International Journal of Economics and Finance; Vol. 6, No. 12; 2014
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

Improve Volatility Forecasting with Realized Semivariance-Evidences from Intra-Day Large Data Sets in Chinese

Lianqian Yin[1,2], Bo Liu[1] & Zhen Du[1]

[1] Finance Department of International Bussiness School, Jinan University, Zhuhai, Guangdong Province, China

[2] Economics and Social Development Research Center of Zhuhai City, Zhuhai, Guangdong Province, China

Correspondence: Lianqian Yin, Finance Department of International Bussiness School, Jinan University, Qianshan road 206#, Zhuhai City, Guangdong Province, Post NO 519070, China. Tel: 86-135-7064-5138. E-mail: lianqian@jnu.edu.cn

Received: September 28, 2014 Accepted: October 15, 2014 Online Published: November 25, 2014

doi:10.5539/ijef.v6n12p64 URL: http://dx.doi.org/10.5539/ijef.v6n12p64

Abstract

Realized semivariance is reported more informative than realized variance. This paper employs a new modeling approach for the realized semivariance inspired by Chou (2005) in order to capture the asymmetry of volatility in financial markets better. With high frequency data from Shanghai stock market in Chinese, the empirical results, which uses four types of volatility proxies including squared daily returns, daily high-low price ranges, realized variance, and realized range consistently, indicate that this model sharpens the forecast power of existing volatility models in terms of GARCH type models. Mincer-Zarnowitz regression and four loss functions are employed for the assessments in out of the sample forecasting.

Keywords: realized semivariance, GARCH, Mincer-Zarnowitz regression, high frequency data

1. Introduction

Volatility has been a traditional tool to measure the risk of the financial market for a long time. It plays a key role in the areas of asset pricing, portfolio allocation, and risk management. In recent years, as transaction data is becoming increasing widely available, great interest has been drawn into the use of high frequency data for measuring and forecasting volatility. This approach is called the realized volatility. One of the advantages to use the new emerging nonparametric volatility approach is that it can fully exploit intraday information and deliver an observable proxy for the volatility. Therefore it makes direct modeling volatility possible and avoids complicated estimation procedures which employ the unobservable volatility approach–the GARCH type and stochastic volatility models.

Barndorff-Nielsen, Kinnebrock and Shephard (2008) introduced a new measure for the variation of asset prices based on high frequency data. It is called realized semivariance (RS) and it is reported more informative than the simple realized variance. Inspired by Chou (2005), the same methodology is adopted in that paper for the realized semivariance to capture the asymmetry in financial markets better.Combined with realized semivariance this modeling approach can sharpen the forecast power of existing volatility models intuitively. It is also confirmed in our empirical study through the comparison of four GARCH-type models for non-negative series, proposed by Engle (2002) and known as Multiplicative Error Model (MEM). We employ Shanghai composite index data of one minute's frequency to obtain our daily and realized volatility estimators. Mincer-Zarnowitz (MZ) regression is a widely accepted method for the model comparison. According to Engle (2005), different volatility proxies contain different information about volatility. Therefore, we use six different volatility proxies of both daily frequency and high frequency as the measure volatility in MZ equation: squared daily returns, absolute daily returns, daily high-low price ranges, realized variance, realized range, and realized bipower variation. They consistently indicate that our modeling approach sharpens the forecast power of non-negative series GARCH type models. Besides MZ equation, we use four loss functions in Hansen and Lunde (2005) as criterions for assessing the forecasting ability of the models. For the one step ahead out of sample tests, we also employ an expanding window estimation procedure to simulate the actual estimation adapted to the data updating process. All in sample, out of sample and the expanding window prediction consistently confirmed our intuition that this modeling approach is able to sharpen the forecast power of non-negative series GARCH type models combined with realized semivariance. The rest of this paper has the following structure. In section 2 we will discuss the theory of realized volatility and semivariance. Section 3 introduces our empirical funds. Section 4 is the model comparisons. Section 5 is the conclusions.

2. Realized Volatility, Realized Semivariance and the Model

Realized variance estimates the ex-post variance of asset prices over a fixed time interval. Since we are going to carry out our empirical analysis based in trading time, we define realized variance as:

$$RV_t = \sum_{i=1}^{I} (P_{t,j+1} - P_{t,j})^2 \tag{1}$$

RV is the sum of squared intraday returns. Although the data arrives into our database at irregular points in time, however according to Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), these irregularly spaced observations can be regarded as being equally spaced observations on a new time-changed process in the same stochastic class. Thus there is no intellectual loss initially considering equally spaced returns. In arbitrage free markets, *P* is often considered to follow a semimartingale process. Then as we have increasing data in one day's time interval *RV* must converge into:

$$RV_t = \sum_{i=1}^{I} (P_{t,j+1} - P_{t,j})^2 \xrightarrow{p} \int_0^1 \sigma_s^2 ds \tag{2}$$

Where $P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s ds$, μ is a locally bounded predictable drift process and σ is a cadlag volatility process, which adapted to some common filtration Ft. Barndorff-Nielsen, Kinnebrock and Shephard (2008) introduced a new measure of variation called realized semivariance. This kind of estimator is solely determined by the single side (upward and downward) moves in high frequency asset prices defined as:

$$RS_t^- = \sum_{i=1}^{I} (P_{t,j+1} - P_{t,j})^2 1_{P_{t,j+1} - P_{t,j} \le 0} \tag{3}$$

$$RS_t^+ = \sum_{i=1}^{I} (P_{t,j+1} - P_{t,j})^2 1_{P_{t,j+1} - P_{t,j} \ge 0} \tag{4}$$

Where 1_P is the indicator function taking the value 1 of the argument is true and 0 otherwise. If P is a semi martingale without jumps as $P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s^2 ds$, then there would be no difference between RS_t^- and RS_t^+. They can both be converged into:

$$RS_t^- \xrightarrow{p} \frac{1}{2}\int_0^1 \sigma_s^2 ds \xleftarrow{p} RS_t^+ \tag{5}$$

Under in-fill asymptotic, the jumps in the process of *P* are:

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s ds + J_t \tag{6}$$

Then the realized variance of *P* converges into:

$$RV_t = \sum_{i=1}^{I} (P_{t,j+1} - P_{t,j})^2 \xrightarrow{p} \int_0^1 \sigma_s^2 ds + \sum_{s \le t} J_t^2 \tag{7}$$

And the downward realized semivariance and upward realized semivariance will converge into different limits under in-fill asymptotic:

$$RS_t^- \xrightarrow{p} \frac{1}{2}\int_0^1 \sigma_s^2 ds + \sum_{s \le t} J_t^2 1_{P_{t,j+1} - P_{t,j} \le 0} \tag{8}$$

$$RS_t^+ \xrightarrow{p} \frac{1}{2}\int_0^1 \sigma_s^2 ds + \sum_{s \le t} J_t^2 1_{P_{t,j+1} - P_{t,j} \ge 0} \tag{9}$$

From above, we can easily figure out that: $RV_t \overset{p}{=} RS_t^- + RS_t^+$. However, since the two components of RV_t can be distinguished, it must be more informative than mixed together. For the purpose of volatility measuring, we also introduce two another realized measures here. The first one is called realized range, proposed by Christensen and Podolskij (2005) and Martens and van Dijk (2007). This estimator is inspired by the idea of Parkinson (1980) that range-based variance estimator is much more efficient than return-based estimator. And this one is indeed reported more efficient and less contaminated by micro noises in empirical study. It is defined as follows:

$$RR_t = \frac{1}{4\log 2}\sum_{i=1}^{I}(H_{t,j+1} - L_{t,j})^2 \xrightarrow{p} \int_0^1 \sigma_s^2 ds + \sum_{s\le t} J_t^2 \tag{10}$$

In a driftless martingale process, this estimator also converges to quadratic variation. Usually for the estimation of one day's volatility, driftless martingale process assumption is not a bad one. The second one is called realized bipower variation. This estimator is proposed by Barndorff-Nielsen and Shephard (2002). It is defined as:

$$RB_t = \frac{1}{\mu_1^2}\sum_{i=1}^{I}| P_{t,j+1} - P_{t,j} || P_{t,j} - P_{t,j-1} | \xrightarrow{p} \int_0^1 \sigma_s^2 ds \tag{11}$$

Where μ_1 is a normalization factor. And in a semimartingale process with finite jumps, realized bipower variation converges to integrated variation but not quadratic variation.

Inspired by Chou (2005), we can know that his model can be naturally extended to model the upward (downward) realized semivariances with a little modification:

$$RS_t^{+(-)} = \lambda_t^{+(-)}\varepsilon_t^{+(-)}, \quad \varepsilon_t^{+(-)} \sim iid.f^{+(-)}(\cdot) \tag{12}$$

$$\lambda_t^+ = \omega^+ + \sum_{i=1}^{p}\alpha_i^+ RS_{t-i}^+ + \sum_{j=1}^{q}\beta_j^+\lambda_{t-j}^+, \lambda_t^- = \omega^- + \sum_{i=1}^{p}\alpha_i^- RS_{t-i}^- + \sum_{j=1}^{q}\beta_j^-\lambda_{t-j}^- \tag{13}$$

$$E[RV_{t+1} | I_t] = E[RS_{t+1}^+ + RS_{t+1}^- | I_t] = E[RS_{t+1}^+ | I_t] + E[RS_{t+1}^- | I_t] = \lambda_t^+ + \lambda_t^- \tag{14}$$

We call this model Asymmetric Multiplicative Error Model (AMEM), according to Engle and Gallo (2006), for Realized Semivariance (AMEM-RS). In the following empirical study, we compare volatility forecasting power in context of out-of-sample forecast of four different models: MEM-RV, MEM-RV with lagged return, AMEM-RS and AMEM-RS with lagged return.

3. Empirical Results

To calibrate our modelling approach, we employ high frequency Shanghai composite index data in this paper. The data contain observations from January 1, 2007 to January 4, 2013. After deleting the days of unavailable and insufficient information, we have 1570 days' observations of 1 minute's frequency data. The data is from the Wind database. Table 1 gives out the descriptive statistics of raw data and daily estimators obtained from raw data in every day.

Table 1. The descriptive statistics of raw and daily data

	Raw prices	Raw returns	Daily returns	Squared returns	Absolute returns	Ranges
Mean	2957	-9.7E-06	-3.1E-04	1.324	0.867	1.685
Median	277	0.000	0.0324	0.526	0.656	1.444
Maximum	6092	2.865	6.876	47.678	6.778	7.638
Minimum	1706	-5.458	-5.820	0.000	0.000	0.425
Std. Dev.	882.5	0.067	1.297	3.164	0.824	0.897
Skewness	1.524	-1.546	0.013	6.170	1.986	1.844
Kurtosis	4.845	24846	5.024	61.894	8.518	9.008
Jarque-Bera	745.0	1.6E+09	314	245278	3169	3467
Probability	0.000	0.000	0.000	0.000	0.000	0.000
Sum	7.8E+08	-5.650	-0.497	2456	1345	2598
Observations	628934	628933	1570	1570	1570	1570

Note. Raw returns, daily returns and range are all multiplied by 100; squared returns and absolute returns are respectively the squared value and absolute value of daily returns.

In order to compare models in terms of their prediction accuracy, we need to use proper proxies for underlying unobservable true volatility. According to Engle and Gallo (2006), there is still no consensus about a “true” or "best" measure of volatility. And “many ways exist to measure and model financial asset volatility”. Here we employ six measures of asset volatility for our model comparison. Three of them are three ordinary daily measures: absolute daily returns, daily Parkinson high-low range estimator and the most usual squared daily returns. We give their statistics description in Table 1. The other three of them are realized volatility measures: realized variance, realized range and realized bipower variation with the most used 5 minutes' frequency. Table 2 gives their statistics description together with RS+ and RS-.

Table 2. The descriptive statistics of realized estimators

	RV	RR	RB	RS+	RS-
Mean	1.642	1.110	1.636	0.811	0.831
Median	1.057	0.758	1.112	0.501	0.493
Maximum	35.663	23.719	21.293	19.202	33.875
Minimum	0.132	0.117	0.157	0.074	0.046
Std. Dev.	2.114	1.393	1.866	1.146	1.375
Skewness	6.303	6.837	4.896	6.321	11.779
Kurtosis	69.475	79.844	38.567	67.219	238.107
Jarque-Bera	299464	398511	89026	280241	3652223
Probability	0.000	0.000	0.000	0.000	0.000
Sum	2577	1743	2569	1273	1304
Observations	1570	1570	1570	1570	1570

Figure 2 presents the time series of RS+ against RS-. These two parts of realized variance do look very different from each other, and therefore two separately models for each of them is necessary. The information may be fruitful.

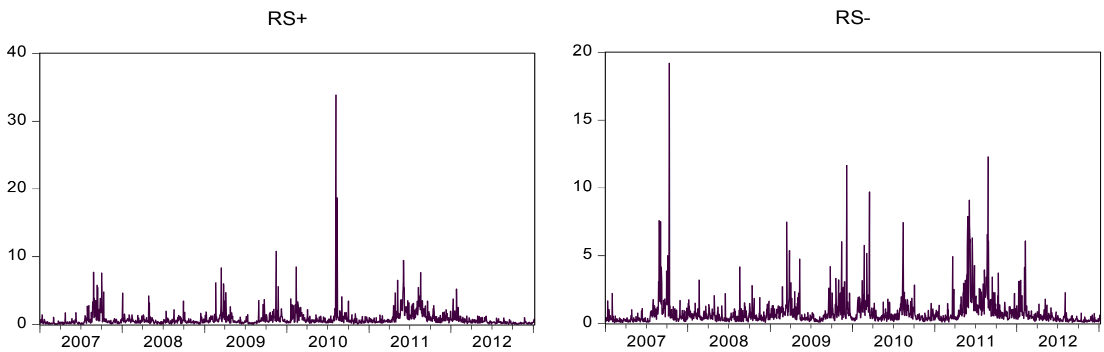

Figure 1. Upside and downside realized semivariance

In order to incorporate the leverage effects of lagged returns better, we estimate four models in this section: MEM-RV, MEM-RV with lagged returns, AMEM-RS and AMEM-RS with lagged returns. We employ the simplest form GARCH model for all of the four models–GARCH (1, 1), which is already adequacy in most applications according to Bollerslev, Chou, and Kroner (1992). Table 3 presents the estimated parameters of the four models.

Table 3. MEM type models for realized volatility and semivariance

	MEM-RV		MEM-RS			
			RS+	RS-	RS+	RS-
Constant	0.037 (0.050)	0.035 (0.032)	0.016 (0.017)	0.017 (0.016)	0.015 (0.011)	0.026 (0.011)**
ARCH	0.367 (0.097)***	0.219 (0.087)**	0.257 (0.078)***	0.282 (0.058)***	0.158 (0.052)***	0.152 (0.054)***
GARCH	0.623 (0.090)***	0.760 (0.086)***	0.731 (0.071)***	0.710 (0.065)***	0.823 (0.051)***	0.814 (0.057)***
Return(-1)		-0.097 (0.049)**				-0.065 (0.015)***
Return(-2)					-0.046 (0.022)*	
Log-L	-2446.0	-2438.9	-1910.8	-1912.5	-1902.5	-1896.1

Note. Model selection is based on AIC and BIC and numbers in parenthesis are the standard deviations, and stars refer to significance level of 10% (*), 5% (**) and 1% (***).

4. Models Comparison

According to Hansen and Lunde (2005) we continue to use the four loss functions employed by them as criterions for model:

$$MSE = n^{-1}\sum_{i=1}^{n}(MV_t - FV_t)^2 \tag{15}$$

$$MAE = n^{-1}\sum_{i=1}^{n}| MV_t - FV_t | \tag{16}$$

$$QLIKE = n^{-1}\sum_{i=1}^{n}(\ln FV_t - MV_t / FV_t)^2 \tag{17}$$

$$R^2LOG = n^{-1}\sum_{i=1}^{n}\ln^2(MV_t / FV_t) \tag{18}$$

The first two loss functions are regular ones. *QLIKE* is proposed by Bollerslev (1994), and is also called Gaussian quasi-maximum likelihood function, which can easily be recognized that it is originated from the likelihood function of GARCH model from its formulation. R^2LOG is proposed by Pagan and Schwert (1990), it aims to give some penalty to the asymmetry of the volatility forecasting. Different from the quadratic loss function, it was a proportional loss function. We focus on the out of sample comparisons for finding useful models in prediction of real world. In table 4, r^2, |r|, range, realized volatility, realized range and realized bipower variance are used as measurement volatility (MV) to judge the performance of the four models' fitting value in last section. It is clear that with most loss functions the lagged realized semivariance (RS-Lag) performs better than other forecasted volatilities (FV).

Table 4. Out of sample forecasting comparisons with different loss functions

	r^2	$\|r\|$	Range	RV	RR	RB
			Loss function: MSE			
RV	1311.01	316.10	173.71	161.84	166.19	147.48
RV-Lag	1291.31	264.08	130.04	159.16	128.32	154.23
RS	1314.82	299.29	156.18	159.38	156.75	151.35
RS-Lag	1275.95	227.06	104.69	159.07	110.93	155.88
			Loss function: MAE			
RV	177.15	110.01	79.57	70.02	75.87	61.68
RV-Lag	175.46	104.57	74.08	69.30	71.58	62.83
RS	178.71	109.48	76.87	70.22	76.95	62.96
RS-Lag	173.93	100.14	69.41	69.15	69.96	62.58
			Loss function: QLIKE			
RV	296.25	137.68	267.57	119.68	79.97	109.97
RV-Lag	280.83	130.42	257.72	120.26	74.55	110.61
RS	274.01	120.77	232.00	104.60	68.97	95.95
RS-Lag	256.58	117.03	233.38	110.12	65.46	101.07
			Loss function: R2LOG			
RV	581.08	184.60	30.35	23.71	45.23	18.45
RV-Lag	575.44	180.40	28.11	23.20	41.84	18.39
RS	586.71	186.37	27.38	24.15	46.95	19.09
RS-Lag	576.70	179.22	25.32	23.73	41.91	18.90

Note. For RS models we use upside RS and downside RS forecasting to synthesize RV forecasting and the model with minimum forecasting errors under four types of loss functions and six types of "true volatility" measurements.

5. Conclusion

Volatility is one of the core problems in many financial practices, but the asymmetry of volatility is often confused in arbitrage and risk management because downside volatility is definitely not equal to upside volatility in these fields. Separately modelling the two sides of volatility would be more informative than just mixing them together.

In this paper, we use a new modelling approach to model the realized semi variance with high frequency data in Chinese financial markets. Then the empirical study shows that when measured by six different volatility proxies, the realized semi variance (RS) performs better than the traditional realized volatility estimator (RV).

These findings reveal that when measuring volatility or fluctuations of financial assets, the usage of our new estimator will increase the performance of many financial practices like pricing or risk management. With the development of the information technology, high frequency data are more and more available. Developing an accurate and robust estimation of assets volatility is increasing important. One feasible way to extend this paper is to incorporate the correlation effects between upside realized semivariance and downside ones.

Table 5. In sample forecasting comparisons with different loss functions

	r^2	|r|	Range	RV	RR	RB
			Loss function: MSE			
RV	917.34	281.96	175.48	329.61	239.66	229.96
RV-Lag	868.62	223.06	126.37	304.65	194.04	206.50
RS	900.61	258.05	152.80	318.47	223.53	220.72
RS-Lag	857.14	195.18	105.52	299.71	178.72	202.70
			Loss function: MAE			
RV	161.01	103.54	80.22	81.53	84.83	72.35
RV-Lag	155.57	97.17	72.51	77.34	78.68	68.61
RS	160.62	102.74	77.13	81.24	85.25	72.52
RS-Lag	154.68	94.45	68.36	76.89	77.82	68.17
			Loss function: QLIKE			
RV	318.62	137.03	265.86	151.53	97.56	136.82
RV-Lag	296.85	129.43	253.80	146.52	88.02	131.02
RS	291.97	120.12	232.62	138.01	85.96	122.61
RS-Lag	280.73	114.25	224.65	133.55	77.94	118.88
			Loss function: R2LOG			
RV	567.85	178.04	32.61	32.08	55.92	25.60
RV-Lag	552.63	170.76	28.97	29.77	50.91	23.26
RS	572.36	179.32	29.69	32.67	57.67	26.03
RS-Lag	556.11	170.91	26.11	30.29	52.32	23.71

Acknowledgments

While remaining responsible for any errors in this paper, the authors would like to thank comments from the 18th International Conference on Macroeconomics Analysis and International Finance. We are also particularly grateful to "Application of Open Source R QuantLib in Financial Engineering Study" program of Jinan University and Social Science Association of Zhuhai city in Guangdong province. Thanks Luoming Fan and Qiuli Gao for their personal assistance in manuscript preparation.

References

Chou, R. (2005). Forecasting financial volatilities with extreme values: the Conditional Autoregressive Range (CARR) Model. *Journal of Money Credit and Banking, 37*(3), 561–582. http://dx.doi.org/10.1353/mcb.2005.0027

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics, 52*, 5–59. http://dx.doi.org/10.1016/0304-4076(92)90064-X

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*, 579–625. http://dx.doi.org/10.1111/1468-0262.00418

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13*, 253–263.

Engle. (2002). New frontiers for arch models. *Journal of Applied Econometrics, 17*(5), 425–446. http://dx.doi.org/10.1002/jae.683

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, (20), 873–889. http://dx.doi.org/10.1002/jae.800

Andersen, T., & Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*, 885–905. http://dx.doi.org/10.2307/2527343

Engle, R., & Russell, J. (1998). Autoregressive conditional duration: a new model for irregular spaced transaction data. *Econometrica, 66*, 1127–1162. http://dx.doi.org/10.2307/2999632

Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics, 131*(1–2), 3–27. http://dx.doi.org/10.1016/j.jeconom.2005.01.018

Ole, E., Barndorff, N., Silja, K., & Neil, S. (2008). *Measuring downside risk–realized semivariance*. Working paper. http://dx.doi.org/10.1093/acprof:oso/9780199549498.003.0007

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *Journal of Finance, 57*, 1047–1092. http://dx.doi.org/10.1111/1540-6261.00454

Corsi, F., & Renò, R. (2008). *Volatility Determinants: Heterogeneity, Leverage, and Jumps*. Retrieved from http://ssrn.com/abstract=1316953

Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics, 131*(1), 3–27. http://dx.doi.org/10.1016/j.jeconom.2005.01.018

Appendix A

In Sample Comparison

In this table, the same research in table 4 is organized and r^2, |r|, range, realized volatility, realized range and realized bipower variance are used as measurement volatility (MV) to judge the performs of the four models' fitting value of in sample forecasts. It also consistently indicates that with most loss functions the lagged realized semivariance (RS-Lag) performs better than other forecasted volatilities (FV).

Copyrights