

Volatility Forecasting Models and HAR-RV Model Group The development of volatility forecasting

Yue Zhao*

Department of Mathematics, University College London, WC1E 6BT, London, the United Kingdom

*Corresponding author: viviann.zhao.19@ucl.ac.uk

Keywords: Volatility forecasting, Realised volatility, GARCH model, HAR-RV model

Abstract: Volatility forecasting has a critical role in pricing options. The paper compares different models on projection theoretically and then discusses the process of conducting derivational HAR-RV models selected. The core of the models is established under the concept of realised volatility based on high-frequency data, which provides more accurate and precise values. Nowadays, HAR-RV is the most commonly used model to predict volatility. The review gives an insight into the difference between models and within the HAR-RV model group. In addition, the paper also proposes potential developing and improving paths. The research on volatility forecasting should be advanced uninterruptedly, since volatility is a decisional part of the option

1. Introduction

Modelling and forecasting for financial assets are some of the most significant and core discussions in academic and practice fields. Volatility is a statistical measure of the dispersion of returns and uncertainties of a security or market index, which is highly applied in modeling. The predicted value of volatility may influence the decisions by investors on option trading, risk management, asset allocation, etc. Stein [1] stated the memorability of the trend of volatility, which means that the volatility does not change sharply from the original fluctuating range and previous direction in the short run. Rubinstein [2] and Engle and Ng [3] found the leverage effect, the negative correlation between an asset's return and its volatility. They also illustrated that the enormous fluctuation appeared if there was any big news. Volatility clustering is another property of volatility. Cont [4] suggested that the volatility keeps a more significant value in a period and stays at a relatively small value in some other duration.

Models like GARCH (General autoregressive conditional heteroskedasticity) and SV (Stochastic volatility) are used in the early stage. They are supposed to provide less convincing and accurate results on volatility research because of low-frequency data [5] which are defined as the daily and weekly data, even monthly and annual data. And High-frequency data is recognised as data in one day, mainly collected each hour, each minute, and even each second. Andersen used high-frequency option prices and Bollerslev [6] to introduce realised volatility (RV), which had few errors in results compared with actual historical data. Corsi [7] proposed the heterogeneous autoregressive model with realised variance (HAR-RV model) based on the Heterogeneous Market Hypothesis. This model described and characterised the long memory of RV. In 2009, Coris [8] discovered the influence of RV by the composition of daily, weekly, and monthly RV by autoregressive process, which stated the continuity of volatility. Andersen et al. [8] decomposed realised volatility into continuous sample path variation and discontinuous jump variation based on the HAR-RV model. They constructed HAR-RV-J and HAR-RV-CJ models, which had greatly improved the accuracy of prediction to future volatility. Black and Scholes [9] obtained the B-S option pricing model using the non-arbitrage pricing principle based on the Risk neutral assumption, which was achieved to estimate volatility. Britten-Jones and Neuberger [10] proposed a model-free method of volatility projection to exclude the impact and errors caused by models. Fair and Shiller [11] suggested the inclusion of regression, which is applicable to compare each model's information.

Inspired by the existing studies and researches, in this paper, I discuss different models on volatility research and review the development of HAR models especially. Section 2 introduces the models studying volatility forecasting. In Section 3, there is a discussion of the development of HAR models. And Section 4 presents some of the potential research paths on volatility.

2. Models on Volatility Study

During the past half-century, researchers have investigated deeply and established different models. The model can be divided into four types, GARCH, SV, Implied Volatility, and HAR-RV. The implied volatility methods estimate future volatility relied on option trading data for projection of future volatility. And the other three types of models use historical stock return data.

2.1. GARCH Model

2.1.1. Traditional GARCH Model

Engle [12] proposed Auto Regressive Conditional Heteroscedasticity Model (ARCH model) based on variance. Later, Bollerslev [13] raised the Generalised ARCH Model (GARCH Model). Then researches came out with different improved models, including Integrated GARCH, Asymmetric GARCH, Exponential GARCH, etc. In these GRACH models, the equations of the rate of return are the same. The difference exists in the method of how residuals are measured. GARCH model shows the clustering of time series.

GARCH models assume that the rate of return satisfies that:

$$R_t = \mu_t + \varepsilon_t \quad (1)$$

where μ_t is the expected rate of return, ε_t is the volatility. Notice that for the variance σ_t^2 and the standard normally distributed innovation sequence z_t , their relation can be expressed as:

$$\varepsilon_t = \sigma_t z_t \quad (2)$$

Usually, it assumes that the variance satisfies the following condition:

$$\sigma_t = \varpi + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (3)$$

where $\varpi > 0$; for all $i > 0, j > 0$; $\alpha_i \geq 0$ and $\beta_j \geq 0$; $\sum_{i=1, j=1}^{\max(p,q)} \alpha_i + \beta_j < 1$

2.1.2. Realised GARCH Model

According to Andersen, etc. [6], realised volatility (RV) is defined as the summation of squares of the high-frequency rate of return. The RV on trading day Day t can be expressed as:

$$RV_t = \sum_{i=1}^n r_{t,j}^2 \quad (4)$$

where $r_{t,j}$ is the log return on Day t time duration j, which the time duration is divided into n parts.

The traditional GARCH Model selects daily data to model volatility, which may ignore significant changes in days. Hansen [14] proposed Realized GARCH model based on high-frequency data by combining RV and GARCH model.

2.2. SV

In the GARCH model, outliers would influence variances, which leads to unstable and less accurate volatility. The Stochastic Volatility Model (SV) is more stabilised as variances follow the stochastic process. Clark [15] first proposed the concept of the Stochastic Volatility Model. Inspired by Brownian motion, he suggested expressing log price as:

$$M_t = W_{\tau_t} \quad (5)$$

where $t > 0$, W refers to Brownian motion, and τ is the change of time. W and τ are independent, and thus, $M_t | \tau_t \sim N(0, \tau_t)$. After adding the adequate remuneration to compensate M [5], the log pricing model of asset is recognised: $Y = M + A$.

2.3. Implied Volatility

2.3.1. Black–Scholes Option Pricing Model

Black and Scholes [9] invented the risk-neutral argument and demonstrated the European option pricing model, the Black-Scholes model. They introduced a mathematical equation to find the theoretical value of European-style options, taking account of the impact of time and other risk factors. The well-known Black–Scholes formula is:

$$C = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad (6)$$

where

$$d_1 = \frac{1}{\sigma \sqrt{T-t}} \left[\ln \left(\frac{S_t}{K} \right) + \left(r + \frac{\sigma^2}{2} \right) (T-t) \right] \quad (7)$$

$$d_2 = d_1 - \sigma \sqrt{T-t} \quad (8)$$

The notation used here are defined:

C : the price of call option

r : the annualised risk-free interest rate;

S_t : the price of the underlying asset at time t ;

t : the time

T : time of option expiration;

σ : the standard deviation of returns, i.e., the implied volatility

K : the strike price of the option

From the formula, it is noticeable that there exist seven parameters, and six of them can be recorded directly from the real market. Therefore, by applying the procedure, the volatility can be conducted reversely.

2.3.2. Model-Free Implied Volatility

The assumptions in the Black-Scholes model are ideal and cannot be achieved in a real-life market. In 2020, Britten-Jones and Neuberger [10] pointed out that model-free implied volatility depends on the current option price and is independent of the model used. Hence, it can prevent errors resulting from the inaccuracy of the model.

Generally, the model-free implied volatility is equal to the expected sum of squared returns under a risk neutral measure. To be more accurate, assume T to denote a limited time duration, h to denote the time interval, S_0 to denote the initial stock price, S_t to denote the stock price at time t , K to denote the strike price and C to denote the call option price. The strike price K is defined as $K = \{L: K = S_0 k^i, i = 0, \pm 1, \pm 2, \dots, \pm M\}$, and $k > 1$.

$$E[\max(S_t - K, 0)] = C(t, K), t \in T, K \in \mathbf{K} \quad (9)$$

$$E \left[\int_0^t \left(\frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(t, K) - \max(S_0 - K, 0)}{K^2} dK \quad (10)$$

The above expressions can derive the implied volatility by using the definition of variance $\sigma^2 = E[X^2] - E[X]^2$.

2.4. HAR-RV Model

Müller et al. [16] first introduced the Heterogeneous Market Hypothesis (HMH) in 1993. The hypothesis believes that different participants in the heterogeneous market have different characteristics. Since traders differ in risk preference, regulations, time horizons, etc., they might interpret the same information given differently and make distinct responses to opportunities.

Based on HMH, Corsi [7] proposed a Heterogeneous Auto-Regressive of Realised Volatility model (HAR-RV). There exist three types of traders in the market, different from the length of the trading time, daily, weekly and monthly. Their behaviours can be detected by daily realised volatility (RV_t), weekly realised volatility (RVW_t), and monthly realised volatility (RVM_t), respectively. The weekly RV is the average value of RV for the past five days, and the monthly RV is the average value of RV for the past twenty-two working days like shown below:

$$RVW_t = \frac{1}{5}(RV_t + RV_{t-1} + RV_{t-2} + RV_{t-3} + RV_{t-4}) \quad (11)$$

$$RVM_t = \frac{1}{22}(RV_t + RV_{t-1} + RV_{t-2} + \dots + RV_{t-20} + RV_{t-21}) \quad (12)$$

The core model expression is:

$$RV_{t+1} = C + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \varpi_t \quad (13)$$

where β are parameters, ϖ represents errors, and C is a constant.

3. The Development of HAR-RV Model

3.1. Leverage Effect

Bollerslev et al. [17] tested the leverage effect in high-frequency data in 2006. In 2008, considering the inherent clustering of the volatility of residuals in HAR models, Corsi et al. [18] interpreted the volatility of RV by giving a GARCH complement to the HAR models and hence introduced the HAR-GARCH model. In 2010, Corsi and Renò [19] included the issue that the probability of the volatility increasing after a negative shock is higher than after a positive shock with identical conditions and thus, constructed the HAR-L model.

Besides, Asei et al. [20], Corsi and Renò [19], and Gong and Lin [21] had proved strongly that the inclusion of the leverage effect could contribute to a better forecasting result for RV.

3.2. Structural Breaks

Hammoudeh and Li [22] found that the change of the rate of return, i.e., volatility, don't vibrate in a fixed range. There existed such a structural change that the change led to the outliers being out of the range. The structural change reduces the continuity of the volatility in the market and consequently highly impacts the prediction result.

Ewing and Malik [23] tried to include the factors of structural breaks in their research of the GARCH model on volatility. They found that the GARCH model considering structural changes performed better in the interpretation and projection of volatility. Inspired by that, Gong et al. [24] combined leverage effect and structural breaks and succeeded in improving the accuracy by adding the two factors.

3.3. Examples of Further HAR-RV Model

In the past twenty years, researchers have proposed and constructed many different improved models based on the core idea of the HAR-RV model. Here, we introduce the three types of models, which are conducted progressively.

3.3.1. HAR-RV-J

HAR-RV model reflects the characteristics of volatility in view of the long term. However, the partial regression coefficient measured the influence on the market's whole volatility by specific traders. Thus, Andersen et al. [8] introduced discontinuous jump variation to the HAR-RV model and constructed HAR-RV-J:

$$RV_{t+1} = C + \beta_d RV_t + \beta_w RVW_t + \beta_m RVM_t + \beta_j J_j + \varpi_t \quad (14)$$

where J refers to the jumping variance.

3.3.2. HAR-RV-CJ

Andersen et al. [8] constructed the HAR-RV-CJ model based on the HAR-RV-J model and significant jump detection tests proposed by Huang and Tauchen [25]. They defined that significant jump component as:

$$CJ_t \equiv I(Z_t > \Phi_\alpha)[RV_t - RBV_t] \quad (15)$$

where parameters:

RVW_t is the realised bipower variance, defined as:

$$BV_t = u_1^{-2} \sum_{j=2}^{\frac{1}{\Delta}} |r_{(t-1)+j^*\Delta}| |r_{(t-1)+(j-1)^*\Delta}| \rightarrow \int_0^t \sigma^2(s) ds \quad (16)$$

Φ_α represents the number at $1/\alpha$ position on standard normal distribution;

Z_t is the statistical value of significant jump detection tests from Huang and Tauchen [25], defined as:

$$Z_t = \delta^{-\frac{1}{2}} \frac{(RV_t - RBV_t)^* RV_t^{-1}}{\sqrt{\left(\frac{\pi^2}{4} + \pi - 5\right) \max\left\{1, \frac{TQ_t}{(RBV_t)^2}\right\}}} \quad (17)$$

TQ_t refers to the realised tripower quarticity, defined as:

$$TQ_t = \Delta^{-1} u_4^{-3} \sum_{j=2}^{\frac{1}{\Delta}} |r_{(t-1)+j^*\Delta}| |r_{(t-1)+(j-1)^*\Delta}| |r_{(t-1)+(j-2)^*\Delta}| \rightarrow \int_0^t \sigma^4(s) ds \quad (18)$$

By using the above expressions, it is simple to have the continuous RV under different significant levels, which the continuous RV is defined as:

$$CRV_{t,\alpha} = I(Z_t \leq \Phi_\alpha) \cdot RV_t + (Z_t > \Phi_\alpha) \cdot RBV_t$$

$CRV_{t,\alpha}$ negated the impact of fluctuation of significant jump on RV. Therefore, the HAR-RV-CJ model is obtained:

$$RV_{t+1} = C + \beta_d CRV_t + \beta_w CRVW_t + \beta_m CRVM_t + \beta_j CJ_j + \varpi_t \quad (19)$$

3.3.3. HAR-RV-TCJ

Corsi et al. [26] showed that when the continuous jump appeared highly frequently in high-frequency data, Z_t proposed by Huang and Tauchen [25] might not be able to detect some jumps. Therefore, Corsi et al. [26] introduced corrected Realized Threshold Multipower (C_TMPV) to give a corrected Z_t , denoted by CT_Z_t .

More detailed and further discussions on these two variables can be found in the research by Corsi et al. [26]. According to the tests of CT_{Z_t} , the significant jump component should be redefined as:

$$TCJ_t \equiv I(CT_{Z_t} > \Phi_\alpha)[RV_t - RBV_t] \quad (20)$$

Also, redefine the continuous RV as:

$$TCRV_{t,\alpha} = I(CT_{Z_t} \leq \Phi_\alpha) \cdot RV_t + (CT_{Z_t} > \Phi_\alpha) \cdot RBV_t \quad (21)$$

Eventually, the HAR-RV-TCJ model's expression is:

$$RV_{t+1} = C + \beta_d TCRV_t + \beta_w TCRVW_t + \beta_m TCRVM_t + \beta_j TCJ_j + \varpi_t \quad (22)$$

4. Future Research Expectation

Looking at the current researches published, mainly the reviews of models focus on the financial index or limited range of futures markets like crude oil and copper. This is because the price of crude oil and copper constantly fluctuate the most and have the most significant impacts on other markets and industries. For example, Gong and Lin [27, 28] explored volatility forecasting with structural breaks of copper futures and crude oil futures in 2018. However, with the deep investigation in the futures market, different research areas may be divided for each commodity.

It is always expected to enhance accuracy. A good example is that Cai and Xiang [29] found that the traditional RV is sensitive to the microstructure noise existing in the copper futures market. Therefore, they employed the model incorporating generalised realised measures and VaR. Microstructure noise is a crucial factor that results in ineligible errors in the prediction. More methodologies are welcomed to be conducted to negate microstructure noise.

Not mentioned in this paper, but in 2020, Maki and Ota [30] demonstrated the introduction of asymmetric properties of RV to models. They claimed the significance of asymmetry but have not recognised the exact unique type of asymmetry which is the most critical to the models.

5. Conclusion

This paper reviews different fundamental models for volatility forecasting and discusses the HAR-RV model and its derivational models. I introduce the mechanisms of the GARCH model, SV model, Black-Scholes Option Pricing Model, Model-Free Implied Volatility, and HAR-RV model. The HAR-RV model is a simple concept but can be derived into different complex models and combined with other methods. As fundamental derived models, HAR-RV-J, HAR-RV-CJ and HAR-RV-TCJ are presented by the process of derivation. Although there are many models and improved versions, the accuracy of forecasted volatility still needs to be improved. There are still areas hidden behind models.

The paper is based on the present outcomes by other researchers, and the models mentioned are not practised with real-life data to verify the properties claimed. Moreover, there are more complicated models like the Vector autoregression model (VAR) and HAR-RV-RS model based on state space. These two types of models add another dimension to simplify or strengthen their original models, respectively. Additionally, we have not analysed the HAR-RV-GARCH model, which combines the HAR-RV model and GARCH model.

References

- [1] Stein J. Overreactions in the option market. *The Journal of Finance*. 1989;44(4):1011-1023.
- [2] Rubinstein M. Implied binomial trees. *The Journal of Finance*. 1994;49(3):771-818.
- [3] Eagle R F, Ng V K, Measuring and testing the impact of news on volatility. *The Journal of Finance*. 1993;48(5):1749-1778.
- [4] Cont R, Volatility clustering in financial markets: Empirical facts and agent-based models in the option market, in: *Long memory in Economics*, 2007, 289-309.
- [5] Zhang Pengyun. Research on the forecasting performance of the HAR-RV model and its expand models-----evidence on the CSI300 high-frequency data. MA thesis, Southwest Jiaotong University; 2014.

- [6] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*. 1998;39(4):885-905
- [7] Corsi F. A simple long memory model of realized volatility. *Journal of Financial Econometrics*. 2009;7(2):174–196
- [8] Andersen T G, Bollerslev T, Diebold F X. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics*. 2007;89(4):701-720.
- [9] Black F, Scholes M. The pricing of options and corporate liabilities. *Journal of Political Economy*. 1973;81(3):637-654.
- [10] Britten-Jones M, Neuberger A. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*. 2000;55(2):839-866.
- [11] Fair R C, Shiller R J. Comparing information in forecasts from econometric models. *American Economic Review*. 1990;80(3):375-389.
- [12] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*. 1982;50(4):987-1008.
- [13] Bollerslev T, Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. 1986;31(3):307—327.
- [14] Hansen P R, Huang Zhuo. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*. 2012;27(6):877-906.
- [15] Clark P K. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*. 1973;41(1):135-156.
- [16] Müller U, Dacorogna M, Dav R, Pictet O, Olsen R, Ward J. Fractals and Intrinsic Time-A Challenge To Econometricians. XXXIXth International AEA Conference on Real Time Econometrics, 14-15 Oct 1993, Luxembourg.
- [17] Bollerslev T, Litvinova J, Tauchen G. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 2006;4(3):353–384.
- [18] Corsi F, Mittnik S, Pigorsch C, Pigorsch U. The volatility of realized volatility. *Econometric Reviews*. 2006;27(1-3):46-78.
- [19] Corsi F, Renò R. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modelling. *Journal of Business & Economic Statistics*. 2012;30(3):368-380.
- [20] Asai M, McAleer M, Medeiros MC. Asymmetry and long memory in volatility modelling. *Journal of Finance Econometrics*, 2012;10(3):495-512.
- [21] Gong X, Lin B. Modeling stock market volatility using new HAR-type models. *Physica A: Statistical Mechanics and its Applications*. 2019;516(C):194-211.
- [22] Hammoudeh S, Li H. Sudden changes in volatility in emerging markets: The case of Gulf Arab stock markets. *International Review of Financial Analysis*. 2008;17(1):47-63.
- [23] Ewing B T, Malik F. Volatility transmission between gold and oil futures under structural breaks. *International Review of Economics & Finance*, 2013;25(C):113-121
- [24] Gong X, Cao J, Wen F, Yang X. The HAR-type models with leverage and structural breaks and their applications to the volatility forecasting of stock market. *Systems Engineering - Theory & Practice*. 2020;40(5):1113-1133.

- [25] Huang X, Tauchen G. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*. 2005;3(4):456—499.
- [26] Corsi F, Pifmo D, Reno R. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*. 2010;159(2):276-288.
- [27] Gong X, Lin B. Structural breaks and volatility forecasting in the copper futures market. *Journal of Futures Markets*. 2018;38(3) 290-339. Available from: <doi.org/10. 1002/fut.21867>
- [28] Gong X, Lin B. Jump Risk, Structural Breaks and Forecasting Crude Oil Futures Volatility. *Chinese Journal of Management Science*. 2018;26(11):11-21.
- [29] Cai G, Xiang L. The Volatility Estimation and VaR Measurement of China's Copper Future Market: Based on Realized HAR GARCH Model Incorporating Generalized Realized Measures. *Chinese Journal of Management Science*. 2021; 29(11): 1-12.
- [30] Maki D, Ota Y. Impacts of asymmetry on forecasting realised volatility in Japanese stock markets. *Economic Modelling*. 2021;101(1):105533. Available from: <doi.org/10.1016/j.econmod.2021.105533>