Forecasting Daily Volatility on Bucharest Stock Exchange using HAR Model

Ramona Radu

Bucharest University of Economic Studies - Faculty of Finance, Assurance, Banking and Stock Exchange, Bucharest, Romania
Email: ramona.radu@fin.ase.ro

Abstract
In this paper we forecast realized volatility using a very liquid equity traded on Bucharest Stock Exchange, Property Fund (FP) which is a company that manages Romanian government real-estate properties and at that time had almost 40% market capitalization. Many research paper use strndard version of HAR model, with or without jumps to forecast volatility on markets with a high level of liquidity. Our result, based on Diebold an Mariano test show that forecast performance increase if we use a modified version of HAR model, allowing for average trade duration form prevoius day to play an important role in analysis.
Keywords: Volatility, Stock Exchange, HAR Model.

Introduction

Unlike financial asset prices that are observable variables, their volatility is a latent variable, hence the need to be estimated. From a risk management perspective the necessity of accurate forecasts of volatility is crucial. During the last decades some major innovation in financial risk estimation have been made, models like ARCH (AutoRegressive Conditional Heteroskedasticity), GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) or SVM (Stochastic Volatility Model) being extremely used in theory but also in practice.

In recent years, a new paradigm that was born due to technological developments, preoccupied academic community, namely predicting the risk of an asset where we have available all transactions in a given day. Thus, we can say that in the last twenty years one of the most important concepts in the financial econometrics is likely introduction of Realized Volatility estimator developed by Andersen and Bollerslev(1998), a concept that allows an unbiased estimation for integrated variance (IV) of a financial asset's price in a particular time, such as a day, performed by summing the squares returns recorded during that trading day. Unlike other models that are taking into account only squared daily returns, like ARCH or GARCH models, the Realized Volatility estimator takes advantage on the all available intra-day information.

Based on Realized Volatility estimator Corsi(2009) proposed the heterogeneous autoregressive (HAR) model to forecast daily volatility. This model is a very good alternative to GARCH models having the advantage of using all available intra-day returns.

Nowadays, HAR model is the standard benchmarkt for modellind and forecasting financial volatility. Authors like Bandi et. all.(2013), Chen et all.(2010), or Lahaye and Shaw (2014) propossed some extension of HAR model thar significantly improved forecast performance. Those extensions where based on different decompositions of realized volatility, in the spirit of Andersen et. all. (2007) who established that realized volatililty can be separated into a continuous sample path and a jump component.

In this paper we are going to extend the approach presented by Corsi(2009) by introducing in linear regression model another term, namely average trade duration. Based on intra-day data for Property Fund (FP) traded on Bucharest Stock Exchange we estimate HAR model with his extension in order to obtain a volatility forecast.

Prevoius studies regarding the efficiency of Romanian capital markets, historical volatility and trading costs from an intra-day perspective were performed by Dragotă et al.(2009), Cepoi (2014a, 2014b), Cepoi and Radu(2014) and Radu and Cepoi(2015). They found that Bucharest Stock Exchange has a week form of efficiency but in comparison with other markets trading cost are much higher.

The paper is organized as follows. Section 2 presents literature review of HAR model and it's extensions. In Section 3, data and estimation results are analyzed. Section 4 concludes this research.

Methodology
Realized Volatility
Let's assume that the logarithmic price of a financial asset is given by the following process

$$p_t = \int_0^t \mu(s) \, ds + \int_0^t \sigma(s) \, dW(s) \tag{1}$$

In equation (1) the mean process, μ(s) is continuous and with finite variation, σ(s) represents the instantaneous volatility while W is a standard Brownian motion. The integrated variance represents the amount of variation accumulated over a past time interval Δ and is given by:

$$IV = \int_{t-\Delta}^t \sigma^2(s) \, ds \tag{2}$$

The cornerstone of this approach is to estimate the integrated variance given by the above equation using a discrete representation. If we assume that we have m intra-day returns the i-th return can be defined as:

$$r_i^{(m)} = p_{i/m} - p_{(i-1)/m} \tag{3}$$

Summing all squared intra-day returns

$$RV^{(m)} = \sum_{i=1}^m \left(r_i^{(m)} \right)^2 \tag{4}$$

we get a natural estimator for the integrated variance denoted realized volatility and was proposed in their pioneering work by Andresen and Bollerslev (1998).However the realized volatility estimator is biased if price series data is contaminated by microstructure noise or jumps. In this sense, Barndorff-Nielsen and Shepard(2004), Zhang *et al.* (2005) *etc.* proposed different kind of high-frequency unbiased variance estimators.

HAR Model

The Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009) has been the most popular model in forecasting daily volatility mostly for it's simplicity and it's ability to accommodates some stylized facts that exists in financial asset volatility moddeling such as multiscaling behavoiur or long memory. The original HAR-RV model specifies RV as a function of a lagged daily realized volatility ($RV_{d,t}$), lagged weekly realized volatility ($RV_{w,t}$), and lagged monthly realized volatility ($RV_{m,t}$) and is expressed as:

$$RV_{t+1,d} = \alpha_0 + \alpha_1 RV_{t,d} + \alpha_2 RV_{w,t} + \alpha_3 RV_{m,t} + \varepsilon_{t+1} \tag{5}$$

where RV_{t+1} denotes the daily realized volatility in day *t+1*, RV_t denotes the daily realized volatility in day *t*, $RV_{w,t} = \frac{1}{5}\sum_{i=0}^{4} RV_{t-i}$ is the weekly realized volatility and $RV_{m,t} = \frac{1}{22}\sum_{i=0}^{21} RV_{t-i}$ is monthly realized volatility.

Andersen et al. (2007) extended this model by explicitly decomposing the realized volatilities into the continuous sample path variability and the jump variation utilizing the separate nonparametric measurements based on a statistical jump test to the HAR-RV-CJ model, which is expressed as:

$$RV_{d,t+1,} = \beta_0 + \beta_{c,d}C_{d,t} + \beta_{c,w}C_{w,t} + \beta_{c,m}C_{m,t} + \beta_{j,d}J_{d,t} + \beta_{j,w}J_{w,t} + \beta_{j,m}J_{m,t} + \varepsilon_{t+1} \tag{6}$$

where $RV_{d,t+1}$ denotes the daily realized volatility in day *t+1, $C_{d,t}$, $J_{d,t}$* are respectively the continuous sample path and jump variation at time t and $C_{w,t} = \frac{1}{5}\sum_{i=0}^{4} C_{d,t-i}$, $C_{m,t} = \frac{1}{22}\sum_{i=0}^{21} C_{d,t-i}$, $J_{w,t} = \frac{1}{5}\sum_{i=0}^{4} J_{d,t-i}$ and $J_{w,t} = \frac{1}{22}\sum_{i=0}^{21} J_{d,t-i}$.

Data and Results

In this section we will present some of the stylized facts of the main data used in our estimations in order to set the stage for the subsequent econometric analysis. We chose one of the top equity traded on Bucharest Stock Exchange, Property Fund (FP) spanning a period from 19 October 2012 to 2 May 2013. FP is from a company which manages Romanian government real-estate properties and at that time had almost 40% market capitalization. The data were extracted from Thomson Reuters Platform. The average numer of thades per day for FP was 92, a tranzaction being made on average at every 4.38 minutes.

Before actually estimating the HAR model and it's extension we need to chek if data series is contaminated with jumps. If yes, we expect to have a better performance in forcasting volatility if we aplly decomposition proposed by (Andersen et. al., 2007). In order to check the exstinece of jumps, we apply the test propossed by Jiang and Oomen (2008). They found that the difference of simple return and logarithmic return can capture one half of integrated variance if there is no jump in the underlying sample path. The null hypothesis of this test is no jumps. The results of Jiang and Oomen test are presented in below:

Table 1
Jinag and Omen jump test

Equity	Test Value	Critical Value	P-value
FP	2.3245	1.96	2.01%

As we can see in Table 1 the null hypothesis of nu jumps in price series is rejected at 5% level. This means that price series is contaminated with jumps and we were expected to have an improvement in regression results when we apply the methodology of Andersen et. all. (2007).

In order to forecast realized volatility we need to estimate first HAR model and his extension. If we denote average trade duration with AVT, ant taking into account equation (5) and (6) we are going to estimate and compare four models:

$$\begin{cases} RV_{t+1,d} = \alpha_0 + \alpha_1 RV_{t,d} + \alpha_2 RV_{w,t} + \alpha_3 RV_{m,t} + \varepsilon_{t+1} \\ RV_{t+1,d} = \alpha_0 + \alpha_1 RV_{t,d} + \alpha_2 RV_{w,t} + \alpha_3 RV_{m,t} + \gamma ATD_t + \varepsilon_{t+1} \\ RV_{d,t+1,} = \beta_0 + \beta_{c,d}C_{d,t} + \beta_{c,w}C_{w,t} + \beta_{c,m}C_{m,t} + \beta_{j,d}J_{d,t} + \beta_{j,w}J_{w,t} + \beta_{j,m}J_{m,t} + \varepsilon_{t+1} \\ RV_{d,t+1,} = \beta_0 + \beta_{c,d}C_{d,t} + \beta_{c,w}C_{w,t} + \beta_{c,m}C_{m,t} + \beta_{j,d}J_{d,t} + \beta_{j,w}J_{w,t} + \beta_{j,m}J_{m,t} + \gamma AVT_t + \varepsilon_{t+1} \end{cases}$$

Estimation results are presented in below

Table 2
Estimation results

FP - models comparison				
Model	1:HAR-RV	2:HAR-RV-D	3:HAR-RV-CJ	4:HAR-RV-CJ-D
α_0	0.0080**	0.0112***	-	-
α_1	0.3500***	0.3026***	-	-
α_2	-0.1190	-0.1207	-	-
α_3	0.0758	0.0766	-	-
γ	-	-0.0335**	-	-0.0374**
β_0	-	-	0.009907***	0.013807***
$\beta_{c,d}$	-	-	0.468335***	0.421828***
$\beta_{c,w}$	-	-	-0.00365	-0.00595
$\beta_{c,m}$	-	-	-0.26295	-0.18782
$\beta_{j,d}$	-	-	-0.14055	-0.17183
$\beta_{j,d}$	-	-	-0.25301	-0.25031
$\beta_{j,d}$	-	-	0.304587	0.123074
R-squared	10.70%	14.72%	19.45%	24.33%
F-statistic	3.79529**	4.059049***	3.701925***	4.179869***
DW-stat	2.019385	2.030128	1.950525	1.955064

The asterix *,** and *** denote rejection of null hypothesis at 10%, 5% and 1% significance level

As we can see in Table 2, only the estimates describing the information from prevois day are statistically significant. This means that volatility in day *(t+1)* can be forecasted if we take a look to the information regarding realized volatility in day (*t*) and avarege trade duration in day (*t*). This statement is true for all four regressions. In all cases an increase in

realized volatility(or in continuous sample path of realized volatility) in day (*t*) will increase next day realized volatility. We can also observe a negative relationship between average trade duration, as a measure for market liquidity and next day realized volatility. Practically, a drecrease in average trade duration in day (*t)* will increase next day realized volatility for both Model 2 and Model 4. However, including average trade duration alongside with Andersen et. all. (2007) decomposition of realized volatility into continuous sample path and a jump component is increasing the value of R-squared with almost 14% (10.70% for Model 1 vs. 24.33% for Model 4)
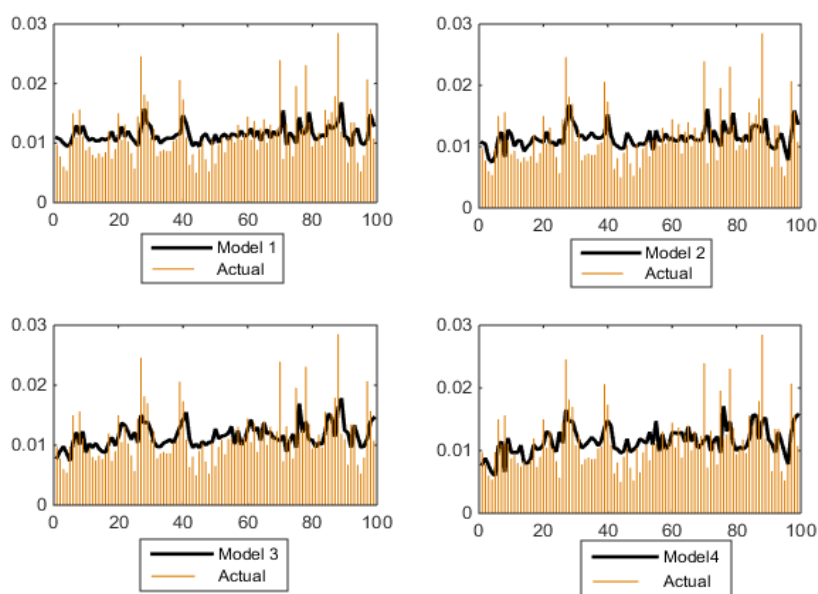

Figure 1. In sample performance comparison

In the next table we present the results of Diebold and Mariano test for forecast accuracy. The null hypothesis of this test is that two models have the same forecast accuracy. The alternative is that the second model performs better than the first one from a forecast perspective.

Table 3
Forecast comparison

Diebold-Mariano Test	In the sample		Out of sample (10 days forecast)	
Models comparison	DM Test Value	P-value	DM Test Value	P-value
Model 2 against Model 1	4.6218	100%	-2.30	97.68%
Model 3 against Model 1	4.1322	100%	-2.91	99.3%
Model 4 against Model 1	4.5827	100%	3.77	0.02%
Model 3 against Model 2	-4.8421	0%	-1.67	93.83%
Model 4 against Model 2	-3.8103	0%	13.63	0%
Model 4 against Model 3	4.2975	100%	15.77	0%

As we cans see from Table 3, Model 4 (HAR-RV-CJ-D) performs better in comparison with all other challenger models if we take into account out of sample forecast but in the sample forecast Model 1, the standard HAR-RV , estimates more accurate the observed Realized Volatility.
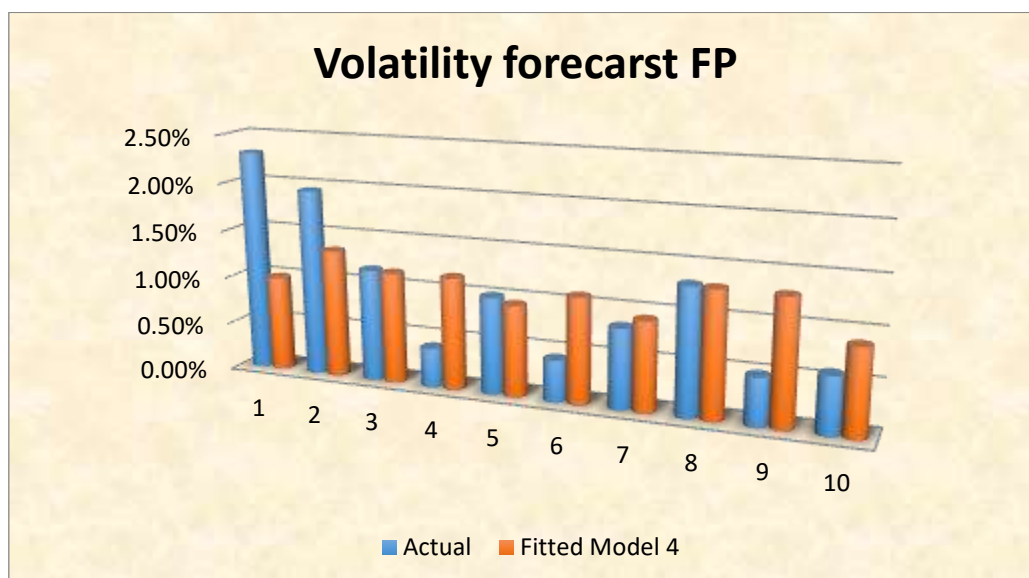

Figure 2. Actual versus fitted

In Figure 2, actual versus fitted values from Model 4 are presented. We can see that in the first four days the actual values are higher than the forecasted ones, but in the next six days this gap is partially eliminated.

Conclusion
The main goal of this paper was to forecast intra-day volatility for a very liquid equity traded at Bucharest Stock Exchange using HAR model and it's extension. The results showed us that this approach is not entirely suited for this equity since weekly and monthly historical volatilities weren't statistically significant. However, if we take into account the average trade duration between two trades from previous day and jumps, then model performance increased with almost 13% from 10.70% to 24.33%. Also, out of sample performance of this model is showing us a very good correlation between forecasted values and actual values.

References

Andersen, T., and Bollerslev, T. (1998): Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, Vol.39, pp: 885-905;

Andersen, T. G., Bollerslev, T., Diebold, F. X. (2007): Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *The Review of Economics and Statistics,* Vol.89, No. (4), pp:701–720;

Bandi, F. M., Russell, J. R., Yang, C. (2013): Realized volatility forecasting in the presence of timevarying noise, *Journal of Business & Economic Statistics* Vol.31, No.(3), pp:331–345;

Barndorff-Nielsen, O. E., Shephard, N. (2004): Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics,* Vol.2, No.(1), pp:1–37;

Cepoi C. (2014a): How Trading costs affect liquidity on Bucharest Stock Exchange?, *Procedia Economics and Finance*, Vol. 15, pp:1153-1158;

Cepoi C. (2014b): Price Leadership in Romanian Capital Market, *Procedia Economics and Finance*, Vol. 15, pp:1185-1191;

Cepoi C., Radu R. (2014): An intraday analysis of market efficiency. The case of Romania, *Monetary, Banking and Ffinancial Issues in Central and Eastern EU Member Countries,* Vol 2. pp: 249-255;

Corsi, F. (2009): A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* Vol.7, No,(2), pp:174–196;

Chen, Y., Härdle, W., and Pigorsch, U. (2010): Localized realized volatility modeling. *Journal of the American Statistical Association*, Vol.105, No.(492), pp:1376–1393;

Dragotă, V., Mitrică, E. (2004): Emergent capital markets efficiency. The case of Romania,*European journal of operational research*, Vol.155, pp:353-360;

Lahaye, J., and Shaw, P. (2014): Can we reject linearity in an HAR-RV model for the S&P 500? Insights from a nonparametric HAR-RV, *Economics Letters*, Vol.125, pp:43-46;

Radu, R., Cepoi C. (2015), Measuring trading costs and noise on Bucharest Stock Exchange. A Bayesian Approach, *Procedia Economics and Finance*, Vol. 32, pp:1459-1468;

Zhang, L., Mykland, P. A., Aït-Sahalia, Y. (2005), A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, Journal of the American Statistical Association*, Vol.100, pp: 1394–1411.