QUANTITATIVE METHODS IN ECONOMICS
Vol. XIII, No 2, 2012, pp. 107 – 114

# COMPARISON OF INTRADAY VOLATILITY FORECASTING MODELS FOR POLISH EQUITIES

**Magdalena E. Sokalska**
Queens College, City University of New York
e-mail: msokalska@qc.cuny.edu

**Abstract:** Several competing intraday volatility forecasting models for equally spaced data have been proposed in the literature. This study reviews a number of models and compares their forecasting performance using data on the market index of the Warsaw Stock Exchange. We also discuss choice criteria and issues specific to volatility forecast evaluation.

**Keywords:** forecasting volatility, ARCH, intraday equity returns

## INTRODUCTION

Volatile asset price fluctuations during the financial crisis of 2008-2009, the flash crash of May 2010 or substantial asset price swings in response to the unfolding of the euro zone crisis underscore the importance of a successful tool to forecast volatility throughout the day. While literature on predicting daily volatility is truly voluminous, intraday forecasting models are still scarce and their relative performance has not been subject to intense research. The aim of this paper is an evaluation of forecasting performance of several intraday volatility models for equally spaced returns on a selection of assets traded at the Warsaw Stock Exchange. Two next sections of the paper contain a review of the models to be evaluated and estimation results. The next chapter discusses evaluation methods and specific problems encountered when assessing volatility forecasts. Presentation of forecasting results is followed by a conclusion.

## INTRADAY VOLATILITY MODELS

Many authors, c.f. [Andersen and Bollerslev 1997], document pronounced periodic patterns in investors' activity, trading volume and return volatility throughout the day. These periodic (diurnal) fluctuations are one of the main reasons why applying standard daily volatility models to intraday data seems inappropriate. In the presence of a regular intraday pattern, unadjusted ARCH-type [Engle 1982] volatility models are misspecified.

In response to this observation, Andersen and Bollerslev [Andersen and Bollerslev 1997, 1998] propose a multiplicative component model for 5-minute returns on Deutschemark-dollar exchange rate and the S&P500 index. Let us assume the following notation. Days in the sample are indexed by $t$ ($t$ =1,…, $T$). Each day is divided into $m$- minute intervals referred to as bins and indexed by $i$ ($i$ =1,…, $N$). Andersen and Bollerslev chose the interval length $m$ = 5, whereas our paper selects $m$ =10. The current period is $(t,i)$. Price of an asset at day $t$ and bin $i$ is denoted by $P_{t,i}$. The continuously compounded return $r_{t,i}$ is modeled as: $r_{t,i} = \ln\left( P_{t,i} / P_{t,i-1} \right)$.

Andersen and Bollerslev assume the conditional variance of intraday asset returns to be a multiplicative product of daily and diurnal components. Intraday equity returns are described by the following process:

$$r_{t,i} = \sqrt{h_t s_i}\,\varepsilon_{t,i} \text{ and } \quad \varepsilon_{t,i} \sim N(0,1) \tag{1}$$

where: $h_t$ is the daily variance component, $s_i$ is the diurnal (periodic) variance pattern, and $\varepsilon_{t,i}$ is an error term. Andersen and Bollerslev [1998] add an additional component which takes account of the influence of macro-economic announcements on the foreign exchange volatility. For most of their models, the intra-daily volatility components are deterministic.

Engle and Sokalska [2012] argue that empirical evidence calls for specification of variance that includes a stochastic intraday component $q_{t,i}$ and extend the model to:

$$r_{t,i} = \sqrt{h_t s_i q_{t,i}}\,\varepsilon_{t,i} \text{ and } \quad \varepsilon_{t,i} \sim N(0,1) \tag{2}$$

Both models (1) and (2) require an exact characterization of the variance components. In our paper, the daily variance component is estimated using ARCH –type specification for a longer sample, going back a number of years. We adopt a GARCH(p,q) process [Bollerslev 1986]:

$$\begin{gathered} r_t = \sqrt{h_t}\,\zeta_t \qquad \zeta_t \sim N(0,1) \\ h_t = w_d + \sum_{k=1}^{p} \beta_k h_{t-k} + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 \end{gathered} \tag{3}$$

where $\zeta_t$ is an error term for daily returns $r_t$, whereas $w_d$, $\alpha_{j\text{-}s}$ and $\beta_{k\text{-}s}$ are parameters of the daily variance equation.

The diurnal component could be modeled in a number of ways. In this paper it is estimated as a variance of returns in each bin after deflating squared returns by the daily variance component.

$$E\left(\frac{r_{i,t}^2}{h_t}\right) = s_i E\left(q_{i,t}\right) = s_i \tag{4}$$

Engle and Sokalska model the residual intraday volatility as a GARCH(p,q) process. Empirical analysis indicates that GARCH(1,1) proves to be the most successful choice:

$$q_{t,i} = \omega + \alpha^{(10)}(r_{t,i-1} / \sqrt{h_t s_{i-1}})^2 + \beta^{(10)} q_{t,i-1} \tag{5}$$

where ω, $\alpha^{(10)}$ and $\beta^{(10)}$ are parameters of the variance equation for the intraday returns adjusted by the daily and periodic components. It has been shown that this multistep estimator is consistent and asymptotically normal.

As mentioned at the beginning of the paper, since GARCH(1,1) does not take account of intraday periodicity, it is clearly misspecified when applied to intraday equally spaced data. Nevertheless, voluminous forecasting literature documents that misspecified but parsimonious models quite often yield superior forecasts in comparison with correctly specified but more complicated models. Therefore it is worthwhile to examine the relative forecasting performance of the model that repeatedly has been shown to be the most successful predictor for daily data. Consequently the third evaluated option involves GARCH (1,1) for the original (non-standardized) intraday returns:

$$\begin{aligned} r_{t,i} &= \sqrt{g_{t,i}}\varepsilon_{t,i} \qquad \varepsilon_{t,i} \sim N(0,1) \\ g_{t,i} &= w_g + \alpha_g r_{t,i-1}^2 + \beta_g g_{t,i-1} \end{aligned} \tag{6}$$

where: $g_{t,i}$ is the intraday conditional variance and $w_g$, $\alpha_g$ and $\beta_g$ are parameters of the variance equation.

## EMPIRICAL RESULTS

### Data

Our dataset is obtained from Bloomberg and consists of both daily and 10-minute intraday logarithmic returns on the broad market index WIG at the Warsaw Stock Exchange. Intraday models are estimated for the period 9 December 2011-30 April 2012. Forecasting is performed for the period 2 May 2012-31 May 2012. The time series on the WIG index for daily component estimation starts on 31 December 2003. The overnight return in bin zero has been deleted and the intraday return for the first bin is a logarithmic difference between the last price for that bin and the opening price. Sokalska [2010] and Engle and Sokalska [2012] give an extended explanation for the reasons and consequences of skipping overnight returns for intraday multiplicative volatility models. Intraday returns on WIG cover the span of the continuous trading at the exchange between hours 9:00am and 5.20pm. This translates into 50 10-minute bins throughout the trading day.

## Estimation Results

Forecasting evaluation will include three models reviewed in Section INTRADAY VOLATILITY MODELS: the model of Engle and Sokalska [2012] (**ES**) described by equations (2)-(5), a variant of the model of Andersen and Bollerslev [1997] (**AB**) described by equations (1),(3) and (4) and GARCH(1,1) for unadjusted intraday returns (**G11**) described by (6).

Table 1 presents estimation results of the daily model (3). GARCH (2,1) seems to fit data best in-sample. The sum of coefficients ($\alpha + \beta_1 + \beta_2$) equals (0.996) and is close to but smaller than one. This indicates high persistence (high degree of volatility clustering) in daily stock returns.

Table 1. GARCH Results for Daily Data

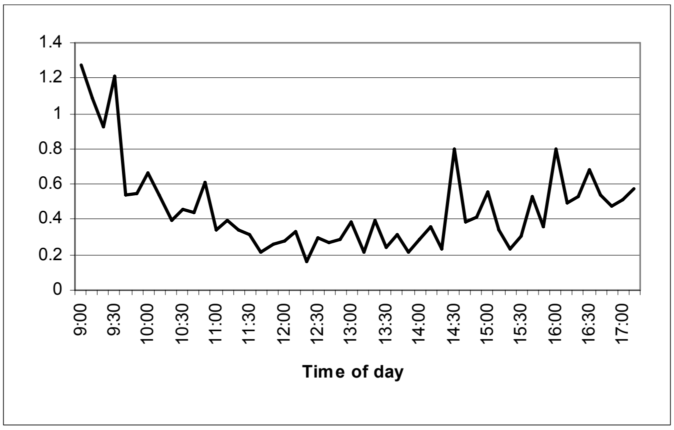
| | Coefficient | Std. Error | t-Statistic |
|---|---|---|---|
| $\omega_d$ | 7.92E-07 | 2.70E-07 | 2.934559 |
| $\alpha$ | 0.035183 | 0.007838 | 4.488603 |
| $\beta_1$ | 1.580650 | 0.101146 | 15.62745 |
| $\beta_2$ | -0.619346 | 0.093202 | -6.645235 |

Source: own calculations

*Notes:* This table presents estimation results for GARCH(2,1) model for daily returns on the WIG index. Sample period: January 2004 -April 2012. Symbols $\alpha$, $\beta_1$, $\beta_2$ and $\omega_d$ denote parameters from the conditional variance equation (3).

One step ahead variance forecasts obtained from the daily model are used to scale intraday returns. Then a periodic (diurnal) component is estimated as variance of returns in each of 50 bins. Figure 2 presents a summary picture of the periodic component estimates for WIG returns.

Figure 2. Diurnal (Periodic) Variance Component




Source: own calculations

Approximately first two hours of the trading session are marked by high volatility. A subsequent quiet period in the middle of the day is followed by a sharp spike in volatility at 2.30 pm. At this time (8.30 Eastern Standard Time) important macroeconomic announcements are scheduled before the opening of North American equity markets. Volatility stays high at the WSE at 3 pm, when the US markets actually open. With an exception of a short period around 3.30 pm volatility remains elevated until the session closes.

Daily and periodic components will be used for calculating variance forecasts for both ES and AB models. In the ES model (2), the third variance component is estimated as a GARCH(1,1) using returns that are adjusted by the daily and diurnal volatility patterns. Table 2 contains estimation results. Attempts to fit higher order models yield statistically insignificant coefficients of lags bigger than one.

Table 2. GARCH Results for Intraday Returns Standardized by Daily and Periodic Components

| | Coefficient | Std. Error | t-Statistic |
|---|---|---|---|
| $\omega$ | 0.169809 | 0.024999 | 6.792545 |
| $\alpha^{(10)}$ | 0.069851 | 0.008502 | 8.215637 |
| $\beta^{(10)}$ | 0.752523 | 0.031315 | 24.03063 |

Source: own calculations

*Notes:* This table presents estimation results for GARCH(1,1) model for intraday WIG returns that have been previously adjusted using daily and diurnal variance components. Sample period 9 December 2010- 30 April 2012. Symbols $\alpha^{(10)}$, $\beta^{(10)}$ and $\omega$ denote GARCH parameters from the variance equation (5).

We may note that the persistence measure is equal to 0.82 and is much lower than in the daily model. This is understandable since intraday returns have already been scaled by daily and periodic components.

Finally for forecasting comparison, we estimate a GARCH model for unadjusted intraday data.

Table 3. GARCH Results for Unadjusted Intraday Returns

| | Coefficient | Std. Error | t-Statistic |
|---|---|---|---|
| $\omega_g$ | 9.24E-08 | 6.80E-09 | 13.59734 |
| $\alpha_g$ | 0.114129 | 0.008186 | 13.94179 |
| $\beta$ | 0.792513 | 0.011945 | 66.34819 |

Source: own calculations

*Notes:* This table presents estimation results for GARCH(1,1) model for unadjusted WIG returns. Sample period 9 December 2010- 30 April 2012. Symbols $\alpha_g$ , $\beta_g$ and $\omega_g$ denote GARCH parameters from the variance equation (6).

Persistence measure is equal to 0.91 and is bigger than for the ES model. We cannot make a meaningful comparison of this measure with daily data because of the exclusion of overnight return. As in the previous case, higher order GARCH processes were ruled out empirically.

## FORECAST EVALUATION

Based on the estimated models presented in the previous section, we forecast volatility of 10-minute logarithmic returns on WIG in the period 2-31 May 2012. Evaluation of volatility forecasts involves an additional difficulty due to fact that the forecasted phenomenon is not directly observable and can be only measured with an error. For our analysis we adopt a popular solution to this problem. We evaluate conditional variance forecasts using the actual squared return over the forecast horizon. Furthermore since we can measure volatility only with an error, this introduces biases in many popular loss functions used for forecast evaluation. Since, under MSE and LIK loss functions optimal forecasts are unbiased (Patton [2011]), this paper uses these two loss functions.

In order to find the preferred model we are looking for a minimum average loss. The mean squared error loss function is defined as $L_{1t,i} = \left( {r_{t,i}}^2 - v_{t,i}^f \right)^2$, where $v_{t,i}^f$ is the forecast of conditional variance of 10-minute logarithmic returns obtained using each of the reviewed models. Since under this loss function the errors are squared, it is very sensitive to large errors. Therefore we add an evaluation criterion more robust to outliers, the out-of-sample likelihood (predictive likelihood-based) loss function calculated as $L_{2t,i} = \ln v_{t,i}^f + \frac{{r_{t,i}}^2}{v_{t,i}^f}$ [Bjørnstad 1990].

For model AB we construct the forecast of conditional variance by multiplying the daily variance component (3), using parameters shown in Table 1, by the periodic variance component (4) depicted by Figure 2 ($v_{t,i}^{AB} = h_t s_i$)

For model ES, following equation (2), the volatility forecast is obtained by multiplication of daily, periodic and intraday components described by (3), (4) and (5), respectively ($v_{t,i}^{ES} = h_t s_i q_{t,i}$). Equation (6) is used to obtain intraday variance for G11 model ($v_{t,i}^{G11} = g_{t,i}$). For all the components that are modeled as GARCH processes, forecasts are obtained in a sequential procedure on the basis of estimated parameters and the volatility forecast calculated at previous bin, as well as actual returns from the previous bin.

Table 4 contains average loss values for two loss functions $L_{1t}$ and $L_{2t}$ and 3 models: $L_l = \frac{1}{\tau}\sum_{ti=1}^{\tau} L_{lti}$ where l=1,2; $\tau$ denotes the length of the forecasting period and $\tau$= 1050 bins. Both $L_1$ and $L_2$ criteria favour the ES model and the misspecified but parsimonious GARCH(1,1) appears to be least successful.

Table 4. Average one-period-ahead forecast errors

| Criterion / Model | ES | AB | GARCH11 |
|---|---|---|---|
| $L_1$ MSE* | 6.89019 | 6.96635 | 7.16529 |
| $L_2$ LIK | -12.68812 | -12.64860 | -12.625177 |

Source: own calculations
* Values for $L_1$ (mean squared error) criterion need to be multiplied by $10^{-12}$

It is, however, difficult to evaluate the practical importance of loss function differentials. Table 5 shows results of a significance test for the differences between evaluation criteria. It contains t-values and p-values of the Diebold-Mariano test [Diebold-Mariano 1995]. For example, the negative value of the t-statistic in the column marked as ES-AB indicates that the mean difference between forecast errors of the first model (ES) and the second model (AB) is negative, that is the ES model tends to yield smaller forecast errors than the AB model. The difference is statistically significant at the 10% level. For the $L_2$ criterion, the difference between the same two models is also negative and significant at the 5% level. This table also indicates that ES yields better forecasts than the simple GARCH(1,1) model, and the differences for both loss functions are significant at the 5% level. Finally, although AB forecasts better than GARCH (1,1) on average, the mean differential is not significant at the 10% level.

Table 5. T-values and p-values for forecast accuracy Diebold-Mariano test

| Criterion / Model | ES - AB | ES - GARCH | AB - GARCH |
|---|---|---|---|
| $L_1$ MSE t-value | -1.80 | -3.01 | -1.48 |
| p-value | 0.072 | 0.003 | 0.139 |
| $L_2$ LIK t value | -2.10 | -2.89 | -0.59 |
| p-value | 0.036 | 0.004 | 0.556 |

Source: own calculations

In sum, the ES model is shown to offer better volatility forecasts than the other two competing models.

## CONCLUSION

This paper reviews several volatility models for equally spaced intraday data and investigates their relative forecasting performance using the example of the broad market index at the Warsaw Stock Exchange. It finds that the multiplicative ES model tends to offer better forecasts than the alternatives. There are a number of ways in which this study could be extended. It would be interesting to investigate if the forecasting performance of the analyzed models depends on asset characteristics, for example, liquidity. Additionally, for the multiplicative models, particular components could be specified in a number of different ways. An investigation of preferable specifications will be the subject of future research.

## REFERENCES

Andersen T.G., Bollerslev T. (1997) Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance, 4, 115-158.

Andersen T.G., Bollerslev T. (1998) DM-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer-run dependencies, Journal of Finance, 53, 219-265.

Bjørnstad J. F. (1990) Predictive likelihood: A review. Statistical Science, 5, 242– 265.

Bollerslev T (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31, 307-327.

Diebold F., Mariano R. (1995) Comparing forecasting accuracy, Journal of Business and Economic Statistics 13, 253-63.

Engle R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica 50(4), 987-1007.

Engle R.F., Sokalska M. (2012), Forecasting intraday volatility in the US equity market. Multiplicative component GARCH, Journal of Financial Econometrics, 10, 54-83.

Patton A. (2011) Volatility forecast comparison using imperfect volatility proxies, Journal of Econometrics, 160, 246-256.

Sokalska M. (2010) Intraday volatility modeling: The example of the Warsaw Stock Exchange, Quantitative Methods in Economics, 11, 139-144.