MEDIAN REALIZED VARIATION AS A ROBUST VOLATILITY MEASURE FOR ESTIMATING HETEROGENEOUS AUTOREGRESSIVE (HAR) MODEL IN THE PRESENCE OF ASYMMETRY, JUMPS AND STRUCTURAL BREAKS

OlaOluwa S. Yaya[1]
Olanrewaju I. Shittu
Femi J. Ayoola
Oluwamilare G. Olutayo

ABSTRACT

This paper investigated the performance of realized volatility measures for jumps on variants of HAR model which included the newly proposed HAR model with asymmetry, jumps, structural breaks and GARCH errors. The study used simulated stochastic diffusion process which mimicked intraday Financial returns and obtained realized volatility estimators that are robust to jumps, such as the Bi-power variation (BV), minimum realized variation (minRV) and medium realized variation (medRV). The results showed that medRV is the best volatility measure for heterogeneous market models such as the types considered in this work. This work has further represented $medRV_t$ as appealing jump-robust estimator of the integrated variance process for intraday (tick-by-tick) financial returns with jumps.

Keyword: GARCH; Heterogeneous market; intraday returns, Monte Carlo Simulation; Structural break

The authors lecture at Economic and Financial Statistics and Computational Statistics Units, Department of Statistics, University of Ibadan, Nigeria.

[1]Corresponding author. Email: os.yaya@ui.edu.ng.

1.INTRODUCTION

Despite the popularity of volatility of asset pricing in financial time series, there appears to be no unique and universally accepted definition to the term. Most empirical studies have considered it as unobservable variable and therefore employed a specified conditional heteroskedasticity model to estimate the latent volatility within the ARCH-GARCH frameworks. At times, the estimation of models for volatility poses much difficulty owing to convergence and unexpected results, which lead to wrong volatility measurements. So, an alternative way out is to construct an observable proxy for the latent volatility in the spirit of Realized Volatility (RV) of Andersen, Bollerslev, Diebold and Labys (2001). The increasing availability of intraday high frequency financial time series has gingered the interest of researchers towards studying the volatility of asset prices by using the RV estimators.

Under the RV measurement, there is the assumption that asset prices originate from a continuous diffusion process with Brownian motion. The realized return (RR), corresponding to the log-return in daily high frequency data is obtained by taking the sum over all the intraday returns while the RV estimator sums squared of non-overlapping intraday returns in order to obtain the daily variance (Andersen et al., 2001). In the absence of jumps, the RV is a consistent estimator of the integrated variance (IV). In practice, price jumps experienced in realized volatility measurements tend to violate the assumption of a continuous diffusion process for asset price. Empirically, the impact of these jumps is temporary in asset pricing and risk management (Duffie et al., 2001). A number of studies have therefore shown that jump continuous stochastic process fits time series data better than does only a continuous process (Eraker et al., 2003; Maheu & McCurdy, 2004). Thus, the RV is weak in the sense that it is too sensitive to microstructure noise in the form of jump when applied to very high frequency data (Andersen et al., 2001). Recent works have proposed the adoption of consistent estimators of IV that are robust to jumps. The bi-power variation of Barndorff-Nielsen & Shephard (2004) and the minimum (minRV) and median realized variance (medRV) estimators of Andersen, Dobrev and Schaumburg (2012) are examples of such proposals.

Based on the framework of RV measurement, one obtains an alternative series similar to that realized from the conditional volatility model that is able to account for all the main features observed in the data. However, in the presence of abrupt jumps (structural breaks), the cumulated sum of products of squared daily returns may not serve as a consistent estimate of integrated volatility (Barndorff-Nielsen & Shephard, 2004). Thus, ignoring the presence of structural breaks may lead to misleading estimates, wrong inference and unreliable forecasts. In the analysis of realized volatility, one can control for the effect of structural breaks by using a jump-robust volatility estimator or embed the structural break feature in the time series models. These models include the

Heterogeneous Autoregressive (HAR) model and HAR-GARCH model of Corsi et al. (2008) and Corsi (2009), HAR-Jump model of Andersen, Bollerslev & Diebold (2007), among others. Generally, the HAR modelling framework assumes that financial markets are made up of heterogeneous market players with short trading horizon investment (noise traders and speculators), medium trading horizon investment (portfolio managers and hedge fund managers) and long term trading horizon investment (long term portfolio managers and pension fund managers). The setup of the model is in line with the Heterogeneous Market Hypothesis (HMH) recommended by Muller et al. (1993) & Dacorogna et al. (2001). Under the modelling framework, the time varying market liquidity can be determined based on the dominant investment horizons. For example, both the short and long investment horizons investors have different sentiments about market participation. A negative inflow of news may be an indicator for short horizon investors to sell an asset, while this might be an avenue for long horizon investor to buy it, and vice versa. If there is a balance between buying and selling, the financial market is can then be in a state of equilibrium. Meanwhile, in a time of economic crisis, this equilibrium is affected, and long horizon investors tend to quit or become short horizon investors. Another interesting attributes of HAR modelling is the realization of long memory volatility for different investment horizon activities. This long memory property diminished during the period of financial turmoil. Thus, HAR modelling framework is becoming another avenue to study volatility and market participation at different investment horizons.

The HAR model mimics the asymmetric propagation of volatility due to the presence of heterogeneous market participation. It is an additive cascade model of different volatility components each of which is generated by actions of different types of market players. This additive volatility cascade generates a simple long AR-type model, as it considers averages of realized measures over different time horizons. In most common HAR specifications, the actual log-realized variance is regressed on its past daily, weekly and monthly averages, together with a leverage term capturing the asymmetric relation between returns and volatility (see Martens et al. 2009), among others. The main advantage of the HAR-type specifications is its estimation simplicity, given that the model can be estimated by ordinary least squares estimation method.

Corsi (2009) employed realized volatility estimators to analyse, model and forecast the time series behaviour of high frequency exchange rate volatility. By proposing the HAR model based on realized volatility measurements, an AR-type model for intraday volatilities is obtained. Surprisingly, in spite of the model simplicity, and the fact that it does not formally belong to the class of long memory models, the HAR-RV model was able to reproduce the same memory persistence observed in volatility as well as many of the other stylized facts of financial data.

The challenge now is how to determine a consistent IV estimator for jumps in the realized volatility estimation of HAR-type model. Noting the role of jumps in asset pricing, large jumps reflect the arrival of surprising news of abnormal large magnitudes (Dobrev, 2007). It was for this reason that the author considered the adoption of jump-robust realized volatility measurements.

This paper investigated the performance of realized volatility measures on variants of HAR model. Specifically, the methodology of the paper involved the inclusion of a new HAR-GARCH model with asymmetry and multiple structural break dummies. The essence was to check the extent to which RV measures capture asymmetry, jumps and structural breaks in the modelling process defined by HAR-GARCH framework. The paper considered the non-jump robust estimator of integrated volatility $\left(RV_t\right)$, jump-robust estimators of integrated volatility, that is, the bi-power variation $\left(BV_t\right)$ and the minimum $\left(\min BV_t\right)$ and median $\left(\text{med}\, BV_t\right)$ which are the nearest neighbour estimators of integrated volatility (Andersen, Dobrev & Schaumburg, 2012). For structural break modelling, the paper applied the sequential multiple structural break test of Bai & Perron (2003) which allows for up to five structural breaks in a given time series.

The rest of the paper contains four sections: Section 2 gives the theoretical framework. Section 3 presents the statistical methodology and description of the Monte Carlo experiment. Section 4 presents results and discussion, while Section 5 renders the conclusion to the paper.

2.THEORETICAL FRAMEWORK

The stochastic diffusion process for logarithmic price, $P(t)$ of an asset is given as,

$$dP(t) = \mu(t)dt + \sigma(t)dW(t), \quad 0 \le t \le N \tag{1}$$

where N is the size of each time process, $\mu(t)$ is the drift, $\sigma(t)$ is the point-in-time or spot volatility and $W(t)$ is the standard Brownian motion. The $\mu(t)$ and $\sigma(t)$ are time-varying, but are assumed to be independent of $dW(t)$. The change of logarithmic price is defined as the continuously compounded intraday returns of day *t* with sampling frequency *M* given as:

$$\begin{aligned} r_t &\equiv P(t) - P(t-1) \\ &= \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s) \\ &\equiv \ln P_{t,j} - \ln P_{t,j-1}, \quad j = 1,\dots,M-1, \end{aligned} \tag{2}$$

Thus, the quadratic variation process for a sequence of partition is equivalent to the integrated volatility as *N* tends to infinity, that is,

$$IV_t \equiv \sum_{i=1}^{M}\left(P_{ti}-P_{ti-1}\right)^2 = \int_{t-1}^{t}\sigma^2(s)\,ds \qquad (3)$$

Clearly, (2) implies that the integrated volatility is an ideal measure of volatility as it gives the consistent estimates of the realized volatility,

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2 \qquad (4)$$

as documented in Andersen & Bollerslev (1998). Recent empirical studies have shown that a continuous diffusion model such as (1) fails to explain fully some inherent properties of asset returns such as excess kurtosis because of jumps/structural breaks observed in returns series. Thus, jump diffusion stochastic volatility models have been proposed in the literature to overcome this deficiency (see Back, 1991; Laurent, 2013). The log-return series is believed to belong to Brownian Semi Martingale with Jumps (BSMJ) families of models. In the BSMJ, the diffusion component captures the smooth variation of the price process, while the jump component captures the large discontinuities in the observed prices. A BSMJ log-price diffusion process is given as,

$$dP(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \qquad 0 \le t \le N \qquad (5)$$

where, in addition to the parameters of the defined in (1), $dq(t)$ is a counting process with $dq(t)=1$ corresponding to a jump at time t, and $dq(t)=0$ at any other time, t. $\kappa(t)$ is the size of the corresponding jump. Jumps in asset pricing are often assumed to follow a Poisson probability law, which is a continuous-time discrete process with parameter l that governs the occurrence of the special event, which is known as the intensity or threshold parameter.

3.METHODOLOGY

3.1 Jump robust estimators of Integrated Variance

Following Andersen & Bollerslev (1998), in the absence of jumps, the integrated volatility IV_t is an ideal measure of volatility as it gives the consistent estimates of the realized volatility,

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2\ , \qquad (6)$$

where $r_{t,j}$ is the intraday log-returns at period *j* for day *t*, obtained by taking differences of logarithm of prices $P_{t,j}$ and $P_{t-1,j}$. In the presence of jumps, this realization may not give consistent estimates. Barndorff-Nielsen & Shephard (2004) indicated that in a family of BSMJ diffusion process, the normalized sum of products of the absolute value of contiguous process gives a consistent estimator for the IV_t. This is a bi-power variation since it involves the product of two absolute intraday returns. Thus,

$$BV_t = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{j=1}^{M}\left|r_{t,j}\right|\left|r_{t,j-1}\right| \quad (7)$$

Unlike the RV_t, BV_t is developed to be robust to jumps because it is formed from the product between two consecutive returns instead of the squared returns.

More recent jump-robust estimators of IV_t are proposed in Andersen, Dobrev & Schaumburg (2012). These are the minimum $(\min BV_t)$ and the median $(\text{med}\, BV_t)$ operators, computed based on the nearest neighbour truncation method. Thus,

$$\min RV_t = \frac{\pi}{\pi-2}\left(\frac{M}{M-1}\right)\sum_{j=1}^{M-1}\left[\min\left(\left|r_{t,j}\right|,\left|r_{t,j+1}\right|\right)\right]^2 \quad (8)$$

and,

$$\text{med}\, RV_t = \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{M}{M-2}\right)\sum_{j=2}^{M-1}\left[\text{med}\left(\left|r_{t,j-1}\right|,\left|r_{t,j}\right|,\left|r_{t,j+1}\right|\right)\right]^2 \quad (9)$$

that is, minRV_t allows for blocks of two intraday returns, while $\text{med}RV_t$ allows for blocks of three intraday returns with more information when $\text{med}\, RV_t$ is used since its structure allows for large returns. Also, $\min RV_t$ suffers for exposure to zero return in finite sample test as in BV_t (Mancini & Calvori, 2012).

3.2 HAR Models

Here, we survey variants of HAR model for realized volatility aiming at modelling the forecasting power of jumps, leverage effects (asymmetry) and structural breaks in IV estimators for jumps in realized volatility measurements. Following Corsi, et al. (2005) and Corsi (2009), the HAR-GARCH model for integrated volatility, $Y_t = \log(IV_t)$ is given as,

$$Y_t = \phi_0 + \phi_D Y_{t-1} + \phi_D Y_{t-1}^W + \phi_D Y_{t-1}^M + \varepsilon_t \quad (10)$$

where $\varepsilon_t | I^{t-1} \sim N(0, \sigma_t^2)$ and $X_t^W = \frac{1}{5}\sum_{j=0}^{4} Y_{t-j}$ and $X_t^W = \frac{1}{22}\sum_{j=0}^{21} Y_{t-j}$ represents the weekly and monthly volatility components, respectively. The conditional variance series ε_t is assumed to follow GARCH(1,1) process given as,

$$r_t = \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim N(0,1) \quad (11)$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (12)$$

where z_t follows the standardized Student-t distribution. Model parameters in (12) are conditioned as $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$ in order to ensure stationarity and positivity of the conditional variance series, σ_t^2. IV transformations $Y_t = IV_t$ and $Y_t = \sqrt{IV_t}$ can also be considered as in Andersen, Bollersley & Diebold (2007), Corsi (2009) & Corsi et al. (2010), whereas, the logged transformation has the advantages of imposing positivity constraints and of improving the normality of the measured variable (Goncalves & Meddahi, 2011).

By introducing asymmetry and structural break components, we obtained the Asymmetric HAR-Structural break-Volatility (A-HAR-StB-V) model specification,

$$Y_t = \phi_0 + \phi_D Y_{t-1} + \phi_W Y_{t-1}^W + \phi_M Y_{t-1}^M + \gamma r_{t-1} I(r_{t-1} < 0) + \sum_{i=1}^{k} D_i B_{it} + \varepsilon_t \quad (13)$$

where the asymmetric component, driven by $I(\cdot)$ is an indicator function that takes value 1 when the daily realized return (RR), $r_{t,j}$ $(j = 1,...,M)$ is negative, and value 0 when the return is positive.[1] Thus, the parameter γ measures the leverage effects. The structural break component is introduced using dummy variables $B_{it}(i = 1,...,k)$, where k, the maximum number of breaks is determined endogenously from the series using Bai & Perron (BP, 2003) multiple structural break test. From each dummy variable, $B_{it} = 1$ if $t \geq T_{Bi}$ where T_{Bi} is the break date, and $B_{it} = 0$, otherwise. Thus, D_i is the break dummy coefficient.

This model is actually presented in four parts, that is the autoregressive (AR), asymmetry component, structural breaks and time varying conditional variance process for ε_t. In the AR part, $\hat{\phi}_0 = \bar{X}$ is to be estimated along with

[1] Daily realized return is the sum of all intraday log-returns.

other parameters, ϕ_D, ϕ_W and ϕ_M for daily, weekly and monthly effects, respectively.

The challenge now is how to determine break dates based on the structural break test. The BP multiple structural break test allows up to five structural breaks to be determined in a time series. The test involves a sequential application of $\sup F_T(l+1|l)$ test in selecting the number of breaks in the time series. It follows this procedure:

(i). We search for the first break date based on the maximum absolute t-value of the break dummy coefficient D_1 to obtain the first break date:

$$\hat{T}_{B1} = \arg_{\hat{T}_{B1}} \left| t_{\hat{D}_1}(T_{B1}) \right|; \tag{14}$$

(ii). By imposing the estimated break date $\hat{T}_{B1}$ in (i), we estimate the second break date $\hat{T}_{B2}$ as,

$$\hat{T}_{B2} = \arg_{\hat{T}_{B2}} \left| t_{\hat{D}_2}\left(\hat{T}_{B1}, T_{B2}\right) \right|. \tag{15}$$

(iii). This process is repeated by imposing the estimated break dates thereby increasing the number of breaks, *l* sequentially until the test $\sup F_T(l+1|l)$ fails to reject more null hypothesis of additional structural breaks.

For consistency, we determine the total number of breaks (NSB≦5), and apply these sequential break dates in the dummy variables B_{it} in (13).

3.3 Monte Carlo Simulation Experiment

This research work applied only Monte Carlo simulation experiment to validate the HAR-GARCH modelling framework using the realized volatility measurements. Recall that Realized Volatility measurement, as part of Time Series Econometrics of Volatility requires intra-day time series which is scarce to obtain. But, in few years to come, such intra-day data would have been made freely available to every time series expert. Thus, the proponent of HAR modelling have considered stochastic process as their Data Generating Process (DGP).

The DGP considered in this work was the continuous stochastic jump process, given simultaneously by the equations below:

$$dP(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t); \quad (16)$$

$$d\sigma^2(t) = \theta\left[\omega - \sigma^2(t)\right]dt + (2\lambda\theta)^{1/2}\sigma^2(t)dW(t); \quad (17)$$

$$\kappa(t) \sim \sigma(t)\sqrt{m}\left([-2,1]\cup[1,2]\right) \text{ and,} \quad (18)$$

$$dq(t) \approx Poisson(l) = \frac{l^m t^m}{m!}\exp(-lt) \quad (19)$$

where $\omega = 0.635$ $\theta = 0.035$ and $\lambda = 0.296$. $\kappa(t) = 0.2$, that is, the product of $\sigma(t)$ and uniformly distributed random variable in (16). The parameter m which determines the magnitude of the jumps was set as *m* = 0.3, 0.5, 0.7, and 1.0 for each break subsample 1175, 1350, 1240 and 750, respectively. These subsamples were then combined to give the overall sample size as 4515. The parameter l was chosen such that there was an average of one jump in every 5 days, and in this case, it was set as $l = 100$. The conditional variance part of the Data Generating Process (DGP), $\sigma^2(t)$ was simulated based on GARCH (1, 1) process. Thus, a total of 288 intraday returns, for 4515 sample size were simulated using the DGP in (16-18) after discarding the first 10 samples which controlled for initialization problem in each subsample.

Having realized the time structured datasets from the DGP in (16-19), the analysis of the time series based on the methodology reviewed in this work started from the exploratory data analysis (EDA). This included the plotting of the time series (Time Plot) and other descriptive measurements. The generation of the realized volatility series, integrated variance series and tests for jumps were also carried out, while model estimation and diagnostic tests (residual analysis) were done using the least square routine included in the software.

4.RESULTS AND DISCUSSION

4.1 Time Plots and Data Description

Time series of RV_t, BV_t, $MinRV_t$ and $MedRV_t$ were obtained from Monte Carlo experiment described in Section 3. In order to obtain time series that are closed to normality and with moderate kurtosis, we obtained the logarithm of each of the series as well. In Figures 1-4, we presented plots of each of the realized volatility measures, as well as their logarithmic transformations. The level series of each realized variance series is plotted on the left panel while that of the logarithmic transformation is plotted on the right panel. As expected, magnitude of jumps increased towards the end of the sample as observed from the level series but logarithmic transformation made the series to be more stationary in Figure 1.

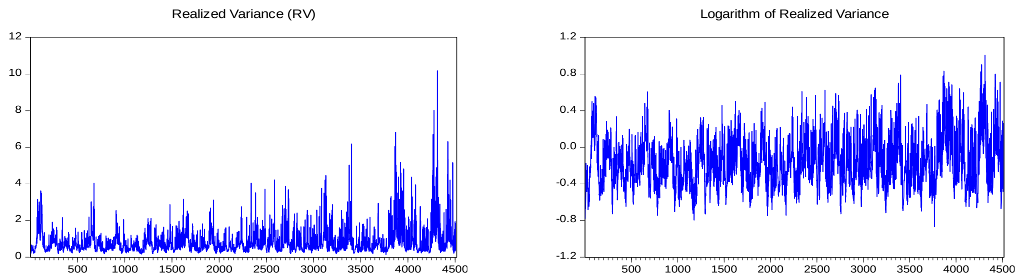

Figure 1: Realized Variance (RV)

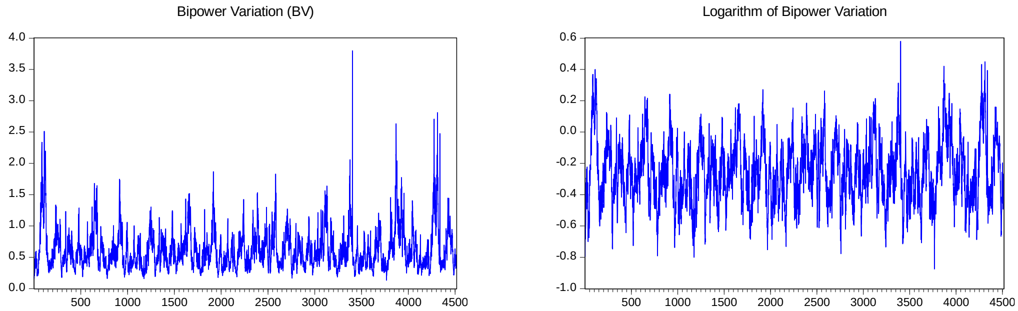

Figure 2: Bipower Variation (BV)

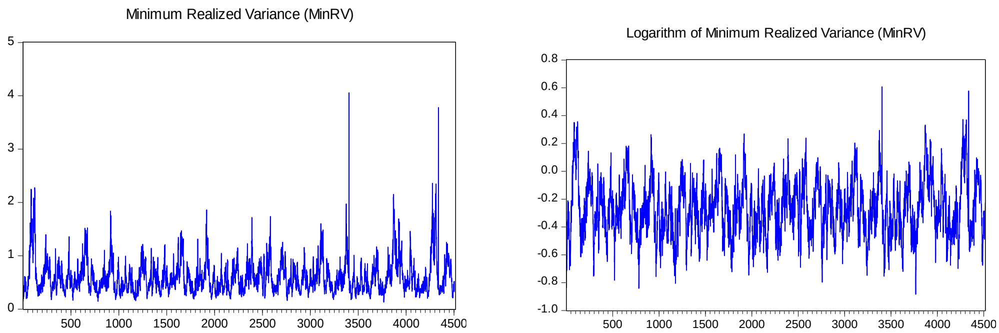

Figure 4: Minimum Realized Variance (MinRV)

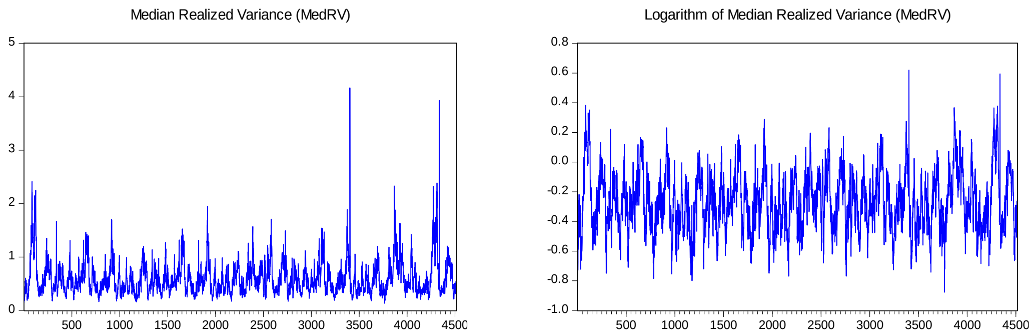

Figure 4: Median Realized Variance (MedRV)

In Figures 2-4, we observed that level series of BV_t, $MinRV_t$ and $MedRV_t$ mimicked one another and there are more significant jump spikes, more conspicuous than that of RV_t. As part of the EDA, we carried out exploratory data analysis on the simulated data and it was discovered that the RV series and the obtained series from other jump robust estimators of IV became more stabilized after taking their natural logarithms. The logarithms of each of these series were then used for the purpose of the analysis. The descriptive measurements on $Log(RV_t)$, $Log(BV_t)$, $Log(minRV_t)$ and $Log(medRV_t)$ are presented here. Mean and median values of the volatility measures are negative in values since they are logarithm of averages for stochastic prices lying between 0 and 1, that is, squared returns. Using the Jarque-Bera statistic, it is observed that the series [Log(RV_t), Log(BV_t), Log($minRV_t$) and Log($medRV_t$)] showed non-normal properties (deviation from normality). The RV_t series, particularly showed a relatively strong deviation (JB = 543.1) as there are more observations around the mean and in the tails i.e. leptokurticity is confirmed (kurtosis was found to be greater than 3). Those jump-robust variations showed relatively smaller (slight) deviations from normality and platykurticity is observed (as the values of kurtosis were less than 3). Furthermore, all the series under study exhibited ARCH effects after the test was conducted (i.e. p-values were less than 5%) which implies that estimation of GARCH model was relevant.

Table 1: Descriptive Statistics in Logarithms of RV_t, BV_t, $minRV_t$ and $medRV_t$

	Log(RV_t)	Log(BV_t)	Log($minRV_t$)	Log($medRV_t$)
Mean	-0.471929	-0.590382	-0.616336	-0.613105
Median	-0.546932	-0.615588	-0.636823	-0.636701
St.Dev	0.582210	0.464776	0.463355	0.461192
Skewness	0.755388	0.288284	0.279034	0.294096
Kurtosis	3.777481	2.925647	2.984098	2.996233
JB	543.1020***	63.57875***	58.63734***	65.08825***
ARCH(5)	113.2336***	1437.053***	1254.183***	1358.581***

*** indicates significance of JB and ARCH test at 5% level

4.2 Results of Model Estimation

Table 2 presents the results of the estimated HAR models for the realized variance measurements. We observed significant daily, weekly and monthly volatility effects at 5% level, and the three models for integrated variance that are robust to jumps outperformed that of realized variance in terms of their abilities to capture well the current realized volatility. We made this judgment based on minimum information criteria (AIC and SBIC), since each of the information criteria of models for Log(BV_t), Log($minRV_t$) or Log($medRV_t$) is less than corresponding information criteria for Log(RV_t). The HAR model for Log($medRV_t$) emerged the best predicting model, followed by that of Log(BV_t) and Log($minRV_t$). As part of the model diagnosis, we obtained Q(5) and $Q^2(5)$, that is estimates of serial correlations of the residuals up to 5 lags, and found these estimates to be highly significant. Thus, there was the need to extend these models to capture salient properties in financial modelling such as asymmetry, heteroscedasticity in residuals and possible structural breaks.

Table 2: Estimated Parameters of HAR model

Estimated parameters	Log(RV_t)	Log(BV_t)	Log($minRV_t$)	Log($medRV_t$)
$\hat{\phi}_0$	-0.030414***	-0.020006***	-0.021445***	-0.019890***

$\hat{\phi}_D$	0.196203***	0.673243***	0.654903***	0.713276***
$\hat{\phi}_W$	0.57669***	0.273777***	0.295223***	0.237665***
$\hat{\phi}_M$	0.077413***	-0.025513**	-0.030724***	-0.026101***
LogL	1159.116	4550.816	4511.455	4779.540
AIC	-0.514185	-2.023955	-2.006435	-2.125769
SBIC	-0.508478	-2.018248	-2.000728	-2.120062
Q(5)	40.476***	66.448***	61.946***	47.166***
$Q^2(5)$	52.205***	226.48***	317.14***	467.35***

***, ** indicate significance of parameters at 5% and 10% level, respectively.

We considered next the case of asymmetric HAR (A-HAR) model, with the results presented in Table 3. The A-HAR model also exhibited daily, weekly and monthly volatility effects for each of the series considered. Significant asymmetric effect is also observed across the four models. Model performance is still ranked in the same order as given in Table 2.

Table 3: Estimated Parameters of A-HAR model

Estimated parameters	Log(RV_t)	Log(BV_t)	Log($minRV_t$)	Log($medRV_t$)
$\hat{\phi}_0$	-0.030400***	-0.019942***	-0.021369***	-0.019815***
$\hat{\phi}_D$	0.196126***	0.673685***	0.655626***	0.714011***
$\hat{\phi}_W$	0.576978***	0.274293***	0.295484***	0.237898***
$\hat{\phi}_M$	0.077213***	-0.026338**	-0.031550***	-0.026904***
$\hat{\gamma}$	-0.001488	-0.005665***	-0.06213***	-0.005785***
LogL	1159.176	4554.743	4516.098	4784.074
AIC	-0.513766	-2.025258	-2.008056	-2.127342
SBIC	0.506633	-2.018125	-2.000922	-2.120209
Q(5)	44.039***	59.562***	59.319***	38.583***
$Q^2(5)$	5.2407	8.5031	6.9572	8.7050

***, ** indicate significance of parameters at 5% and 10% level, respectively.

Now, by carrying out Bai and Perron (2003) multiple structural break test on all the considered IV estimators, we obtained the results presented in Table 6 at the appendix. After imposing the condition that the maximum number of structural breaks is 5, the Log(RV_t) series exhibited 4 significant structural breaks, other three integrated volatility estimators exhibited structural breaks (StBs) at 5 different points in time, in the series. Since structural break effects are present in all the series, we then proceeded to estimate another HAR variants that included structural break dummies and GARCH errors. Thus, this is the Asymmetric-HAR-StB-GARCH model.[2] This time around, model diagnostic measures are presented in this paper for all HAR-type models with GARCH errors, and the results presented in Table 4.

[2] Full models parameter estimation results are available upon request.

From Table 4, Log(medRV$_t$) volatility measurements indicated HAR-type models with minimum information criteria, compared with the corresponding models for the three other realized volatility measures. The results further showed wider performance gap from non-jump robust RV [Log(RV$_t$)] and the three jump IV estimators in terms of model fitness (AIC and SBIC), and forecasts performance (RMSFE). Now, focusing on the three jump IV estimators, Log(BV$_t$) is ranked next to Log(medRV$_t$), then, next is Log(minRV$_t$).

Table 4: Summary of Diagnostic test and Forecasts Evaluation

	Log(RV$_t$)				Log(BV$_t$)			
Model	**LogL**	**AIC**	**SBIC**	**RMSFE**	**LogL**	**AIC**	**SBIC**	**RMSFE**
A-HAR-StB-GARCH(1,1)	1330.4	-0.5869	-0.5697	0.2480	4585.0	-2.0352	-2.0166	0.1970
A-HAR-GARCH(1,1)	1328.2	-0.5877	-0.5763	0.2536	4581.5	-2.0358	-2.0244	0.2017
A-HAR	1159.2	-0.5138	-0.5066	0.2523	4554.7	-2.0253	-2.0181	0.2017
HAR	1159.1	-0.5142	-0.5085	0.2522	4550.8	-2.0240	-2.0182	0.2011
	Log(minRV$_t$)				**Log(medRV$_t$)**			
Model	**LogL**	**AIC**	**SBIC**	**RMSFE**	**LogL**	**AIC**	**SBIC**	**RMSFE**
A-HAR-StB-GARCH(1,1)	4555.36	-2.0220	-2.0034	0.1968	***4843.23***	***-2.1501***	***-2.1316***	***0.1959***
A-HAR-GARCH(1,1)	4551.84	-2.0226	-2.0112	0.2012	***4840.28***	***-2.1510***	***-2.1396***	***0.2003***
A-HAR	4516.10	-2.0081	-2.0009	0.2012	***4784.07***	***-2.1273***	***-2.1202***	***0.2002***
HAR	4511.46	-2.0064	2.0007	0.2005	***4779.54***	***-2.1258***	***-2.1201***	***0.1996***

To check for possible influence of GARCH-types and model orders on model performance, we considered EGARCH model of Nelson (1991) and APARCH model of Ding et al. (1993). These are asymmetric GARCH models. Firstly, we found order (1,1) as optimal order capturing serial correlations in the residuals by means of GARCH, EGARCH and APARCH modelling of HAR residuals. Table 5 presents the results of model diagnosis, as we observe Log(medRV$_t$) as the best performing IV estimator.

Table 5: Robustness to EGARCH(1,1) and APARCH(1,1) Errors

	Log(RV$_t$)				Log(BV$_t$)			
Model	**LogL**	**AIC**	**SBIC**	**RMSFE**	**LogL**	**AIC**	**SBIC**	**RMSFE**
A-HAR-StB-EGARCH(1,1)	1335.92	-0.5893	-0.5722	0.2507	4592.77	-2.0391	-2.0219	0.1989
A-HAR-EGARCH(1,1)	1335.20	-0.5903	-0.5775	0.2523	4590.67	-2.0395	-2.0266	0.2021
A-HAR-StB-APARCH(1,1)	1329.10	-0.5858	-0.5673	0.2485	4594.23	-2.0393	-2.0207	0.1987

A-HAR-APARCH(1,1)	1329.14	-0.5872	-0.5729	0.2526	4591.67	-2.0395	-2.0252	0.2019
		Log($minRV_t$)				**Log($medRV_t$)**		
Model	**LogL**	**AIC**	**SBIC**	**RMSFE**	**LogL**	**AIC**	**SBIC**	**RMSFE**
A-HAR-StB-EGARCH(1,1)	4603.24	-2.0433	-2.0247	0.1982	***4839.73***	***-2.1490***	***-2.1319***	***0.1989***
A-HAR-EGARCH(1,1)	4548.68	-2.0208	-2.0079	0.2013	***4838.15***	***-2.1496***	***-2.1368***	***0.2005***
A-HAR-StB-APARCH(1,1)	4615.30	-2.0482	-2.0282	0.1982	***4843.55***	***-2.1503***	***-2.1317***	***0.1985***
A-HAR-APARCH(1,1)	4557.21	-2.0241	-2.0099	0.2012	***4842.31***	***-2.1510***	***-2.1368***	***0.2003***

5.CONCLUDING REMARKS

We have considered IV estimators for realized volatility in the presence of stochastic jumps in this paper. Having considered a stochastic diffusion process in the simulation of RV_t, BV_t, $minRV_t$ and $medRV_t$, the preliminary analysis indicated smaller and consistent skewness and kurtosis values for BV_t, $minRV_t$ and $medRV_t$, and in modelling, similar results were obtained with glaring ranking among the three. In terms of model fitness and forecast ability, we found Log($medRV_t$) as the best realized volatility measure for HAR modelling in the presence of asymmetry, jumps and structural breaks.

Jump variation is an important factor in forecasting future volatility, with volatility attributable to negative jumps leading to significantly higher future volatility, while positive jumps lead to significantly *lower* volatility. Actually, jumps are of limited use in forecasting future volatility using HAR models but assessing the usefulness of realized volatility via jump realized variations in financial applications such as portfolio management and pricing may be an interesting area of future research spurred by findings in this work. In this regards, other financial models apart from HAR models can be considered. Finally, this work has further represented $medRV_t$ as appealing jump-robust estimator of the integrated variance process for intraday (tick-by-tick) financial returns with jumps. Though we were limited to simulated intraday returns but we hope to be able to apply this methodology on intraday returns of financial series in the nearest future. This is left as future research.

Appendix:

Table 6: Result of Bai-Perron Multiple Structural Break

Estimator	No of StBs	Breaks dates (with sequential F-test statistic)				
		$\hat{T}_{B1}$	$\hat{T}_{B2}$	$\hat{T}_{B3}$	$\hat{T}_{B4}$	$\hat{T}_{B5}$
$Log(RV_t)$	4	680 (132.75)***	2318 (12.86)***	3116 (27.59)***	3793 (24.68)***	---- (11.93)^{NS}
$Log(BV_t)$	5	680 (80.52)***	1357 (39.62)***	2320 (24.09)***	3124 (28.93)***	3801 (28.15)***
$Log(minRV_t)$	5	684 (69.83)***	1357 (44.33)***	2322 (22.08)***	3125 (32.46)***	3802 (27.22)***
$Log(medRV_t)$	5	680 (69.40)***	1467 (43.40)***	2322 (22.02)***	3125 (32.11)***	3802 (27.58)***

Note, significant break dates based on Bai-Perron (2003) test are given in the table with corresponding sequential F test in parenthesis with critical values 8.58, 10.13, 11.14, 11.83 and 12.25 for break dates *B*1 to *B*5, respectively.

REFERENCES

Andersen, T.G. & Bollerslev, T. (1998). "Answering the skeptics: yes, standard volatility models do provide accurate forecasts". *International Economic Review*, 39, 885–906.

Andersen, T.G., Bollerslev, T., Diebold, F.X. & Labys, P. (2001). "The distribution of exchange rate volatility". *Journal of the America Statistical Association,* 96, 42-55.

Andersen, T.G., Bollerslev, T. & Diebold, F.X. (2007). "Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility". *Review of Economic Statistics*, 89, 701–720.

Andersen T.G, Dobrev D, Schaumburg E (2012). "Jump-robust volatility estimation using nearest neighbor truncation". *Journal of Econometrics*, 169(1), 75–93.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). "The distribution of stock return volatility". *Journal of Financial Economics*, 6, 43–76.

Back, K. (1991). "Asset pricing for general processes". *Journal of Mathematical Economics*, 20, 371-395.

Bai, J. & Perron, P. (2003). "Computation and analysis of multiple structural change models". *Journal of Applied Economics*, 18, 1-22.

Barndorff-Nielsen, O.E. & Shephard, N. (2004). "Power and bi-power variation with stochastic volatility jumps". *Journal of Financial Econometrics*, 2(1): 1-37.

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

Corsi, R., Mittnik, S., Pigorsch, C. & Pigorsch, U. (2008). "The volatility of realized volatility". *Econometric Reviews,* 27, 46-78.

Corsi, F., Vecchia, La & Fusari, N. (2010). "Realizing Smiles: Pricing options with realized volatility". *Journal of Financial Economics* (forthcoming).

Corsi, F., Pirino, D. & Reno, R. (2010). "Threshold Bipower variation and the impact of jumps on volatility forecasting". *Journal of Econometrics*, 159, 276-288.

Dacorogna, M., Ulrich, M., Richard, O. & Oliveier, P. (2001). "Defining efficiency in heterogeneous markets". *Quantitative Finance*, 1, 198–201.

Ding, Z., Granger, C. W. J. & Engle, R. F. (1993). "A long memory property of stock market returns and a new model". *Journal of Empirical Finance*, 1, 83-106.

Dobrev, D. (2007). "Capturing volatility from large price moves: generalized range theory and applications". Working paper, Kellog School of Management, Northwestern University.

Duffie, D., Pan, J., & Singleton, K. (2001). "Transform analysis and asset pricing for affine jump diffusions". *Econometrica*, 68,1343–1376.

Eraker, B., Johannes, M. S., & Polson, N. G. (2003). "The impact of jumps in returns and volatility". *Journal of Finance*, 53(3),1269–1300.

Goncalves, S. & Meddahi, N. (2011). "Box-Cox transforms for realized volatility". *Journal of Econometrics*, 160, 129-144.

Laurent, S. (2013). *Estimating and forecasting ARCH models using G@RCH 7.* London: Timberlake Consultants Press.

Maheu, J. M. & McCurdy, T. H. (2004). "News arrival, jump dynamics, and volatility components for individual stock returns". *Journal of Finance,* LIX,755–793.

Mancini, C. & Calvori, F. (2012). "Jumps". In L. Bauwens, C. Hafner & S. Laurent (Eds.). *Handbook of Volatility models and their Applications.* 403-445. London, John Wiley & Sons, UK.

Martens, M., Van Dijk, D., & De Pooter, M. (2009). "Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements". *International Journal of Forecasting*, 25, 282–303.

Muller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O. & Ward, J. (1993). "Fractals and intrinsic time – a challenge to econometricians". *34th Iinternational AEA conference on real time econometrics*, 14–15.

Nelson, D.B. (1991). "Conditional heteroscedasticity in asset returns: A new approach". *Econometrica* 59, 347-370.