

Université de Montréal

Rapport de recherche

“Empirical Analysis of Jumps Contribution to
Volatility Forecasting Using High Frequency Data”

Rédigé par :
Imhof, Adolfo

Dirigé par :
Kalnina, Ilze

Département de sciences économiques
Faculté des arts et des sciences

16 avril 2013

Abstract

The aim of this project is to empirically validate, using high frequency data, whether jumps help or not to improve volatility forecasting by comparing the results of two models: the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV), proposed by Corsi F., 2009, and a version of the Heterogeneous Auto-Regressive with Continuous Volatility and Jumps (HAR-CJ) model described in Corsi et al., 2010¹. Both models assume that log-price follows a jump-diffusion process² and consider volatilities realized over different interval horizons in order to capture the heterogeneity of the different participants of the market and the asymmetric propagation of volatility. The difference between them is that the HAR-RV does not split the quadratic variation of the cumulative return process into its continuous and discrete parts, but the HAR-CJ model does. The analysis of the data shows that after a large jump, the daily realized volatility is significantly higher than usual suggesting that jumps may have a positive impact in future volatility. The results we obtain confirm this idea because we demonstrate that the forecasted volatility is improved when the volatility is broken down into continuous variations and jumps. In other words, the volatility forecasting precision of the HAR-CJ model is better than that of the HAR-RV model.

1. Introduction

Volatility forecasting is an important factor for option trading and value at risk calculation of a financial position in risk management and also has many other financial applications³. Consequently, it is one of the most important and more studied subjects in financial econometrics. Standard GARCH and stochastic volatility models are not able to reproduce certain features of financial data; long-memory volatility models, FIGARCH and ARFIMA, are

¹ *Instead of using the estimator proposed by Corsi et al., 2010 for the integrated volatility, which is the threshold bipower variation, we use the bipower variation (BPV), which is the most popular estimator.*

² *Although Corsi F., 2009 describes his model as if no jumps were a requirement, we describe it differently.*

³ *Tsay R.S., 2010, p. 109 and 110.*

complex and not easily extendible to multivariate processes⁴. Thus, we have decided to use the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) and the Heterogeneous Auto-Regressive with Continuous volatility and Jumps (HAR-CJ) models, which in spite of their simplicity are able to produce rich dynamics for returns and volatilities which closely reproduce the observed empirical data and show good forecasting performance. While some studies show that jumps are irrelevant or even have a negative impact in volatility forecasting, other studies show that jumps have a positive and significant impact on future volatility⁵.

First of all, we start with the definition of the models and explain how the forecasting performance is evaluated. We continue with the empirical analysis, identifying the data and explaining the data cleaning process and how the fixed intervals are constructed. Some descriptive statistics are also added. Then, we produce the estimation of the parameters and forecast the daily realized volatility one day ahead for each of the two models. Finally, we measure their performance in terms of forecasting precision and conclude.

2. The Models

We first review the definitions that are applicable to both models. We suppose that the log price follows the standard continuous time process

$$(1) \quad dp(t) = \mu(t) dt + \sigma(t) dW(t) + \kappa(t) dq(t), \quad 0 \leq t \leq T,$$

where $dp(t)$ is the logarithm of instantaneous price,

$\mu(t)$ is cadlág finite variation process.

$W(t)$ is a standard Brownian motion

$\sigma(t)$ is a stochastic process independent of $W(t)$

$dq(t)=1$ if there is a jump at time t and $dq(t)=0$ otherwise and

$\kappa(t)$ refers to the corresponding jump size

⁴ Corsi F., 2009 p. 175.

⁵ Corsi et al., 2010 p. 276.

The quadratic variation of the cumulative return process, $r(t) = p(t) - p(0)$, is

$$(2) \quad [r, r]_t = \int_0^t \sigma^2(s) ds + \sum_{0 < s \leq t} \kappa^2(s)$$

Equation (2) splits the quadratic variation of the cumulative return process into its continuous variation, the so-called integrated variance (first term on the right side of the equation), and its discrete variation or the cumulative sum of squared jumps (second term on the right side of the equation).

The integrated variance associated with day $t+1$ is the integral of the spot or instantaneous variance over one-day interval $[t, t+1d]$, where a full-trading day is represented by the time interval $1d$.

$$(3) \quad IV_{t+1} = \int_t^{t+1} \sigma^2(s) ds$$

Consequently, the integrated volatility is

$$(4) \quad \sigma_{t+1} = (IV_{t+1})^{1/2}$$

Let the discretely sampled Δ -period returns be denoted by $r_{t, \Delta} \equiv p(t) - p(t - \Delta)$.

An estimator for the square root of the quadratic variation of the cumulative return process is the realized volatility, which for one trading day is defined by the square root of the summation of the corresponding $1/\Delta$ high-frequency intraday squared returns.

$$(5) \quad RV_{t+1}^{(d)} \equiv \sqrt{\sum_{j=1}^{1/\Delta} r_{t+j\Delta, \Delta}^2}$$

This realized volatility converges uniformly in probability to the corresponding increment to the quadratic variation process, i.e. for $\Delta \rightarrow 0$,

$$(6) \quad RV_{t+1}^{(d)} \rightarrow \sqrt{[r,r]_{t+1} - [r,r]_t} = \sqrt{\int_t^{t+1} \sigma^2(s) ds + \sum_{t < s \leq t+1} \kappa^2(s)}$$

In the absence of jumps, the realized volatility is thus consistent for the integrated volatility and if jumps are present the realized volatility is consistent for the square root of the sum of the integrated variance and the cumulative sum of squared jumps.

Note that the realized volatility only takes into account the returns within the trading day (intraday returns) but not the overnight returns; thus, the potential overnight jumps are excluded.

The weekly realized volatility at time t is a simple average of the daily volatilities.

$$(7) \quad RV_t^{(w)} = \frac{1}{5} (RV_t^{(d)} + RV_{t-1d}^{(d)} + RV_{t-2d}^{(d)} + RV_{t-3d}^{(d)} + RV_{t-4d}^{(d)})$$

And the monthly realized volatility is obtained applying the same formula but considering 20 days instead of 5 days.

The purpose of measuring the realized volatility at different time intervals is to capture the heterogeneity of the different participants of the market and the asymmetric propagation of volatility (see Corsi F., 2009).

1. The HAR-RV model

To forecast the daily realized volatility one period ahead using the HAR-RV model we consider the following equation:

$$(8) \quad RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}$$

where the c (*the constant*) and β are the parameters and $\omega t + Id$ is the error term. The model has a simple autoregressive structure in the realized volatility but with the feature of considering volatilities realized over different interval sizes; this model could be seen as HAR(3)-RV.

2. The HAR-CJ model

Now, we present a version of the Heterogeneous Auto-Regressive with Continuous volatility and Jumps (HAR-CJ) model described in Corsi et al., 2010. Instead of using the estimator proposed by Corsi et al., 2010 for the integrated volatility, which is the threshold bipower variation, we use the bipower variation (BPV), which is the most popular estimator. The advantage is that a threshold 'c' does not need to be chosen. The disadvantage is that it might produce worst results in finite samples.⁶

The forecasted realized volatility with the Heterogeneous Auto-Regressive with Continuous Volatility and Jumps (HAR-CJ) is⁷

$$(9) \quad RV_{t:t+h-1} = \beta_0 + \beta_d \hat{C}_{t-1} + \beta_w \hat{C}_{t-5:t-1} + \beta_m \hat{C}_{t-20:t-1} + \beta_j \hat{J}_{t-1} + \varepsilon_t$$

where \hat{J}_t and \hat{C}_t are defined below.

The daily continuous part of the quadratic variation is defined by

$$(10) \quad \hat{C}_t = RV_t - \hat{J}_t.$$

The weekly and monthly continuous parts of the quadratic variation at time t are simple averages of the daily ones.

⁶ For the HAR-CJ model, we follow Huang et al., 2005, Andersen et al., 2007 and Corsi et al., 2010.

⁷ Please note that when $h=1$, we are forecasting the daily realized volatility one period ahead.

The daily jump is defined as

$$(11) \quad \widehat{J}_t = I_{\{z_t > \Phi_\alpha\}} \cdot (RV_t^{(d)} - BPV_t)^+$$

where z_t and BPV_t are defined below and Φ_α is the cumulative distribution function at a confidence level α . I is an indicator function that has two possible outcomes: 1, when $z_t > \Phi_\alpha$ and a jump is detected and; 0 otherwise, when there is no jump. The symbol plus (+) indicates that the subtraction on the right side can be either zero, when the results is zero or negative, or a positive number.

Jump Detection Test

We use the following jump detection test⁸ to test for daily jumps. For each t , $z_t \xrightarrow{D} N(0,1)$ as $M \rightarrow \infty$ on the assumption of no jumps.

$$(12) \quad z_t = \frac{RV_t - BV_t}{\sqrt{(v_{bb} - v_{qq}) \frac{1}{M} TP_t}}$$

where BV_t , v_{bb} , v_{qq} and TP_t are defined below.

The bipower variation equation is

$$(13) \quad BV_t = \sqrt{\mu_1^{-2} \left(\frac{M}{M-1} \right) \sum_{j=2}^M |r_{t,j-1}| |r_{t,j}|} = \sqrt{\frac{\pi}{2} \left(\frac{M}{M-1} \right) \sum_{j=2}^M |r_{t,j-1}| |r_{t,j}|}$$

The bipower variation (asymptotically) measures the integrated volatility attributable to the diffusive component.

⁸ See Huang, X. et al., 2005, equation (5).

The tri-power quarticity statistics is

$$(14) \quad TP_t = M\mu_{4/3}^{-3} \left(\frac{M}{M-2} \right) \sum_{j=3}^M |r_{t,j-2}|^{4/3} |r_{t,j-1}|^{4/3} |r_{t,j}|^{4/3}$$

and

$$(15) \quad v_{qq} = 2,$$

$$(16) \quad v_{bb} = \left(\frac{\pi}{2} \right)^2 + \pi - 3.$$

Forecasting Performance

To evaluate the forecasting performance of the two models, we are going to use the heteroskedasticity adjusted root mean square error suggested in Bollerslev and Ghysels (1996) and the QLIKE loss function.

$$(18) \quad \text{HRMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{RV_t - \widehat{RV}_t}{RV_t} \right)^2}$$

where RV_t is the ex-post and \widehat{RV}_t is the ex-ante realized volatility at time t .

$$(19) \quad \text{QLIKE} = \frac{1}{T} \sum_{t=1}^T \left(\log RV_t + \frac{\widehat{RV}_t}{RV_t} \right)$$

3. Empirical analysis

1. Data

In this section we identify the data that have been used as well as indicate how the fixed intervals are constructed and also produce some descriptive statistics.

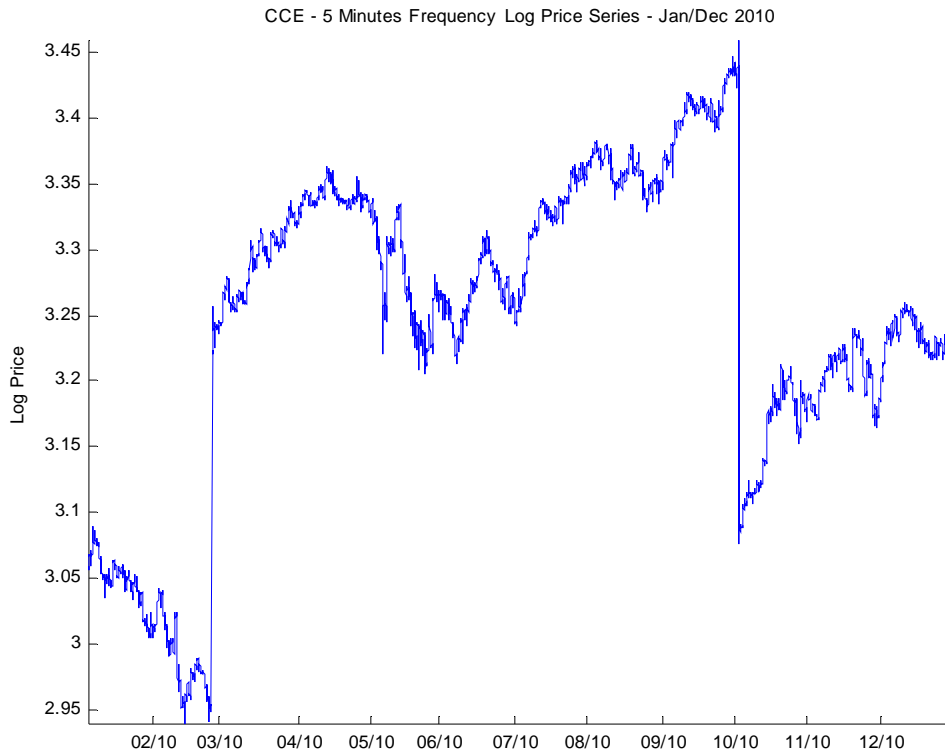
We have used the TAQ (transaction and quotes database) from the NYSE (New York Stock Exchange) price series for CCE (Coca Cola Enterprises) for 2010.

For identifying and removing outliers, we have followed the procedure described in Brownlee C.T. et al., 2006.

Then, we sampled the data at 5 minutes fixed intervals within the official trade hours. Therefore, each day has 79 observations, or 78 intervals 5 minute each, from 9:30am to 4pm. To create the equally spaced time intervals, we have taken the price of the last observation of each interval.

Fig. 1 shows the 5 minutes frequency log prices for CCE for the year 2010 (almost 20000 observations for 2010). We can clearly distinguish 2 jumps in log price. The first one is upward on February 25 and the second one is downward on October 3. On February 24, 2010, The Coca-Cola Company and CCE (Coca-Cola Enterprises) entered talks about selling CCE's North American division to Coca-Cola. Coca-Cola paid over \$15 billion, including a redemption of Coca-Cola's 33% shareholding in CCE. The acquisition closed on October 3, 2010.

Fig. 1.



Figures below show the nominal tick-by-tick transaction price series (Fig.2a) and the log price series for the 79 sampled observations (Fig.2b) of the first day of the sample, January 4th, 2010. The number of transactions for the CCE stock total 4692 for that day and almost 1150 thousands for all the 252 trading days in 2010. We can see that there are some trades reported before and after the official NYSE opening and closing trade hours, which are 9:30 am and 4 pm respectively.

Fig. 2a.

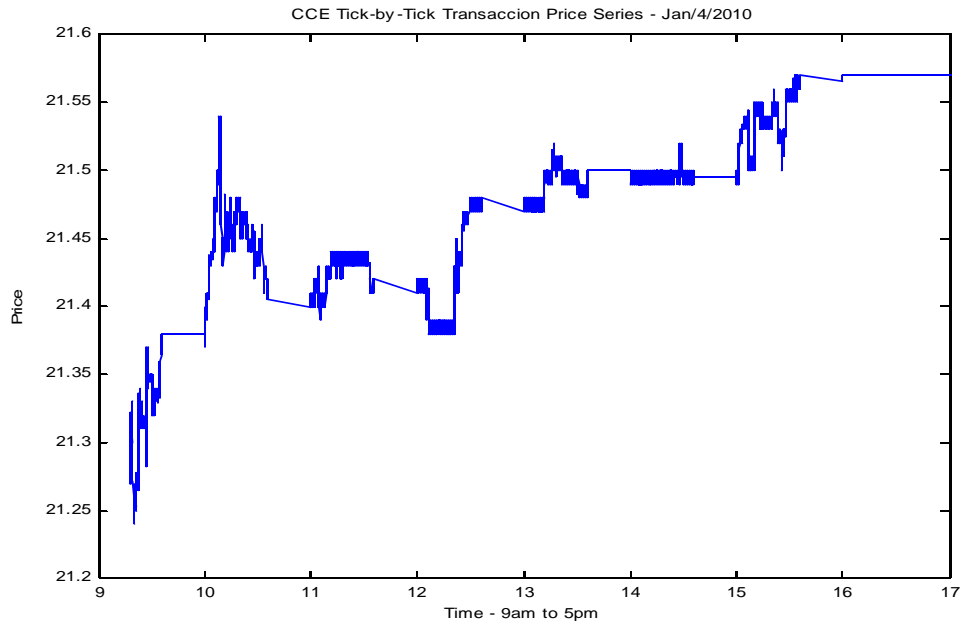


Fig. 2b

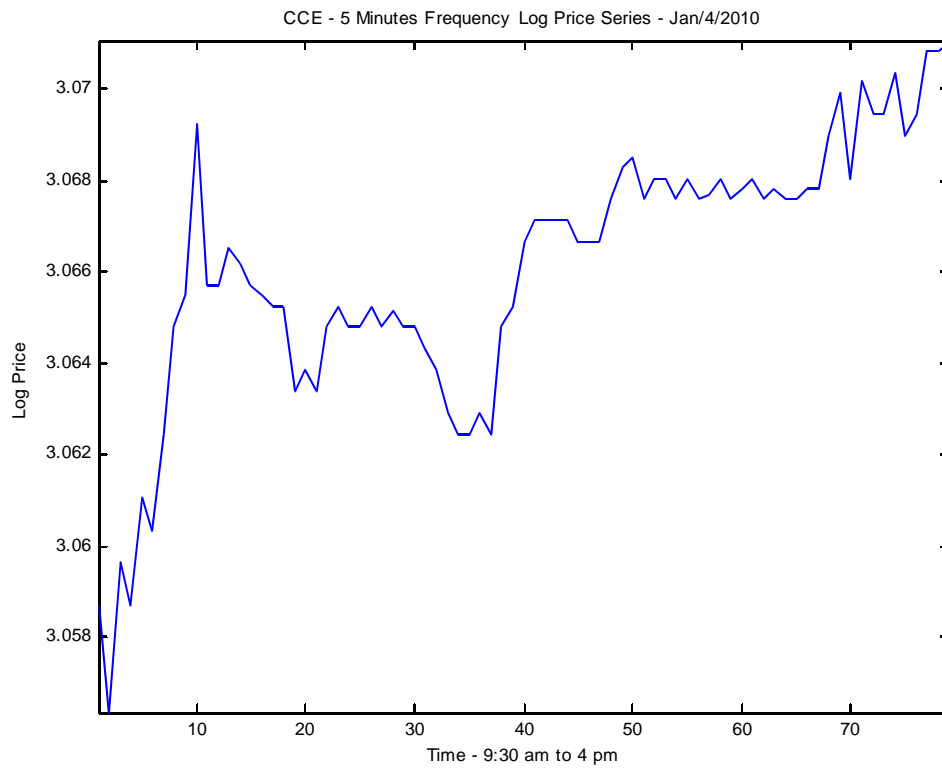
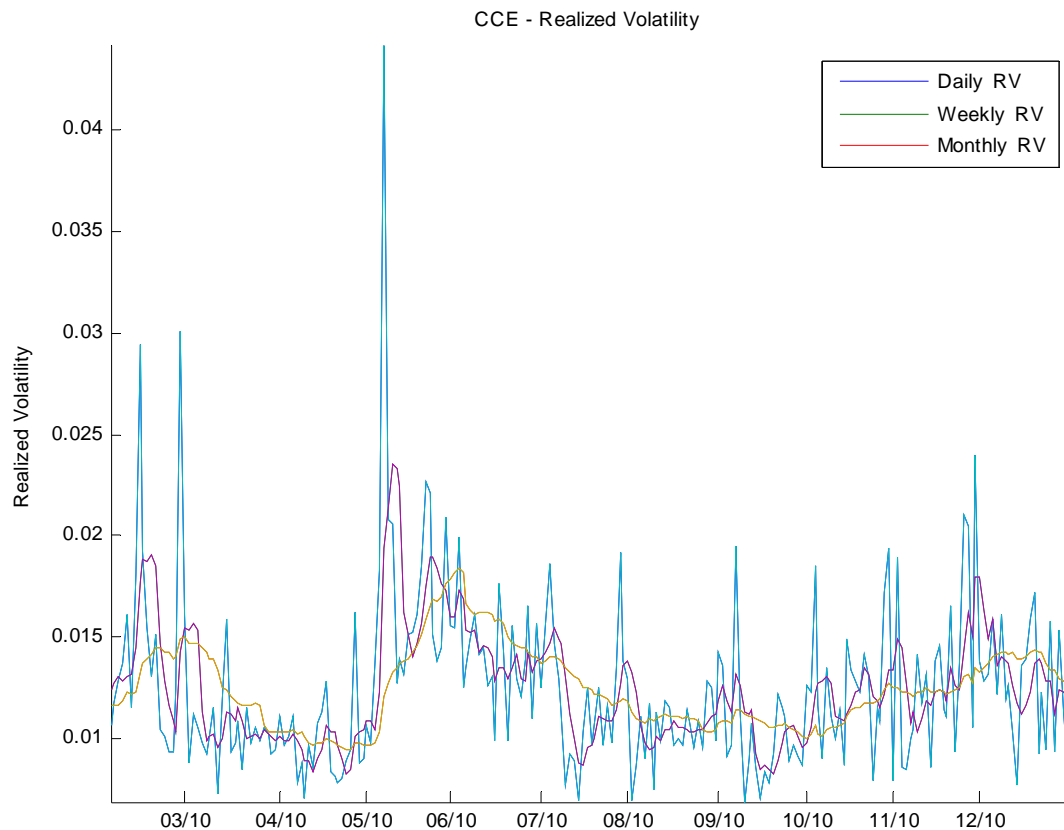


Figure 3 shows the daily, weekly and monthly realized volatilities. The daily realized volatility shows three outliers on 10 February, 25 February and 6 May. These are intraday changes in the asset price. The outlier on 25 February is produced just after the overnight jump in price for the acquisition of CCE by Coca-Cola. Hence, this may confirm the idea that jumps have a positive and mostly significant impact on future volatility as stated in Corsi et al., 2010.

Fig. 3



The autocorrelation function of daily realized volatilities (Fig.4) is becoming small very slowly. While this situation does not bias the OLS coefficient estimates, the standard errors tend to be underestimated.

Fig. 4.

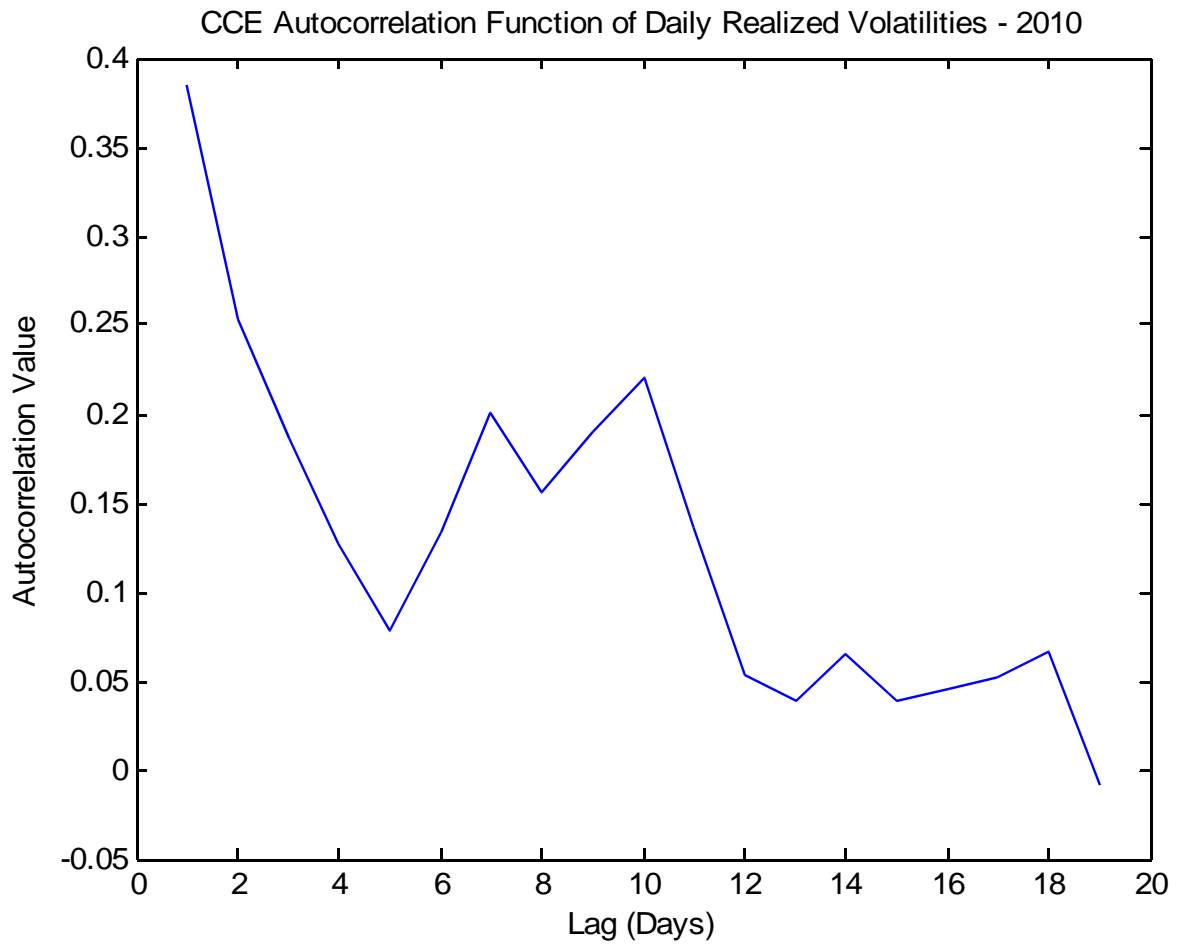


Figure 5 shows the results statistics for the regression of all the available observations (232 days).

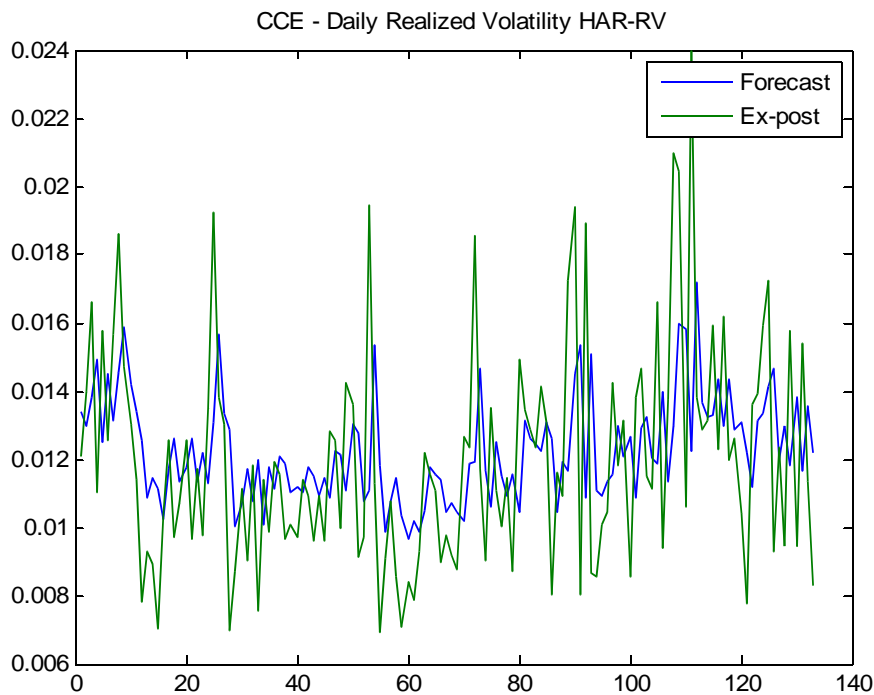
4. Estimation and Forecast

Now, we estimate the parameters and produce the forecasted volatility one day ahead.

We consider equations (8) and (9)⁹ and estimate the parameters by simple linear regression (OLS)¹⁰.

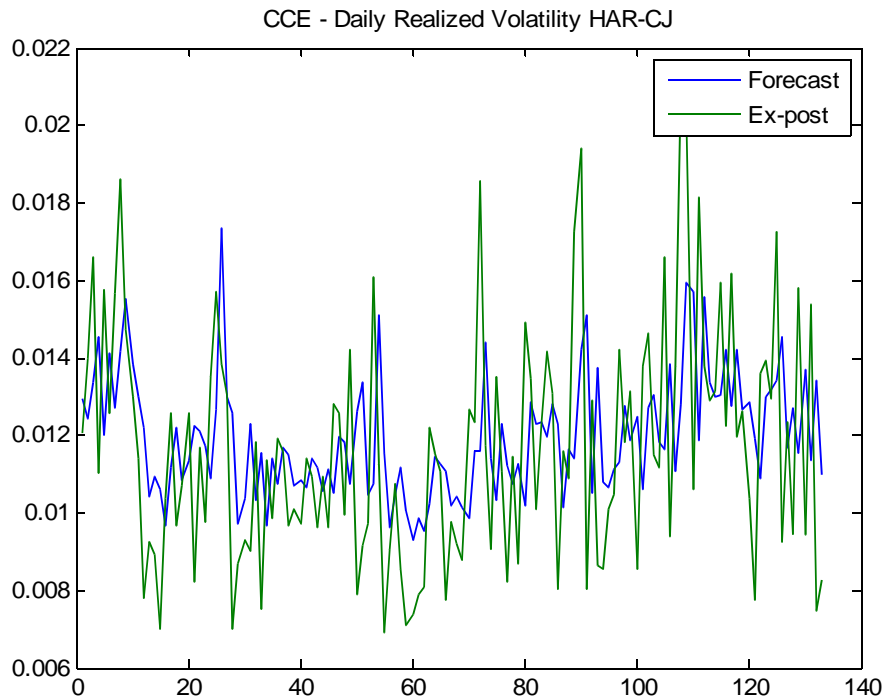
The comparison between the *ex-ante* (forecast) and the *ex-post* figures is shown in the charts below. The first chart was obtained with the HAR-RV model and the second chart with the HAR-CJ model. We can see that the time series for the forecasted daily realized volatility smooths the *ex-post* time series.

Fig. 5.



⁹ For equation (9), we consider $h=1$.

¹⁰ We have taken as a base the first 100 days of the time series available and have produced 133 forecasted volatilities one day ahead.



5. Forecasting precision

Once the forecasted volatilities are produced, we assess the performance of the models in term of forecasting precision. We consider the equations (18) and (19) and obtain the results of the RMSE and QLIKE loss function equation. For the HAR-CJ model, the significant daily jumps are computed using a critical value of $\alpha = 99\%$ -see equation (11)-.

Model	RMSE	QLIKE
HAR-RV	0.0290	-3.3735
HAR-CJ	0.0283	-3.3990

Based on the results of the RMSE and QLIKE loss function equation, the forecasting precision of the HAR-CJ model is better than that of the HAR-RV model.

6. Conclusion

We have considered the HAR-RV and HAR-CJ models to forecast volatility one day ahead using one year of high frequency data for Coca Cola Enterprises (CCE). The analysis of data shows that within the day after a large jump, as for example the overnight jump produced between the 24 and 25 February 2010, the daily realized volatility is significantly higher than usual. This situation may indicate that jumps should have a positive impact in future volatility. The results we have obtained confirm this idea and we have demonstrated that when the volatility is broken down into continuous variations and jumps, the forecasted volatility is improved. In other words, the volatility forecasting precision of the HAR-CJ model is better than that of the HAR-RV model.

REFERENCES

- Andersen, T., Bollerslev, T., Dobrev, D., 2007. “No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and IID noise: theory and testable distributional implications”. *Journal of Econometrics* 138 (1), 125–180.
- Bollerslev, T., Ghysels, E., 1996. Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics* 14 (2), 139–151.
- Brownlees, C.T., Gallo, G.M. 2006. “Financial econometric analysis at ultra-high frequency: Data Hangling concerns.”, *Science Direct – Computational Statistics & Data Analysis* 51, 2232-2245.
- Corsi, F., 2009. “A Simple Approximate Long-Memory Model of Realized Volatility.”, *Journal of Financial Econometrics*. Vol.7, No. 2, 174-196.
- Corsi, F., Pirino, D., Reno, R., 2010. “Threshold bipower variation and the impact of jumps on volatility forecasting.”, *Journal of Econometrics* 159, 276-288.
- Huang, X., Tauchen, G., 2005. “The relative contribution of jumps to total price variance”. *Journal of Financial Econometrics* 3 (4), 456–499.
- Tsay R.S., 2010. “Analysis of Financial Time Series”. 3rd edition, Wiley.
- “The Coca-Cola Company Completes Transaction of the Coca-Cola Enterprises”, [BevNET.com](http://www.bevnet.com), October 2010, [<http://www.bevnet.com/news/2010/10-4-2010-coca-cola-aquistion>] (accessed on February 19, 2010)