ATLANTIS PRESS

Advances in Economics, Business and Management Research, volume 203

Proceedings of the 2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021)

Forecast on S&P 500 Index Based on HAR-RV Model
With VIX and Day-of-the-Week Effect

Qiannan Xiong[1,*]

[1] *Department of Economics, University of Macau, Macau SAR, China*
[*] *Corresponding author. Email: [1] sb82422@um.edu.mo*

ABSTRACT
The S&P 500 is an essential indicator for the U.S. and even the global stock markets. Meanwhile, the HAR-RV model is a new testing model, so predicting the realized volatility of S&P is significant to analyze using the HAR-RV model. This article will use the HAR-RV model to predict the S&P 500 index. Moreover, to make the model more accurate, the report adds the VIX and day-of-the-week effect into the formula. Finally, we get that VIX has a noticeable impact on the prediction of the S&P 500, but there is not enough evidence that the day-of-the-effect existed.

Keywords: *S&P 500, HAR-RV model, VIX, day-of-the-week-effect.*

1. INTRODUCTION

When we talk about the stock market in the US, it is expected to mention the standard and poor's 500(S&P 500), which is a significant indicator for the U.S. stock market. S&P 500 comprises 500 large-cap stocks, which represent the leading industries of the U.S. economy. These stocks are chosen by comparing the market size and liquidity. Their market capitalizations are more than $10 billion by using their stock price and the shares outstanding to calculate. Moreover, their liquidity is high, so they can easily to in the market buy or sell without influencing the asset price. Investors can see the relationship between risk and return by analysing the S&P 500. This article chooses the S&P 500 instead of the Dow Jones index because the Dow Jones only contains 30 companies, a much smaller amount than the S&P 500. The fewer companies there are, the less representative the U.S. stock market is.

Complex and changeable volatility has always been one of the typical characteristics of financial markets. The volatility of asset price changes plays a vital role in financial asset pricing, asset portfolio allocation, and financial risk management. Engle [1], Bollerslev [2], Nelson [3], and others have extensive and in-depth research on the measurement and modelling of the volatility of asset price, and they believe that the volatility of financial markets has a particular time-varying nature. Then, they introduced the ARCH/GARCH model to capture the aggregation effect on market volatility. Scholars conduct in-depth research from theoretical and empirical perspectives and have achieved fruitful research results. Nevertheless, Taylor [4] proposed the SV model (Stochastic Volatility, SV), which can describe the time-varying nature of market volatility and has better elasticity than the ARCH type. Therefore, it has also been widely used in the field of volatility research.

Domestic and foreign research institutions have unanimously recognized the traditional GARCH model, SV model, and other research results based on low-frequency financial data on asset price fluctuations. However, with the introduction and broad application of high-frequency financial data, the realized volatility, the realized amplitude, and the realized double power variation based on high-frequency data measurement contain more market information than the low-frequency model volatility. The volatility characterization accuracy also shows apparent advantages. Domestic and foreign scholars have combined their research to model high-frequency volatility from different perspectives. Corsi [5] published an article that raised the HAR-RV model based on the theory of heterogeneous markets. The first-order autoregressive volatility process is realized, representing the heterogeneous trading behaviour of traders in the market. Andersen and Bollerslev et al. [6] built a new HAR model(HAR-RV-CJ) based on the original one, based on the quadratic mutation theory to decompose the realized volatility into continuous sample path variance and jump variance to analyze the impact of volatility. To develop the HAR-RV-CJ model, Corsi and Reno [7] considered the effect of leverage on the realized volatility. They constructed the LHAR-CJ

Copyright © 2021 The Authors. Published by Atlantis Press International B.V.
This is an open access article distributed under the CC BY-NC 4.0 license -http://creativecommons.org/licenses/by-nc/4.0/.

model to capture the dynamic characteristics of financial asset volatility. According to Wen, F., Liu, X., &Tang, H, et al. [8]'s research, the LHAR-RV-V model was introduced to characterize the asymmetry of Chinese stock market volatility. Their study was based on the HAR-RV model, which considers the leverage effect and discusses the direct connection between volume and price. Additionally, Li, W., Cheng, Y., & Fang, Q. [9] used both the HAR-RV model and the HAR-RV-CJ model mentioned before to test the volatility of silver futures.

In summary, modelling based on high-frequency volatility has made significant progress both theoretically and empirically. In short, at present, high-frequency volatility modelling is mainly based on the two basic models of ARFIMA and HAR-RV, which are continuously expanding. While enriching the financial market theory, it also reveals more financial market volatility and microstructure characteristics.

In this article, the VIX and day-of-week effect will be added to make the prediction more accurate. VIX has the full name Chicago Board Options Exchange Volatility Index, but it has a more common name by using its stock code-the VIX. It was created by the Chicago Board Options Exchange (CBOE) in 1990 and is maintained by CBOE Global Markets to measure the market risk and investor's sentiments. The principle of the VIX is to reflect market participants' expectations of volatility over the next 30 days. VIX always goes in the opposite direction with S&P 500, but it may not reflect the S&P 500 index volatility so well. For example, Vodenska and Chambers [10] have proved that VIX overrates the S&P 500 index volatility during the stable period while underestimates it throughout a high volatility period. Moreover, for S&P 500 and VIX, Auinger [11] has proved no statistically significant causal relationship between them. However, VIX is still an essential component to predict the volatility of the S&P 500 index, so this article will add the VIX into the HAR-RV model to calculate.

The day-of-the-week effect is that the yield of the stock market varies within a week, and different stock markets will also have other day-of-the-week effects. Moreover, even in the same market, the day-of-the-week effect is different in different periods. For example, Kie Ann Wong, Tak Hee Hui, and Choy Yin Chan [12] show that Singapore, Malaysia, Hong Kong, and Thailand have negative average returns on Mondays or Tuesdays and favourable returns on Fridays. Additionally, in China, Feng, L.[13] have proved the Shanghai and Shenzhen stock markets both have negative "Tuesday effect" and significant "main effect".

In this paper, we will only concentrate on the HAR-RV model. This article has five sections. One is the introduction above to show the background and the motivation of this article. The next part is to show the data used in the paper. The third section is the methodologies and the models we use to run the regression, the next part is the result analysis, and the last part is the conclusion.

2. DATA

The impact of financial market information on asset prices is a continuous process. When the data sampling frequency is low, it will cause the data to be scattered, which will lead to varying degrees of information loss. Thus, to improve the accuracy, this article chooses the high-frequency data, 5-minute-frequency trading data to measure the price volatility of the S&P 500 price. The period of the data is from January 22, 2015, to May 28, 2021, with 12499 raw five-minute frequency data, and this article finally gets 1600 trading date's data after calculating the RV and eliminating the vacancy value. The data used in this article are from Global Financial Information Service(GFIS) and Wind, which are two financial data and information analysis platforms.

Figure 1 shows the six-year trend chart of the S&P 500 price, which is a rising trend. From January 2015 to September 2015, the price trend was relatively flat. From September 2015 to May 2016, there were small continuous fluctuations. Beginning in May 2016 and continuing until February 2018, the S&P 500 price rose steadily without significant changes. From February 2018 to February 2020, there have been varying degrees of turbulence in the two years. The most extensive fluctuation period is from February 2020 to September 2020, and the price difference is around 1500. After that, it continued until May 2021 and maintained steady growth.

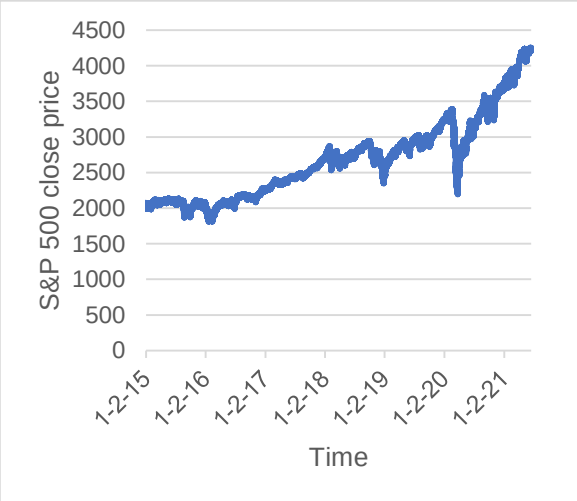

Fig. 1. Five minutes high-frequency price dynamics of S&P500.

Figure 2 shows the daily price volatility dynamics of S&P500, a steady trend with little volatility. There was a considerable large fluctuation from February 2020 to May 2020, and the RV reached 112. There was a

slightly significant fluctuation from September to October 2015. RV is more than 40. The remaining volatility is no more than 40, and the volatility trend is a relatively smooth straight line. Some structure breakpoints and periods may exist, but this article will not test the structure break effect.

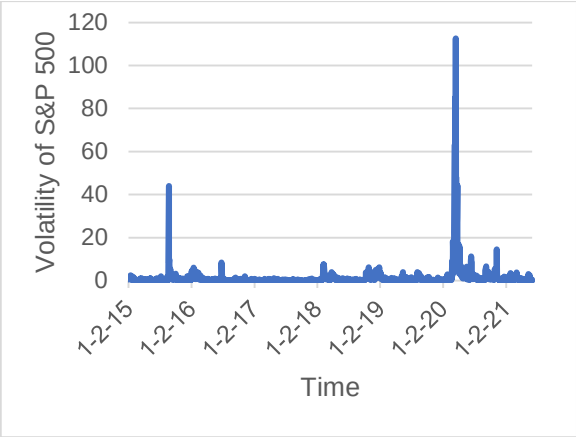

Fig. 2. Daily price volatility dynamics of S&P500

3. ECONOMETRIC MODELS

3.1. HAR-RV model

Andersen and Bollerslev [14] proposed a new nonparametric volatility estimation method——Realized Volatility(RV) for the first time. The HAR-RV model divides the data into daily data, weekly data, and monthly data, respectively representing the traders' trading data of traders with different holding period preferences. The HAR-RV model uses high-frequency data of 5 minutes to calculate the three kinds of data. The formula is as follows:

$$RV_{d,t} = \sum_{i=1}^{M} r_{t,i}^2 \tag{1}$$

$r_{t,i}^2$ is the square sequence of the intraday interval return rate $r_{t,i}$, and we multiply by 100 to make the result easily observed.

$$r_{t,i} = \left(LnP_{t,i} - LnP_{t,i-1}\right) \times 100 \tag{2}$$

The weekly and monthly data should also be calculated by the weighted average of daily data calculated above.

$$RV_{w,t} = \frac{RV_{d,t}+RV_{d,t-1}+RV_{d,t-2}+\cdots+RV_{d,t-4}}{5} \tag{3}$$

$$RV_{m,t} = \frac{RV_{d,t}+RV_{d,t-1}+RV_{d,t-2}+\cdots+RV_{d,t-20}}{21} \tag{4}$$

The target is to predict the S&P 500 index, so we need to find the average RV from day t to (t + H) using the following model. In simple terms, the formula delays the daily, weekly, and monthly data obtained before by one order.

$$\overline{RV}_{t+H} = \frac{1}{H}\sum_{i=1}^{H} RV_{d,t+i} \tag{5}$$

Moreover, this article chooses the logarithmic form of the HAR model to do the forecast. The formula is expressed below:

$$Ln\overline{RV}_{t+H}=\beta_0 + \beta_d RV_{d,t} + \beta_w RV_{w,t} + \beta_m RV_{m,t} \tag{6}$$

3.2. HAR-RV-V model

The VIX index, compiled by the CBOE in the US, is an annualized measure of the volatility of implied short-term expectations based on the price of options on the S&P 500. It is an index to reflect participants' views on the volatility of the market. In general, the 'fear index' known as the VIX tends to fall when stocks rise, and there is usually a negative correlation between them. Thus, VIX is a contrarian indicator with S&P 500. This model is to identify how VIX influences the performance of the S&P 500. We also use the logarithmic form as below:

$$Ln\overline{RV}_{t+H}=\beta_0 + \beta_d RV_{d,t} + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \beta_v VIX \tag{7}$$

3.3. HAR-RV-VW model

The model adds dummy variables to test the weekly effect. The weekly effect means that the average yield of one trading day is hugely lower than the average yield of other trading days and is statistically significantly negative. Thus, we set the data of Friday as the dummy, which means Fri=1, and Mon=Tue=Wed=Thu=0. After running the regression model, we can verify whether there is a weekly effect through the dummy variable coefficients obtained after the regression. For instance, if the coefficient of one dummy variable θ_i is positive, the volatility of this working day is significantly greater than the volatility of the other four days. Otherwise, there is no weekly effect. The logarithmic formula is as follows:

$$Ln\overline{RV}_{t+H}=\beta_0 + \beta_d RV_{d,t} + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \theta_1 Mon + \theta_2 Tue + \theta_3 Wed + \theta_4 Thu \tag{8}$$

3.4. HAR-RV-V-VW model

This model is the last HAR-RV derivative formula, combining the previous three HAR-type models,

including VIX and weekly effects. We expect the accuracy of the prediction to increase as the variable increase. The following logarithmic regression is shown below:

$$Ln\overline{RV}_{t+H}=\beta_0+\beta_d RV_{d,t}+\beta_w RV_{w,t}+$$
$$Ln\overline{RV}_{t+H}=\beta_0+\beta_d RV_{d,t}+\beta_w RV_{w,t}+$$
$$\beta_m RV_{m,t}+\beta_v VIX+\theta_1 Mon+\theta_2 Tue+\theta_3 Wed+\theta_4 Thu \quad (9)$$

4. RESULT

Refer to the following descriptive statistical analysis(Table 1), and we can see RV_d, RV_w, RV_m all have significant and robust volatility. The range of RV_d is from 0.014 to 122.5, the range of RV_w is from 0.046 to 57.89, range from $RV_m$is from 0.073 to 27.54. The volatility is keeping narrowing. Moreover, although the mean of RV_d, RV_w, and RV_m are almost the same, the standard deviation of these three are on the downward trend, which means that the volatility of daily RV is more significant than that of weekly RV and that of weekly RV is more prominent than monthly RV. Additionally, VIX has the most extensive range from 9.14 to 82.69, and the standard deviation is also much more significant than any other three variables. Median is increasing follow the rank(RV_d, RV_w, RV_m, and VIX).

Table 1 Descriptive statistical analysis

	N	mean	p50	sd	min	max
rvd	1600	1.248917	0.375727	4.949281	0.0139037	112.522
rvw	1600	1.249148	0.4327655	4.031468	0.0462939	57.88881
rvm	1600	1.249155	0.4904204	3.153021	0.0728765	27.53982
vix	1600	17.71774	15.21	8.137713	9.14	82.69

4.1. Parameter estimations

Table 2. Parameter estimation results of HAR-RV-type models

	HAR-RV				HAR-RV-V			HAR-RV-VW			HAR-RV-V-VW	
	1-day	1-week	1-month	1-day	1-week	1-month	1-day	1-week	1-month	1-day	1-week	1-month
β_0	-0.196***	0.002	0.004*	-0.547***	-0.191***	-0.065***	-0.131***	0.008	-0.012***	-0.485***	-0.184***	-0.057***
	(0.019)	(0.006)	(0.003)	(0.0867)	(0.029)	(0.011)	(0.037)	(0.012)	(0.005)	(0.091)	(0.031)	(0.012)
$lnRV_t^d$	0.021	0.090***	0.005	0.002	-0.079***	0.001	0.021	0.090***	0.006	0.002	0.079***	0.002
	(0.026)	(0.009)	(0.03)	(0.027)	(0.009)	(0.004)	(0.026)	(0.009)	(0.004)	(0.027)	(0.001)	(0.004)
$lnRV_t^w$	0.935***	0.823***	0.024***	0.894***	-0.801***	0.015***	0.935***	0.824***	0.023***	0.893***	0.801***	0.015***
	(0.042)	(0.014)	(0.006)	(0.043)	(0.014)	(0.006)	(0.042)	(0.014)	(0.006)	(0.043)	(0.014)	(0.006)
$lnRV_t^m$	0.020***	0.081***	0.969***	-0.028	-0.055***	0.959***	0.019	0.081***	0.969***	-0.029	0.055***	0.959***
	(0.035)	(0.012)	(0.005)	(0.037)	(0.012)	(0.005)	(0.035)	(0.012)	(0.005)	(0.037)	(0.012)	(0.005)
VIX				0.016***	0.009***	0.003***				0.016***	0.009***	0.003***

				(0.004)	(0.001)	(0.001)				(0.004)	(0.001)	(0.001)
Mon							-0.091*	-0.001	-0.007	-0.098**	-0.005	-0.008
							(0.050)	(0.017)	(0.007)	(0.050)	(0.017)	(0.007)
Tue							-0.112**	-0.009	-0.010	-0.116**	-0.011	-0.011*
							(0.049)	(0.017)	(0.007)	(0.049)	(0.016)	(0.006)
Wed							-0.074	-0.022	-0.009	-0.076	-0.022	-0.010
							(0.049)	(0.017)	(0.007)	(0.049)	(0.016)	(0.006)
Thu							-0.047	0.002	-0.015**	-0.050	-0.000	-0.016**
							(0.049)	(0.017)	(0.007)	(0.049)	(0.016)	(0.006)
Adj_R^2	0.7442	0.9631	0.9934	0.7468	0.9641	0.9935	0.7446	0.9630	0.9934	0.7472	0.9641	0.9935

In table 2, we formed the four regression results into one table to directly compare the differences. There are three horizontal lines——the daily result, weekly result, and monthly result. For the HAR-RV model, no matter 1-day, 1-week, or 1-month results, all daily RV, weekly RV, and monthly RV have a positive relationship with the future S&P 500 RV. Moreover, both weekly RV and monthly RV have significant correlations. For the HAR-RV model, the association is either positive or negative for daily RV, weekly VR, and monthly RV. In predicting weekly and monthly results, weekly RV and monthly RV still play a significant role. Still, in the prediction of daily outcomes, the weekly RV did not show a significant correlation. For the HAR-RV-VW model, all daily RV, weekly RV, and monthly RV positively correlate with RV prediction. The significance of the correlation is the same as that of the HAR-RV-V model. For the HAR-RV-V-VW model, daily RV, weekly RV, and monthly RV positively correlate with the forecast except for the monthly RV on a 1-day RV prediction. The significance of the correlation is the same as the HAR-RV-VW model.

For the VIX index, wherever in HAR-RV-V model and HAR-RV-V-VW, it is positively significant with 1-day, 1-week, and 1-month RV prediction, which is the same result as our assumption. Because the VIX p moves inversely to the S&P 500, their relationship has proved to be very tight. Hence, the significant positive result shows that the 1 unit change of the VIX will cause the volatility of the S&P 500 increase.

For the day-of-week effect, the regression result of the dummy variable shows that in both HAR-RV-VW and HAR-RV-V-VW models, the dummy variable is not so significant. In the HAR-RV-VW model, all dummy variables show a negative result except the Thu dummy variable in 1-week prediction shows a positive impact, which means that trading on Thursday makes the future weekly S&P 500 index volatile more. There is only a Mon dummy, and the Tue dummy in 1-day prediction and Thu dummy in the 1-month forecast have a significant correlation. Moreover, in the HAR-RV-V-VW model, all four dummy variables are negatively correlated, and this means that when the trading date is these four days, the volatility of the S&P 500 is less than on Friday. However, only the Mon and Tue dummy in 1-day prediction and the Thu dummy in 1-month

prognosis significantly correlate. Thus, in our article, most dummy variables show a negative relationship with little significance.

From the adjusted-R square, we can find 1-month prediction has a giant R square than that of the 1-week forecast than that of the 1-day forecast, which means the model fits the 1-month forecast the most and is followed by the 1-week prediction and 1-day prediction.

5. CONCLUSION

This article uses the Standard & Poor's 500 Index, an index composed of 500 large-cap stocks representing the leading industries in the US economy, as our data indicator. This article is interested in how the realized volatility data can predict future volatility. The data used in this article is from WIND and GFIS, which are big platforms to select data. For the methodology and the model, this article uses five-minute high-frequency data to avoid losing information.

Based on the HAR-RV model, we first study the original daily, weekly and monthly data. Then, we add the VIX index and the day-of-week effect step by step to form the HAR-RV-V model, HAR-RV-VW model, and VAR-RV-V-VW model.

The regression results show that the monthly model fitted the best, followed by the weekly and daily models. Moreover, although we did not do the robust test, our test is stable since our model already lagged the data by one order. The HAR-RV model fits the S&P 500 well, so we can look forward to using this model for more volatility tests on other indexes.

REFERENCES

[1] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.

[3] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

[4] Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, 4(2), 183-204.

[5] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174-196.

[6] Andersen, T. G., Bollerslev, T., & Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high‐frequency data and realized volatilities. *Econometrica*, 73(1), 279-296.

[7] Corsi, F., Pirino, D., & Reno, R. (2008). Volatility forecasting: The jumps do matter. Università di Siena.

[8] Wen, F., Liu, X., &Tang, H, et al. Study on the volatility of Chinese Stock Market based on LHAR-RV-V model [J]. *Journal of Management Sciences in China*,2012(6):59-67.

[9] Li, W., Cheng, Y., & Fang, Q. (2020). Forecast on silver futures linked with structural breaks and day-of-the-week effect. *The North American Journal of Economics and Finance*, *53*, 101192.

[10] Vodenska, I., & Chambers, W. J. (2013, August). Understanding the relationship between VIX and the S&P 500 index volatility. *In 26th Australasian Finance and Banking Conference.*

[11] Auinger, F. (2015). *The Causal Relationship between the S&P 500 and the VIX Index: Critical Analysis of Financial Market Volatility and Its Predictability.* Springer.

[12] Wong, K. A., Hui, T. H., & Chan, C. Y. (1992). Day-of-the-week effects: evidence from developing stock markets. *Applied Financial Economics*, 2(1), 49-56.

[13] Licheng Feng. (2000). The day-of-the-week effect in China's stock market. *Economic Research Journal*, (11), 50-57.

[14] T.G. Andersen, T. Bollerslev. Heterogeneous information arrivals of return volatility dynamics: increasing the long run in high frequency returns. Journal of Finance, 52 (1997), pp. 975-1006