^{(k)}, \quad k = 1, \dots, K \quad (7)$$

where $X_t^{(k)}$ denotes k th daily logarithm of RV in day t computed using the k th subsample. Following ABDL (2003) 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

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

to obtain the combined forecast. We compare Subsample-Averaging forecast $\overline{RV}_{T+1}^{(\Delta)}$ with the benchmark forecast, which 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 Corsi (2009) 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

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

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

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

where

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

is the weekly aggregated RV, and

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

is the monthly aggregation.

Same as in Section 3.1 illustrated above, we consider $\Delta = 15, 30, 60$. The subsample-averaging method for the HAR model goes in exactly the same way as for 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 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 Corsi (2009), thus MSFE values of Y will appear in different scale than that of $ARFI(p, d)$, when we show out-of-sample results in the next section.

4 Results

Evaluating volatility forecast involves the selection of volatility proxy given that the true underlying volatility is latent and subject to a researcher's own definition. Recent papers addressing this

include Patton (2011) and Hansen and Lunde (2006b). Generally, the suggested volatility proxy is the realized volatility in some particular form, coupled with a MSFE evaluation criterion.

Our empirical work involves the comparison of forecasting performances of Subsample-averaging 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. 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-Averaging is taken over the K subsample forecasts. Benchmark uses only the last subsample forecast.

In Table 1 where the ARFI(p, d) model is used to forecast RV, we find that Subsample-averaging improves upon the benchmark in all three cases for Subsample-Averaging, quite substantially for Case A. In Table 2 where the HAR model is used, Subsample-Averaging is substantially better than the benchmark for all three cases.

5 Conclusions

We propose a forecasting methodology that uses subsample-averaging to forecast daily realized volatility. While the subsample-averaging has been proposed and used in estimating RV, this paper is the first that uses the subsample-averaging for forecasting RV. In this paper, we show that the 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 substantially improves upon forecasts from 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

- Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005), "How Often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies* 18, 351-416.
- Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* 96, 42-55.
- Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica* 71, 579-625.
- Bandi, F.M. and Russell, J.R. (2008), "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies* 75, 339-369
- Barndorff-Nielsen, O.E. and Shephard, N. (2002), "Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society Series B* 64, 853-223.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., and Shephard, N. (2008), "Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise," *Econometrica* 76(6), 1481-1536
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., and Shephard, N. (2011), "Subsampling Realised Kernels," *Journal of Econometrics* 160, 204-219
- Corsi, F. (2009), "A Simple Approximate Long Memory Model of Realized Volatility," *Journal of Financial Econometrics* 7, 174-196.
- Geweke, J. and Porter-Hudak, S. (1984), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis* 4, 221-238.
- Hansen, P.R. and Lunde, A. (2006a), "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics* 24, 127-218.
- Hansen, P.R. and Lunde, A. (2006b), "Consistent Ranking of Volatility Models," *Journal of Econometrics* 131, 97-121.
- Patton, A.J. (2011), "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics* 160(1), 246-256.
- Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association* 100, 1394-1411.
- Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business and Economic Statistics* 14, 45-52.

Table 1. Using ARFI model for RV

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

Notes: This table compares the performance of the ARFI model for daily realized volatility (RV) forecast without and with subsampling average. We use ARFI(p, d) model proposed by ABDL (2003) where $p = 1$ and d is estimated by the GPH method dynamically for each case.

Table 2. Using HAR model for Sqrt(RV)

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

Notes: This table compares the performance of the HAR model for the square root of daily realized volatility (RV) forecast without and with subsampling average. We use HAR model proposed by Corsi (2009), which is applied to the square root of RV and estimated dynamically for each case.