3. HETEROGENEOUS MARKET HYPOTHESIS EVALUATIONS USING VARIOUS JUMP-ROBUST REALIZED VOLATILITY

Chin Wen CHEONG[1]
Lee Min CHERNG[2]
Grace Lee Ching YAP[3]

Abstract

The availability of high frequency data has promoted the usage of realized volatility as the unobservable latent volatility in financial markets. However, the traditional realized volatility (RV) representation is not robust to abrupt jumps in nowadays volatile globalized financial markets. This study includes other alternatives of jump-robust realized volatilities namely the bipower, minimum and median nearest neighbor truncation (NNT) volatility proxies in the examination of the heterogeneous market hypothesis (HMH) through the extension of heterogeneous autoregressive (HAR) model specifications. The empirical results show that the aforementioned alternative realized volatilities provide better forecast evaluations as compared to the standard realized volatility. Thus, the alternative realized volatility proxies are better explained the heterogeneous market hypothesis. In addition, the combination forecast models using three weighting schemes indicated better forecast performance as compared to the individual forecast. To complete this study, we illustrate a value-at-risk determination for the emerging Brazilian stock exchange.

Keywords: nearest neighbor truncation estimation, heterogeneous market hypothesis, realized volatility, heterogeneous autoregressive models, value-at-risk

JEL Classification: C22, C52, C58, G14

[1] *Faculty of Management, SIG Quantitative Economics and Finance, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia. E-mail: wcchin@mmu.edu.my*

[2] *Department of Mathematical and Actuarial Science, University Tunku Abdul Rahman, 43300 Kajang, Selangor, Malaysia.*

[3] *Faculty of Engineering, University of Nottingham (Malaysia Campus), 43500 Semenyih, Selangor, Malaysia.*

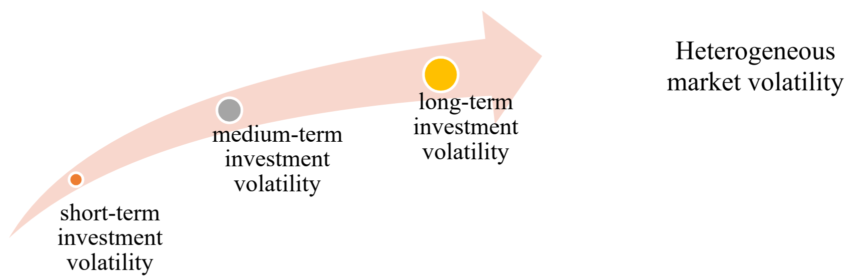
I. Introduction

For the past several decades, the informational efficient market hypothesis (EMH) has been intensively studied theoretically and empirically (Fama, 1998; Malkiel, 2003) using financial markets data. In an ideal efficient market, the market prices reflect all the relevant market information; hence there will be no investors that are able to beat the markets even using any financial strategies such as optimal asset selections or market timing strategy.There are two major approaches that can be used to improve the analysis of EMH. These include new definitions of EMH in terms of theoretical framework as well as empirical methodologies as the yardstick to either support or concluded contradictory against the EMH. For example, chaos theory (Mandelbrot, 2005) and behavioral finance theory (Shiller, 2006) have been used to further explain the traditional EMH. Some of the new definitions that complement the classic EMH are such as fractal market hypothesis (Peters, 1994), heterogeneous market hypothesis (Muller *et al*., 1993; Dacorogna, 1998) and adaptive market hypothesis (Lo, 2005).

Heterogeneous market hypothesis (HMH) is among the new concepts that suggested non-homogeneous market participants in the market efficiency literature. The empirical study had been conducted by Muller *et al*. (1993) in FOREX and stock markets by Dacorogna *et al*. (2001). Instead of homogeneity among market participants, the HMH claimed that the heterogeneity of market participants interpreted same information in different ways according to their trading preferences and opportunities. This heterogeneity has created an additive volatility with various different trading activities duration such as short, medium and long term investments. In other words, a financial market is composed by investors with various investment strategies ranging from short to long durations. The combinations of these various duration volatilities have produced the long memory property in financial markets. Besides the HMH concept, the fractionally integrated (Andersen *et al*., 2006) ARMA approach is also often used to capture the long memory. However, this study does not include this approach because it is more to a mathematical model that without any theoretical financial. For graphical illustration, Figure 1 shows the structure of heterogeneous market volatility.

Figure 1

Structure of Heterogeneous Market Volatility



Based on the HMH structure, the combination of volatilities can be constructed using an additive hierarchical structure of various duration investment volatilities. The HMH heterogeneity has been studied with different approaches by researchers such as Lux and Marchesi (1999), Andersen and Bollerslev (1997), Muller *et al*. (1997), Cheong *et al*. (2007), Corsi *et al*. (2008) and Corsi (2009). Most of the aforementioned studies are conducted using high frequency data (or intraday data) which collected minutely from the daily trading activities in a specific financial market. With the heavy trading activities, financial markets are normally facilitated with information technology facilities which led to enormous amounts of intraday information for data analysis. After Andersen and Bollerslev (1998) and Blair *et al*. (2001) have shown that the high-frequency forecast provided better performance over the traditional daily forecast in foreign exchange and stock markets, the community of high-frequency researchers has expanded intensively over the years.

One of the very important empirical studies is conducted by Andersen and Bollerslev (1998). They estimated the latent volatility by cumulating the sum of products of return within a day or more commonly named as the realized volatility (RV). Some vital theoretical properties of RV are studied by Andersen *et al*. (2003) and Barndorff-Nielsen and Shephard (2002). Nevertheless the RV estimator is facing the biasness and inconsistency issues by the microstructure effect (Hansen and Lunde, 2006, Andersen *et al*., 2011). Besides, the RV has some issues (Barndorff-Nielsen and Shephard, 2004; Andersen *et al*., 2012) when abrupt jumps occurred in the financial markets. In order to overcome this shortcoming, Barndorff-Nielsen and Shephard (2004) has introduced the bipower variation volatility proxy with the cumulative sum of products of adjacent absolute returns. Although the bipower variation (BV) measure is able to lessen the noise which leads to more consistent estimation, it is still sensitive and bias to the presence of very small returns. Alternately, two jump-robust estimators are proposed by Andersen et al. (2010) using the nearest neighbor truncation approach to battle the estimation issue. The first volatility estimator, namely the minimum realized volatility (minRV), is constructed by scaling the square of the minimum of two consecutive absolute returns. With the presence of jump during an interval, the minRV will eliminate it and compute based on the adjacent diffusive returns. Again, minRV is also sensitive to very small returns and leads to efficiency issue. Consequently, to improve the robustness to jump, the latter estimator, median realized volatility (medRV) uses the median operator to square the median of three consecutive absolute returns. In other words, the minimum and median operator intended to eliminate the noise of the volatility.

For this specific study, we intend to re-examine the HMH using a variety of RV estimators through the autoregressive heterogeneous model (Corsi, 2009). Unlike prior studies using realized volatility only, we have included the BV, minRV and medRV as the jump-robust volatility proxies under the assumption of heavy-tailed with student-t distributed innovations. Thus, this study attempts to add the empirical literature of EMH by using various standard and jump-robust volatility estimators in the HMH. Using a more robust volatility estimator should help explain and model the HMH in a better way. The remaining of this research is organized as follows: Section 2 explains the formation of RV, BV, minRV and medRV and the modified heterogeneous autoregressive models; Section 3 discusses the Brazilian stock exchange data and results and finally, Section 4 concludes and summarizes the study.

II. Methodology

High frequency integrated volatility estimation is widely used to measure the latent financial volatility which cannot be directly observed from the raw data. The high frequency data consist of more trading information as compared to daily closed data and have significant impact to the accuracy in portfolio analysis and risk management. From the efficient market hypothesis analysis point of view, availability of high frequency data provides further advantages in the empirical study of informational efficiency. For this particular study, we attempt to explore the HMH using various high frequency volatility estimators which are robust to jumps and market micro structural noise. For the empirical study, we have selected the Brazilian stock exchange.

For one day interval high frequency data, the continuously compounded intraday returns of day *t* with *N* observations is defined as $r_{t,j} = 100\left(\ln P_{t,j}^{close} - \ln P_{t,j-1}^{close}\right)$ where $j = 1,\ldots,N$ and $t = 1, \ldots, T$. Hence, for 5-minute interval daily observation consists of $N\times\delta = 78$ minutes with *N* equally-spaced subintervals. Whereas the daily closed return is defined as $r_t = 100\left(\ln P_t^{close} - \ln P_{t-1}^{close}\right)$. For high frequency volatility estimation, Andersen and Bollerslev (1998) accumulate the daily squared return as $\sigma_{RV,t}^2 = \sum_{j=1}^{N} r_{t,j}^2$. This estimator is well-known as realized volatility (RV) and converges uniformly in probability to the quadratic variation process as the sampling frequency approaches infinity, $\sigma_{RV,t}^2 \to \int_{t-1}^{t} \sigma^2(t)dt$. According to Barndorff-Nielsen and Shephard (2002), *RV* is a consistent estimator for integrated volatility (IV) in the absence of jump. Although high sampling frequency may reduce the RV's variance, it may increase its biasness component. Under the presence of abrupt jumps, the RV is no longer consistent estimate for IV. Due to this, Barndorff-Nielsen and Shephard (2004) has recommended a jump-robust estimator, Bipower variation (BV) volatility estimator to deal with this issue as follow:

$$\sigma_{BV,t}^2 = \frac{\pi}{2}\frac{t}{t-1}\sum_{j=1}^{t-1}\left|r_{t,j}\right|\left|r_{t,j+1}\right| \quad . \tag{1}$$

Although the BV is able to smooth the impact of jump by multiplying two consecutive returns, it is not able to reduce the magnitude of two consecutive jumps. Another issue of BV is its sensitivity and biasness to the presence of very small returns. In order to enhance these estimators, Andersen et al. (2010) proposed two estimators based on minimum (minRV) and median (medRV) operators based on the nearest neighbor truncation (NTT) approach as follows:

$$MINRV_t = \frac{\pi}{\pi-2}\frac{t}{t-1}\sum_{j=1}^{t-1}\left[min\left|r_{t,j}\right|, \left|r_{t,j+1}\right|\right]^2 \tag{2}$$

$$MEDRV_t = \frac{\pi}{6-4\sqrt{3}+\pi}\frac{t}{t-1}\sum_{j=2}^{t-1}\left[med\left|r_{t,j-1}\right|, \left|r_{t,j}\right|, \left|r_{t,j+1}\right|\right]^2 \tag{3}$$

The minimum realized volatility (minRV) will eliminate a jump for a given block of two consecutive returns and compute based on the adjacent diffusive returns whereas the median realized volatility (medRV) uses the median operator to square the median of three consecutive absolute returns. As a comparison, BV smoothes a possible jump whereas NTT estimators eliminate it from the block of returns. It is proven that (Andersen *et al*., 2010), the NTT estimators are more efficient and robust under the presence of jumps.

Intraday Volatility Model Using Jump-Robust Estimators

This study considers the fundamental heterogeneous autoregressive (HAR) model proposed by Corsi (2009). Following the HMH concept, the HAR current volatility consists of multiple past autoregressive components for daily, weekly and monthly volatilities. In order to accommodate the non-gaussianity and time-varying volatility in the RV, we have followed the model specification suggested by Cheong et al. (2007) and Corsi *et al*. (2008). Instead of using the standard RV only, this modified HAR includes the bipower variation and nearest neighbor truncated volatility estimators as the proxy of latent volatility. The specification of this jump-robust HAR-RV-GARCH(1,1) model is expressed as follows,

$$\ln(VOL_{i,t}^{day}) = \theta_{i,0} + \theta_{i,d}\ln(VOL_{i,t-1}^{day}) + \theta_{i,w}\ln(VOL_{i,t-1}^{week}) + \theta_{i,m}\ln(VOL_{i,t-1}^{month}) + a_{i,t}$$
$$a_{i,t} = \sigma_{i,t}\varepsilon_{i,t} \quad \varepsilon_{i,t}|\Omega_{t-1} \sim \text{student}-\text{t}(v),$$
$$\sigma(VOL)_{i,t}^2 = \alpha_{i,0} + \alpha_{i,1}a_{i,t-1}^2 + \beta_{i,1}\sigma(VOL)_{i,t-1}^2 \tag{4}$$

where: a_t follows a conditional density with time-varying RV with the HAR components $VOL_{t-1}^{week} = \frac{1}{5}\sum_{j=1}^{5}\ln(VOL_{t-j}^{day})$ and $VOL_{t-1}^{month} = \frac{1}{22}\sum_{j=1}^{22} ln(VOL_{t-j}^{day})$. The subscription i =1, 2, 3 and 4 indicates the standard RV, BV, MINRV and MEDRV respectively. The $\sigma(VOL)_{i,t}^2$ is interpreted as the volatility of RV (Corsi et al., 2008). Due to the non-gaussianity of financial time series, the error a_t is assumed to be followed a student-t with the density function

$$f_{student-t}(a_t \mid v) = \frac{\Gamma\left(\frac{v+1}{2}\right)1}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)}}\left(1+\frac{a_t^2}{v-2}\right)^{-\frac{v+1}{2}}, \ v>2 \tag{5}$$

where: *v* is the degree of freedom and Γ(·) is a gamma function. For 246 out-of-sample one-day ahead forecasts, the model is re-estimated every day based on a fix rolling sample of 1689 (1st February 2008 until 31st December 2014) starts from 1st January 2015 to 31st December 2015. The various one-day-ahead logarithmic RV forecasts are computed as follows:

$$\ln(VOL_{i,t+1|t}^{day}) = \theta_{i,0}^t + \theta_{i,d}^t\ln(VOL_{i,t}^{day}) + \theta_{i,w}^t\ln(VOL_{i,t}^{week}) + \theta_{i,m}^t\ln(VOL_{i,t}^{month})$$
$$\sigma(VOL)_{i,t}^2 = \alpha_{i,0}^t + \alpha_{i,1}^t a_{i,t}^2 + \beta_{i,1}^t\sigma(VOL)_{i,t-1}^2 \tag{6}$$

where: $VOL_t^{week} = \frac{1}{5}\sum_{j=1}^{5}\ln(VOL_{t-j+1}^{day})$ and $VOL_t^{month} = \frac{1}{22}\sum_{j=1}^{22} ln(VOL_{t-j+1}^{day})$. Consider the parameter vector to be estimated at each day t is $\Theta^{(t)} = (\theta_i^{(t)}, \alpha_i^{(t)}, \beta_i^{(t)})'$; therefore, the vector $\Theta^{(t)}$ is re-estimated every day for t = h, h+1,…,h+T-1 days.

Combination of Forecast Evaluations

For out-of-sample forecast evaluations, we have selected three loss function criteria namely the root-mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). Besides the individual model performance, we also include the combination forecast for all the four models using RV, BV, minRV and medRV. According to Timmermann (2006), combining forecasts into a single forecast can outperform the individual benchmark model. In this study the combination forecasts are based on the simple-average (SA), $W^r = \frac{1}{S}$, where every forecast is given the similar weight, Least Squares (LS) with the weights are estimated using ordinary least-squares regression (Granger and Ramanathan, 1984), $Y_{t+h} = W^0 + W^1Y_{t+h}^1 + W^2Y_{t+h}^2 + \cdots + W^SY_{t+h}^S + \epsilon_{t+h}$ and MSE ranks $W^r = \frac{1/RANK_{r,t+h}}{\sum_{l=1}^{S} 1/RANK_{l,t+h}}$, where the smallest MSE will has the rank 1 (Aiolfi and Timmermann, 2006). Assume that the *h*-step-ahead forecasts, Y_{t+h}^A can be composed as Y_{t+h}^r for $r = 1, \ldots, S$ using the aforementioned weight scheme as follow:

$$Y_{t+h}^A = W^1Y_{t+h}^1 + W^2Y_{t+h}^2 + \cdots + W^SY_{t+h}^S \qquad (7)$$

III. Empirical Study Using the Brazilian Stock Exchange Index

This study selects the emerging market Brazil BOVESPA index which consists of the top 381 active companies that serves as the barometer for Brazil economic performance. In year 2008, the Sao Paulo stock exchange and the Brazilian Mercantile and Future Exchange merged and established the BM&FOVESPA. The empirical data is collected from year 2008 to year 2015 with approximately 800,000 5-minutely data from trading hour 10.00 to 17.30. In order to construct the daily, weekly and monthly volatility components, the estimation is started from February 2008 and ended at December 2014 (1689 days). For forecast evaluations, we utilized the data from January 2015 until December 2015 (246 days).

Figure 2 indicates the plots for all the volatility estimators namely the standard realized volatility (RV), bipower variation (BV), nearest neighbor truncated minimum (minRV) and median realized volatility (minRV) respectively. It is found that the RV (y-axis with maximum scale 0.008) shows the noisiest estimator among the rest whereas BV, minRV and medRV have indicated similar magnitude over the analysis periods. This is because the smoothing (averaging) process by BV and eliminations by minRV and medRV has lessened the fluctuations of the estimated volatility.

Figure 2

Plots of Various Realized Volatility Estimators

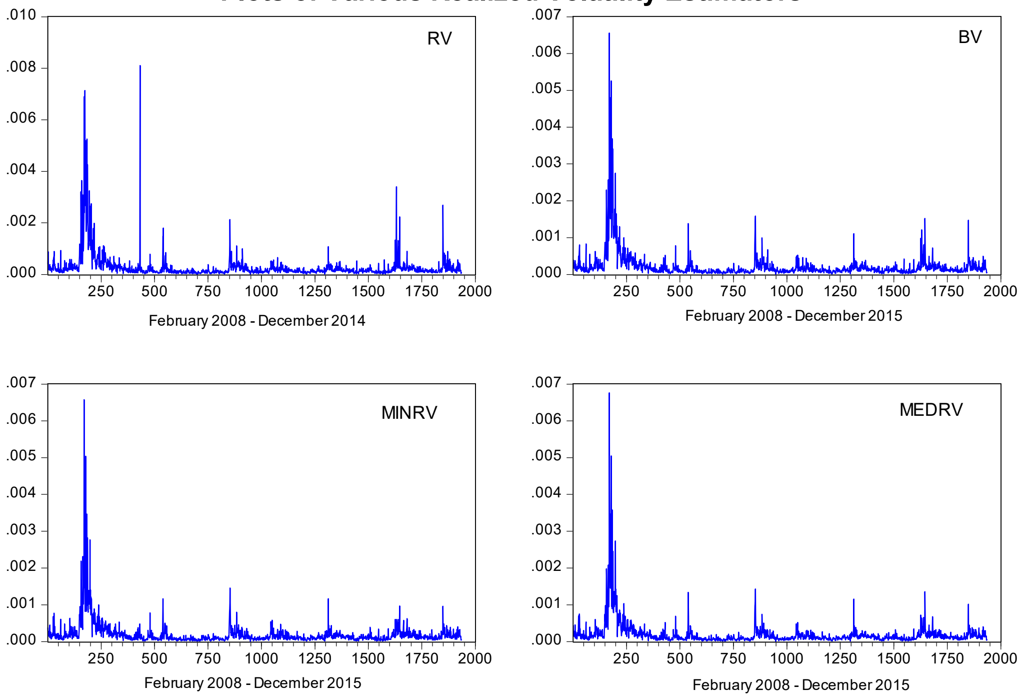


Table 1 shows the first four moment statistics of the logarithmic realized volatility. For Jacque-Bera normality tests, it is found that all the tests rejected the tests at 5% level of significance. Thus, the non-gaussianity assumption should be included in the model specification

Table 1

Descriptive Statistics for Logarithmic Volatility Estimators

Statistic	Log(RV)	Log(BV)	Log(minRV)	Log(medRV)
Mean	-8.687089	-8.687089	-8.976885	-8.983210
Std. Dev.	0.831292	0.831292	0.833403	0.825781
Skewness	0.909236	0.909236	0.704640	0.723439
Kurtosis	5.124847	5.124847	4.473609	4.589749
Jarque-Bera	630.6339*	630.6339*	335.2059*	372.5489*

Note: Jacque-Bera test, H_0: normality; * significant at 5% level.

HAR-GARCH Estimation Results

Table 2 illustrates the maximum likelihood estimations for four logarithmic heavy-tailed HAR-GARCH models with the additive volatility cascade of different time horizons namely daily, weekly and monthly under the student-t distributed error assumption. All the models indicated the tail index with degree of freedom, *v* above 2. Although the distributed errors of the volatility shown heavy-tail property, the intensity is considered as median if compared to the returns error normally fall within 3 to 6 degree of freedom (Dufour & Kurz-Kim, 2014). To include the conditional heteroskedasticity of realized volatility, the GARCH coefficients are all statistically different from zero. These results show that the presence of volatility in realized volatility (Corsi *et al*., 2009).

Table 2

The Maximum Likelihood Estimations

Estimation	HAR-RV-GARCH (1,1)	HAR-BV-GARCH (1,1)	HAR-minRV-GARCH (1,1)	HAR-medRV-GARCH (1,1)
θ_0	-0.873785* (0.146326)	-0.724907* (0.145266)	-0.737020* (0.157871)	-0.702630* (0.148530)
$\theta_{day,t-1}$	0.282411* (0.028770)	0.352403* (0.028698)	0.336853* (0.028771)	0.375843* (0.028238)
$\theta_{day,t-2}$	0.169790* (0.032126)	0.104746* (0.032463)	0.066448* (0.032176)	
$\theta_{week,t-1}$	0.221988* (0.055529)	0.216432* (0.056887)	0.252337* (0.058926)	0.309919* (0.044852)
$\theta_{month,t-1}$	0.232936* (0.036384)	0.251482* (0.036570)	0.269244* (0.039041)	0.242439* (0.036778)
α_0	0.034692* (0.013088)	0.027362* (0.013212)	0.018649 (0.010265)	0.013132 (0.007173)
α_1	0.079795* (0.023852)	0.056456* (0.020048)	0.041731* (0.016190)	0.040166* (0.014453)
β_1	0.793674* (0.062590)	0.831143* (0.066690)	0.888733* (0.048273)	0.905910* (0.038295)
ν	6.608347* (1.012813)	7.472163* (1.359337)	10.04029* (2.419972)	8.115834* (1.651082)
Model selection				
AIC	1.470220	1.375053	1.493118	1.385854
SIC	1.499192	1.404025	1.522091	1.411594
HIC	1.480950	1.385782	1.503848	1.395386
Diagnostic				
$\tilde{a}_t$, LB (12)	8.4830	7.5108	8.0394	10.549
$\tilde{a}_t^2$, LB (12)	16.892	12.913	12.734	13.835

Notes: *1. $\tilde{a}_t$ represents the standardized residual.*

2. The parentheses values represent standard error

*3. * denotes 5% level of significance.*

For HAR-RV-GARCH(1,1) and HAR-BV-GARCH(1,1) models, the estimation requires past daily volatility up to lag 2 whereas lag one for HAR-MINRV-GARCH(1,1) and HAR-MEDRV-GARCH(1,1) models in order to pass the Ljung-Box serial correlation tests for standardized and squared standardized residuals. For HAR-RV-GARCH(1,1) only, the impact of prior volatility are almost equally distributed by daily, weekly and monthly horizons. On the other hand, the HAR-BV-GARCH(1,1), HAR-MINRV-GARCH(1,1) and HAR-MEDRV-GARCH(1,1) models observed that the strength of the impact of past volatility are in the descending order of daily, weekly and monthly. This finding explained that the nearest past fluctuations of market returns have the highest impact to the recent volatility movements. It is found that all the coefficients of different time horizons are statistically different from zero at 5% level of significance. From the economic perspective, the empirical findings are supporting the heterogeneous market hypothesis where market participants with different investment time horizons have different ways to interpret market information differently. Using the additive components of various volatilities framework, the real structure of Brazilian financial stock market can be better explained and understanded by the long memory volatility behavior. This statistical element is an important finding in portfolio strategy planning and further explores the efficient market hypothesis.

In model diagnostic, all the models failed to reject the Ljung-Box serial correlations for standardized and squared standardized residuals under the null hypothesis of serially uncorrelated series. For model estimation performance, we refer to Akaike information criterion (AIC), Schwarz information criterion (SIC) and Hannan-Quinn information criterion (HIC). Table 2 shows that HAR-minRV indicates the highest values over the three criteria, followed by HAR-RV, HAR-medRV and HAR-BV. As a comparison, the RV performs slightly better than the minRV in the HAR modelling. This is because under the student-t assumption, the RV indicates the lowest tail index (measured in degree of freedom) with a value of 6.608347 as compared to minRV with the tail index of 10.04029. Since the t-distribution approaches the normal as the degree of freedom getting larger, therefore the RV fits better than minRV in the HAR modeling. However, good estimation result does not always provides outperform forecast results. Each of the models will be evaluated based on several loss functions in the forecast evaluations. In short, the jump-robust realized volatilities are out-performed the standard realized volatility in the estimation performance except for the minRV volatility representation.

Forecast Evaluations

Figure 3 and Figure 4 present the forecasts plot for combination forecasts and individual forecasts for all the seven models with the actual volatility proxy, logMEDRV whereas Table 3 shows the forecast evaluations using MAE, RMSE and MAPE with alternately four proxies as the actual volatility. Overall, in general all the volatility estimators indicated improvements (smaller MAE, RMSE and MAPE) when the actual volatility proxies shifted from RV to medRV. Overall, the combination forecast especially under the MSE ranks scheme achieved most frequent best forecast performance as compared to its counterparts. For individual model, logBV and logMEDRV only managed to score the best once for each for all the evaluations. In other words, the combination forecasts are proven to be more accurate for this particular study. It is also worth noting that when RV acted as the actual volatility proxy, all the estimators shown the highest MSE and RMSE. These findings are under expectation due to RV's higher intensity of noisiness as compared to the other three counterparts which either smoothen or eliminated the possible jumps (noisy observations). The noisy proxy of RV has caused inconsistent forecast performances (first row for MAE, RMSE and MAPE evaluations) with the other three estimators as indicated in Table 4. Besides the RV's acting as the volatility proxy, the ranking are very consistent for both the MAE and RMSE evaluations. Thus, the robustness (Patton, 2011) of the MAE, RMSE and MAPE evaluations are considered acceptable since the ranking are consistent no matter what type of proxies are being used. As a summary, the jump-robust estimators such as BV, MINRV and MEDRV performed better than the standard RV over the 7 models with 4 individual models and 3 combination models.

Figure 3

246 One-Day-Ahead Forecasts for Combined Forecasts

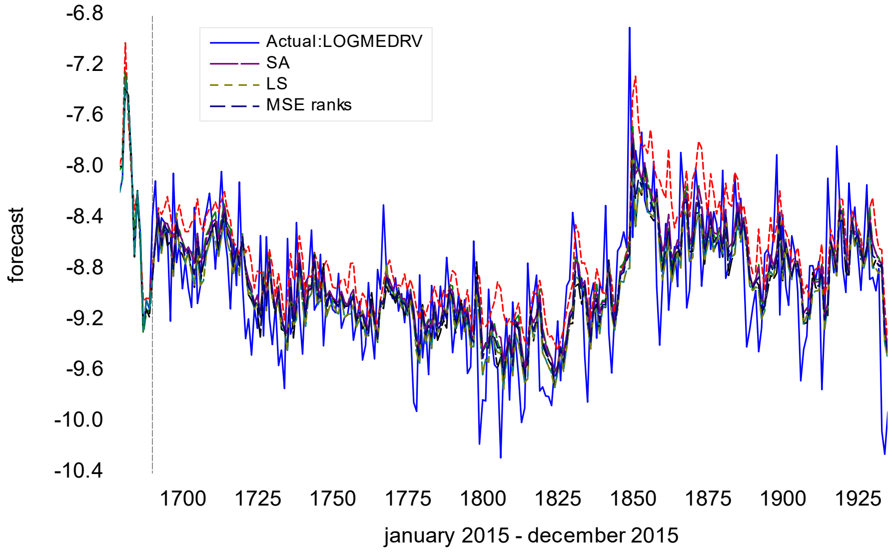


Figure 4

246 One-Day-Ahead Forecasts for Individual Forecasts

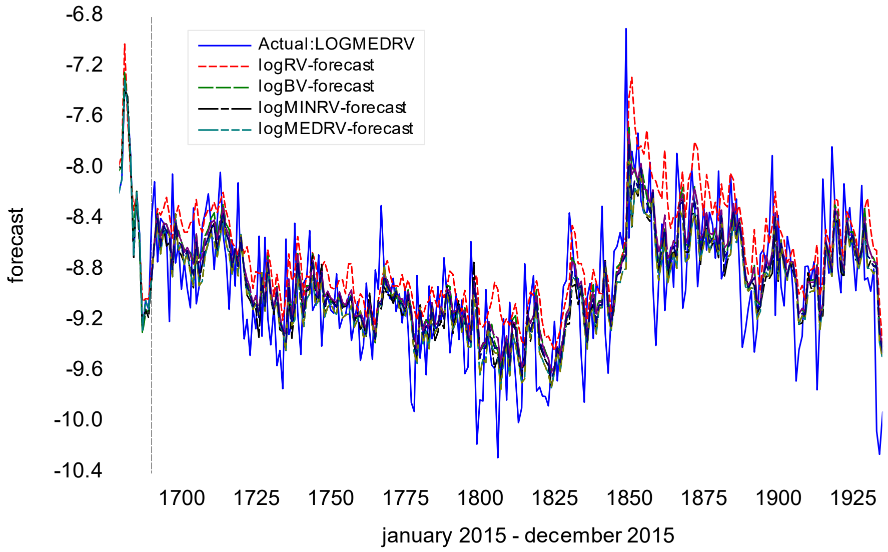


Table 3

Forecast Evaluations

Actual: logRV	Forecast evaluation		
Forecast using	RMSE	MAE	MAPE
logRV	0.49494	0.386536	4.562688
logBV	0.513392	0.387586	4.651566
logMINRV	0.54654	0.412809	4.977481
logMEDRV	0.554792	0.417078	5.037177
SA	0.513512	0.387999	4.657217
LS	**0.49494***	0.386536	4.562688
MSE ranks	0.501931	**0.381047***	**4.556527***
Actual: logBV	Forecast evaluation		
Forecast using	RMSE	MAE	MAPE
logRV	0.475343	0.378748	4.290004
logBV	**0.43176***	0.336388	3.866766
logMINRV	0.442661	0.340336	3.939538
logMEDRV	0.446142	0.341171	3.959907
SA	0.432175	0.335924	**3.861288***
LS	0.431760	0.336388	3.866766
MSE ranks	0.43279	**0.335453***	3.865176
Actual: logMINRV	Forecast evaluation		
Forecast using	RMSE	MAE	MAPE
logRV	0.528484	0.423925	4.722033
logBV	0.463652	0.370561	4.177745
logMINRV	0.461944	0.370603	4.206564
logMEDRV	0.462554	0.369928	4.210572
SA	0.464047	0.371904	4.192648
LS	0.461944	0.370603	4.206564
MSE ranks	**0.459905***	**0.368896***	**4.175272***
Actual: logMEDRV	Forecast evaluation		
Forecast using	RMSE	MAE	MAPE
logRV	0.50672	0.40612	4.503422
logBV	0.43245	0.342417	3.838837
logMINRV	0.429478	0.338921	3.826041
logMEDRV	0.428011	**0.336361***	3.807387
SA	0.433349	0.344358	3.860172
LS	0.428011	0.336361	3.807387
MSE ranks	0.**427348***	0.336963	**3.793044***

Note: * indicates the best model (smallest error)

Applications in Finance

For market risk evaluation, we determine the value-at-risk (VaR) based on the heavy-tailed HAR-GARCH which using alternately the RV, BV, minRV and medRV representation. The VaR is one of the important indicators (Jorion, 2006) in quantifying the market risk for financial and actuarial industries. According to Tsay (2005) probabilistic framework of VaR, the $\Delta r(\tau)$ is defined as the changes of the returns in stocks market from t to $t+\tau$ in a stock market. Also defines the $F_\tau(x)$, as the cumulative distribution function of $\Delta r(\tau)$, the VaR of a long position over the time horizon τ with probability α can be written as

$$F_{long\text{-}position}(\text{VaR}) = P[\Delta r(\tau) \leq \text{VaR}] = \alpha \,. \tag{8}$$

For heavy-tailed HAR-GARCH model, the long position for BOVESPA market α % quintile VaR is defined as

$$VaR_t = capital \times (Student_v \times \hat{\sigma}_t) \tag{9}$$

where: $student_v$ and $\hat{\sigma}_t$ represent the α-*th* quintile of a student-t distributed returns with tail parameter *v* and the forecasted volatility, respectively. The long position investors buy a stock, hold it while it appreciates, and sell it for profit. The market risk they are facing is when the price of the stock plunges. Therefore, long position investment concerns about the left tail of the financial return time.

Table 4

Value-at-Risk Determination

Volatility	RV	BV	minRV	medRV
One-day ahead forecast, $\widehat{RV}$	0.01227695	0.01182422	0.01185671	0.01196265
Student-t				
5% quantile	-0.01249139	-0.01148265	-0.01155505	-0.01179109
5% Value-at-risk	-12491	-11483	-11555	-11791
1% quantile	-0.02404564	-0.02261082	-0.02271381	-0.02304954
1% Value-at-risk	-24046	-22611	-22714	-23050
Normal				
1% quantile	-0.00886631	-0.00781311	-0.0078887	-0.00813514
1% Value-at-risk	-8866	-7813	-7889	-8135

Note: Long financial position value-at-risk with capital of \$1 million. The 5% and 1% critical values for *student-t* (degree of freedom, *v*=10.5248) are 2.22814 and 3.24984 respectively.

Assume that an investor holding a long financial position of the BOVESPA stock market with a capital of \$1 million. The 5% quantile for one-day ahead HAR(BV)-GARCH, student-t (*v*=10.61954) distributed return is

$$\begin{aligned} quantile_{1\%} &= student_{10.5248} \times \hat{\sigma}_{BV}(1) \\ &= -2.22814 \times 0.01182422 \\ &= -0.011483 \end{aligned}$$

For long position trading, the quintile often indicated in negative value, and it is understood that it signifies a loss which positioned at the left tail of the return distribution. The VaR with probability 0.05 is 0.011483×\$1000000 = \$$11483$. This result indicates that with probability 95%, the potential loss for the next day is \$$11483$. Similarly, the VaR with probability 0.01 is \$22611. From Table 4, as expected the most volatile RV indicated the greatest VaR, followed by BV, MINRV and lastly the MEDRV. As a comparison, we also conducted the forecasted volatility based on normality assumption. The 1% quintile for one-day ahead HAR(BV)-GARCH normal distributed return is

$$quantile_{1\%} = normal_{0.05} \times \hat{\sigma}_{norml}(1)$$
$$= -1.644854 \times 0.00024503$$
$$= -0.00781311$$

Thus the 1% VaR under the normality assumption is $7813 which is significantly smaller than student-t VaR with a value of $22611. Similar results have been shown for other volatility estimators as well. In other words, the inappropriate parametric distribution assumption against the empirical student-t distribution often faces the underestimation issue in VaR determination. Table 4 shows the overall results of VaR evaluations for all the volatility models.

IV. Conclusion

This study re-examines the heterogeneous market hypothesis using a modified heterogeneous autoregressive with various high frequency realized volatilities. The empirical findings show that the jump-robust volatility estimators outperformed the standard realized volatility in model specifications and forecast evaluations. In addition, the combination forecasts using three specific schemes indicated better forecast evaluations results over the individual models. To end this study, the forecasted volatilities are used in the value-at-risk determinations. As a conclusion, this study adds to the literature of efficient market hypothesis using high frequency data under the heterogeneous market hypothesis framework. The outcomes of this study also provide better forecasts and market risk determinations for the financial industries that involve with risk management and investment portfolio analysis.

Acknowledgement

The authors would like to thank Malaysia *Ministry of Higher Education (MOHE) for the financial support* under the Fundamental Research Grant Scheme (FRGS).

References

Aiolfi, M. and Timmermann, A., 2004. *Structural breaks and the performance of forecast combinations. Unpublished, Bocconi University.*

Andersen, T., and Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39, pp. 885-906.

Andersen T.G., Bollerslev, T., 1997. Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *The Journal of Finance*, 52, pp. 975-1005.

Andersen T.G., Bollerslev, T. and Diebold, F.X., 2006. Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. *Review of Economics and Statistics*, 89, pp. 701-720.

Andersen T. G., Bollerslev, T., Frederiksen, P. and Nielsen, M.O., 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25, pp. 233-261.

Andersen T.G., Bollerslev, T., Diebold, F.X. and Labys, P., 2003. Modeling and Forecasting Realized Volatility. *Econometrica,* 71, pp. 529–626.

Andersen T.G., Dobrev, D. and Schaumburg, E., 2011. A Functional Filtering and Neighborhood Truncation Approach to Integrated Quarticity Estimation. *National Bureau of Economic Research Working Papers*.

Andersen T. G., Dobrev, D. and Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169, pp. 75-93.

Barndorff-Nielsen, O.E. and Shephard, N., 2002. Estimating Quadratic Variation Using Realised Volatility. *Journal of Applied Econometrics*, 17, pp. 457-477.

Barndorff-Nielsen, O.E. and Shephard, N., 2004. Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2, pp. 1-37.

Blair, J.B., Poon, S.H., and Taylor, S.J., 2001. Forecasting S&P100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns. *Journal of Econometrics*, 105, pp. 5-26.

Cheong, C.W., 2013. The Computational of Stock Market Volatility from the Perspective of Heterogeneous Market Hypothesis. *Eco. Comp. & Econ. Cyber. Stu. & Res.*, 47, pp. 247-260.

Cheong, C.W., Isa, Z., Abu Hassan S.M.N., 2007. Modelling Financial Observable-Volatility using Long Memory Models. Applied Financial Economics Letters, 3, pp.201-208.

Corsi, F., 2009. A Simple Approximate Long Memory Modelo Realized Volatility. *Journal of Financial Econometrics*, 7, pp. 174–196.

Corsi, R., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The Volatility of Realized Volatility. *Econometric Reviews*, 27, pp. 46-78.

Dacorogna, M., Muller, U., Dav, R., Olsen, R. and Pictet, O.V., 1998. Modelling Short-Term Volatility with Garch and Harch models. *Nonlinear Modelling of High Frequency Finacial Time Series,* pp. 161-176.

Dacorogna M., Muller, U., Olsen, R. and Pictet, O.V., 2001. Defining efficiency in heterogeneous markets. *Quantitative Finance*, 1, pp. 198-201.

Dufour J.M. and Kurz-Kim J.R., 2014. Heavy Tails and Stable Paretian Distributions in Econometrics. *Journal of Econometrics*, 181, pp. 1-2.

Fama E., (1998). Market Efficiency, Long-term Returns, and Behavioral Finance. *Journal of Financial Economics*, 49, pp. 283-306.

Granger, C.W.J. and Ramanathan, R., 1984. Improved methods of combining forecasts. *Journal of Forecasting,* 3, pp. 197-204.

Hansen, P.R. and Lunde, A., 2006. Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics*, 24, pp. 127-218.

Jorion P., 2006. Value-at-Risk: The new benchmark for controlling market risk. Third Edition. McGraw-Hill, Chicago.

Lux, T. and Marchesi, M., 1999, Scaling and Criticality in A Stochastic Multi-Agent Model of Financial Market. *Nature,* 397, pp. 498-500.

Lo, A., 2005. Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *Journal of Investment Consulting*, 7, pp. 21-44.

Malkiel, B.G., 2003. The Efficient Market Hypothesis and Its Critics. *The Journal of Economic Perspectives, 17, pp. 59-82.*

Mandelbrot, B., 2008. How Fractals Can Explain What's Wrong with Wall Street. *Scientific American*.

Muller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R. and Ward, J., 1993. Fractals and Intrinsic Time - A Challenge To Econometricians. XXXIXth International AEA Conference on Real Time Econometrics, Luxembourg.

Muller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O. and von Weizsacker, J., 1997. Volatilities of Different Time Resolutions - Analysing the Dynamics of Market Components. *Journal of Empirical Finance,* 4, pp. 213-239.

Patton, A.J., 2011. Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics*, 160, pp. 246-256.

Peters, E.E., 1994. *Fractal market analysis: Applying Chaos Theory to Investment and Economics*. John Wiley & Sons, Inc.

Shiller, R.J., 2006. Tools for Financial Innovation: Neoclassical versus Behavioral Finance. *Financial Review*, 41, pp. 1-8.

Timmermann, A., 2006. Chapter 4 Forecast Combinations. *Handbook of Economic Forecasting*, 1, pp. 135-196.

Tsay, R.S., 2005. *Analysis of Financial Time Series*. John Wiley & Sons.