Forecasting intraday traded volume with the Weibull ACV model: an application to Polish stocks

Roman Huptas[1]

Abstract

The trading volume is one of the basic measures of intensity of trading activity and plays a crucial role in the financial market microstructure literature. The aim of the paper is to examine the out-of-sample point and density forecasting performance of the Bayesian linear Autoregressive Conditional Volume (ACV) model with Weibull distribution for the error term for intra-daily volume data. The analysis follows a rolling window scheme and is based on real-time 5-minute intra-daily traded volume data for stocks quoted on the Warsaw Stock Exchange, which is a leading stock market in Central and Eastern Europe. It is concluded that in terms of point forecasts the considered Bayesian linear ACV model significantly outperforms such benchmarks as the naïve or the Rolling Means methods. Moreover, the exponential error ACV models generate more accurate point forecasts than the structures with the Weibull distribution, but the differences between forecast errors are not in all cases statistically significant. The results obtained from analysis of density forecasts indicate that in most cases the linear ACV model with the Weibull distribution provides significantly better density forecasts as compared to the linear ACV model with exponential innovations in terms of the log-predictive score.

Keywords: *trading volume, forecasting, ACV model, market microstructure, Bayesian inference*

JEL Classification: C11, C22, C53

1 Introduction

The trading volume is one of the key characteristics of liquidity on stock markets and plays a very important role in the literature on financial market microstructure. It can be very important in order to understand stock trading and behaviour of market participants.

In existing literature on the subject a lot of attention is paid to the issue of examining the dependencies between trade size and other financial variables, such as price (Easley and O'Hara, 1987; Foster and Viswanathan, 1990 for example) or volatility (Tauchen and Pitts, 1983; Karpoff and Boyd, 1987; Andersen, 1996; Manganelli, 2005 for example). In turn, forecasting of trading volume in stock markets definitely has not been a central point of financial econometrics for years. Only few works deal exclusively with the modelling and predicting volume on stock exchanges. Kaastra and Boyd (1995) designed neural networks to forecast monthly futures trading volume. Lo and Wang (2000) applied the principal component analysis to a decomposition of the volume. Białkowski et al. (2008) presented

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: huptasr@uek.krakow.pl.

a new methodology for modelling the dynamics of intraday volume, which allows for a significant reduction of the execution risk in Volume Weighted Average Price (VWAP) orders. Brownlees et al. (2011) used a Component Multiplicative Error Model for intra-daily volumes, in which the conditional expectation of volume is the product of three multiplicative elements: a daily component, an intra-daily periodic component and an intra-daily dynamic non-periodic component. Predicting intraday trading volume is also a topic of interest of Satish et al. (2014). Finally, Ito (2016) developed and applied the Spline-DCS model to forecasting the high-frequency traded volume of selected equity and foreign currency exchange pairs.

The aim of this paper is to present results of a pilot study in which the quality of the out-of-sample point and density forecasts of intra-daily volume data generated by the linear Autoregressive Conditional Volume (ACV) model with the Weibull distribution for the error term is examined. Forecasting models are developed on the basis of the 5-minute volume data of some representative, widely traded stocks quoted on the Warsaw Stock Exchange. Considered models follow a rolling window forecast scheme. The evaluation of point forecast accuracy is performed by comparison forecast errors of the considered linear ACV models and the benchmark models (the random walk without drift and the Rolling Means technique). To compare point forecasts generated by all forecasting models analysed in the study we use standard forecast accuracy measures: the root mean square forecast error (RMSFE) and the mean absolute forecast error (MAFE), and Diebold and Mariano (1995) test. The comparison of density forecasts is obtained by using the log-predictive scores (LPSs) and the Amisano and Giacomini (2007) test.

The rest of the paper is scheduled as follows. In Section 2, we briefly present in more details the Autoregressive Conditional Volume model and the Bayesian estimation for the models under consideration, along with relevant MCMC methods is also considered. Then, Section 3 describes the forecasting performance. The results of the empirical study are discussed in Section 4. Finally, the last section presents the conclusions of the paper.

2 Research methodology

In the presented analysis the Autoregressive Conditional Volume (ACV) model suggested by Manganelli (2005) is a basic econometric tool. The ACV model is definitely a volume counterpart of the Autoregressive Conditional Duration (ACD) model, introduced in the seminal paper of Engle and Russell (1998) and applied to model the dynamics of the time intervals between successive events of the transaction process.

The general ACV model for the volume v_i , $i = 0, 1, 2,...,N$ (with N standing for the total number of observations) is defined multiplicatively as:

$$v_i = \Phi_i \cdot \varepsilon_i \,, \tag{1}$$

$$\Phi_i = E(v_i \mid \Im_{i-1}, \theta) = \Phi_i(v_{i-1}, \ldots, v_1; \theta) \tag{2}$$

where Φ_i represents the conditional mean value of trading volume, $\Im_{i-1}$ is the information set available at time t_{i-1}, θ is the vector of unknown parameters, and ε_i is a sequence of positive, identically and independently distributed random variables with density function $f_\varepsilon(\varepsilon_i)$ and mean value $E(\varepsilon_i) = 1$. Equations (1)-(2) formulate a very general setup that allows for a variety of specific models. In this study the following particular, linear ACV model, which remains in line with Manganelli (2005) is used:

$$v_i = \Phi_i \cdot \varepsilon_i \,, \tag{3}$$

$$\Phi_i = \omega + \alpha \cdot v_{i-1} + \beta \cdot \Phi_{i-1} \,, \tag{4}$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$. These inequality restrictions are imposed in order to ensure positive conditional volumes for all possible realizations of random variables v_i, existence of the unconditional mean of trading volume and stationarity of the model. Moreover, we assume that ε_i follows either the exponential, or the Weibull distribution, yielding ACV specifications denoted as Exp-ACV and W-ACV. The corresponding density functions (under the assumption that the error terms have unit expectation) are given by:

1. $\varepsilon_i \sim Exp(1)$:

$$f_\varepsilon(\varepsilon_i) = \exp(-\varepsilon_i),\ \varepsilon_i > 0 \,, \tag{5}$$

2. $\varepsilon_i \sim W(\lambda, \gamma)$ with parameter $\lambda = 1/\Gamma(1 + 1/\gamma)$:

$$f_\varepsilon(\varepsilon_i) = \frac{\gamma}{\lambda} \cdot \left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma-1} \cdot \exp\left[-\left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma}\right], \varepsilon_i > 0 \,, \gamma > 0 \,. \tag{6}$$

In order to estimate parameters of the considered ACV models, the Bayesian approach is applied. It is worth noting that the literature on the Bayesian inference in modelling and forecasting financial high-frequency data with the use of the ACD-type models is still limited (e.g. Brownlees and Vannucci, 2013; Huptas, 2014, 2016; Gerlach et al., 2016). Bayesian estimation of the ACV models outlined above requires certain prior assumptions. We assume that all parameters – whenever possible – are *a priori* independent. Moreover, in order to express the lack of prior knowledge, fairly diffuse prior distributions are assumed, so that the data dominates the inference about the parameters through the likelihood function.

Specifically, for all parameters of Equation (4) we propose the normal distributions with zero mean and standard deviation of five, adequately truncated, due to relevant restrictions imposed on the parameters in each model. For the ACV model with the Weibull innovations, the prior density for parameter γ is also specified as density of the normal distribution with zero mean and standard deviation of five adequately truncated.

The inference was conducted using MCMC techniques. The Metropolis-Hastings (MH) algorithm (Gamerman and Lopes, 2006) with a multivariate Student's *t* candidate generating distribution with three degrees of freedom and the expected value equal to the previous state of the Markov chain was used to generate a pseudo-random sample from the posterior distribution. The covariance matrix was obtained from initial cycles, which were performed to calibrate the sampling mechanism. Convergence of chain was carefully examined by starting the MCMC scheme from different initial points and checking trace plots of iterates for convergence to the same posterior. Acceptance rates were sufficiently high and always exceed 50%, indicating good mixing properties of the posterior sampler. The final results and conclusions were based on 100,000 draws, preceded by 50,000 burn-in cycles. All codes were implemented by the author and ran using the GAUSS software, version 13.0.

3 Forecasting procedure

Forecasts are determined using a rolling window prediction scheme with a fixed-length window of 4816 observations. First, we estimate ACV models and then generate one- and multi-step-ahead point and density forecasts for horizons h=2, 3, 5, 10 in each model. Then the estimation window is moved by one 5-minute bin in relation to the earlier one, the models are re-estimated and future volumes are predicted. The forecasting procedure is repeated 200 times. This means that the out-of-sample period used to evaluate forecast performance contains 200 observations.

The point forecasts are calculated as arithmetic means of draws from the predictive distributions. For assessing the quality of the point predictions, the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) are used. Further, the pair-wise Diebold and Mariano (1995) tests of equal point forecast accuracy are carried out. In this paper, to benchmark our Bayesian ACV models in terms of the one-step-ahead point forecasts we use a naïve strategy based on a random walk model and a Rolling Means (RM) method.

This research focuses also on density forecasts for traded volumes. The density forecast is a predictive density function evaluated at realization of intra-daily volume. In turn, the Bayesian density forecast is obtained as the arithmetic mean of the density forecasts corresponding to draws from the posterior distribution. Taking the logarithm of density forecast, we obtain the so-called log predictive score (LPS). To compare the performance of two different density forecasts, we calculate the difference between the corresponding LPSs and we follow the Amisano and Giacomini (2007) test.

4 Data and empirical results

The empirical analysis and forecasting study are based on 5-minute intra-daily volume data of three actively traded companies listed in the WIG20 Index of the Warsaw Stock Exchange (WSE), namely: the Polish Telecommunications (TPSA, currently Orange Polska S.A.), the PKOBP SA bank (PKOBP) and the KGHM SA company (KGHM). The data set comprises intra-daily observations spanning 60 consecutive trading days from March 23, 2009 to June 18, 2009. It must be stressed that each 5-minute bin volume was computed as the sum of all transaction volumes exchanged within the time interval covered by the bin. Additionally, the analysis covers only transactions carried out in the continuous trading phase which in the case of the WSE in 2009 fall on between 10:00 and 16:10. Taking into account that the analysis is conducted in a rolling window scheme, the first subsample starts on the first 5-minute bin of March 23, 2009 and ends on the 86th 5-minute bin of June 12, 2009. The second estimation window is moved towards the first one by one bin, so the second subsample starts on the second 5-minute bin of March 23, 2009 and ends on the first 5-minute bin of June 15, 2009. The last 200th subsample starts on the 28th 5-minute bin of March 25, 2009 and ends on the 27th 5-minute bin of June 17, 2009.

It is well known in the financial literature that trading volumes exhibit a well-pronounced intraday periodic pattern. In a way similar to Bauwens and Veredas (2004), the intraday periodic pattern for intra-daily volumes was estimated using the Nadaraya-Watson kernel estimator of regression of the trading volume on the time of the day. Following Engle and Russell (1998) the time-of-day adjusted volumes were computed by dividing plain intra-daily volumes by estimated periodic component.

Table 1 presents the values of the MAFEs for all models for TPSA, PKOBP and KGHM. It also contains the results of the one-sided Diebold-Mariano tests. Firstly, the point forecasts from the Exp-ACV models lead to the lowest MAFEs for all stocks and for all horizons. In the case of the one-step-ahead forecasts the ACV models always significantly outperform the naïve and the Rolling Means benchmarks. Moreover, the forecasts from the Exp-ACV models are significantly more accurate than the point forecasts from the W-ACV models for almost all the horizons and assets. However, it emerges that the point forecasts in both ACV specifications are not significantly different from each other in the case of KGHM for horizons h=1, 5, 10 and in the case of PKOBP for horizon h=5.

In the following part of the article, there are discussed results of the RMSFEs. These forecast accuracy measures (in levels) for all models for TPSA, KGHM and PKOBP are presented in Table 2. Furthermore, we also report the results of the Diebold and Mariano tests. An analysis of values of these measures lead to the conclusion that, similarly to the MAFEs, the smallest RMSFE errors are obtained for the Exp-ACV models for all stocks and for almost all horizons with the exception of h=5 for PKOBP. In the case of the one-step-ahead point forecasts

Model	h=1	h=2	h=3	h=5	h=10
TPSA					
Naïve	76009.7 *	-	-	-	-
RM	78353.5 ***	-	-	-	-
Exp-ACV	**60749.6**	**61078.8**	**58855.4**	**58738.1**	**57484.2**
W-ACV	61781.5 **	62103.1 **	59791.4 **	59878.8 ***	58521.5 ***
PKOBP					
Naïve	20259.3 **	-	-	-	-
RM	61573.3 ***	-	-	-	-
Exp-ACV	**18112.9**	**19112.9**	**19279.3**	**20082.0**	**21952.2**
W-ACV	18127.0 *	19128.3 *	19302.2 **	20088.6	21974.1 **
KGHM					
Naïve	14776.8 **	-	-	-	-
RM	19546.2 ***	-	-	-	-
Exp-ACV	**12367.1**	**12949.1**	**12668.7**	**12555.9**	**12634.4**
W-ACV	12379.6	12980.4 **	12700.8 **	12559.4	12643.6

Notes: The lowest MAFEs are shown in bold type. *, ** and *** indicate significance of the one-sided Diebold-Mariano test at the 0.1, 0.05 and 0.01 significance level, respectively. The Diebold-Mariano test of model with the lowest MAFE against a given model in a row.

Table 1. Mean absolute forecast errors for TPSA, PKOBP and KGHM.

the ACV models always significantly outperform the naïve and the Rolling Means benchmarks except for RM for TPSA. Moreover, it should be noted that, for TPSA and at all the horizons, the point forecasts from the Exp-ACV model are not significantly superior (in terms of RMSFEs) as compared with the W-ACV specification. In turn, in the case of PKOBP and KGHM forecasts generated by the Exp-ACV models are statistically more accurate than the W-ACV forecasts only at h=2, 3, 10 for PKOBP and at h=2, 3 for KGHM. For the remaining horizons the differences between root mean squared errors are statistically insignificant.

Table 3 shows the differences between the average values of the LPSs of both ACV models for all considered assets. It also presents the results of the Amisano and Giacomini tests. On the basis of the information included in Table 3, it can be noted that in the case of PKOBP and KGHM, the density forecasts from the W-ACV model are significantly superior

Model	**h=1**	**h=2**	**h=3**	**h=5**	**h=10**
TPSA					
Naïve	260661.6 *	-	-	-	-
RM	196736.0	-	-	-	-
Exp-ACV	**192449.6**	**190223.6**	**187175.1**	**184995.8**	**185900.5**
W-ACV	194331.8	191882.3	188354.9	185772.3	186280.8
PKOBP					
Naïve	28726.1 ***	-	-	-	-
RM	68847.2 ***	-	-	-	-
Exp-ACV	**23343.4**	**24146.4**	**24046.1**	25453.9	**26408.2**
W-ACV	23356.5	24163.6 *	24069.2 **	**25453.0**	26428.4 **
KGHM					
Naïve	24127.4 **	-	-	-	-
RM	23815.4 ***	-	-	-	-
Exp-ACV	**18517.9**	**19401.9**	**19271.1**	**19475.7**	**19890.9**
W-ACV	18539.9	19455.4 **	19322.1 *	19498.2	19898.6

Notes: The lowest RMSFEs are shown in bold type. *, ** and *** indicate significance of the one-sided Diebold-Mariano test at the 0.1, 0.05 and 0.01 significance level, respectively. The Diebold-Mariano test of model with the lowest RMSFE against a given model in a row.

Table 2. Root mean squared forecast errors for TPSA, PKOBP and KGHM.

Model	h=1	h=2	h=3	h=5	h=10
TPSA					
W-ACV vs. Exp-ACV	0.1773 *	0.0471	0.0185	0.0031	-0.0129
PKOBP					
W-ACV vs. Exp-ACV	0.0025 ***	0.0021 **	0.0023 **	-0.0049	0.0017 *
KGHM					
W-ACV vs. Exp-ACV	-0.0094 **	-0.0085 **	-0.0084 **	-0.0092 **	-0.0120 ***

Notes: Positive (negative) values indicate that the first model has a higher (lower) average LPS than the second model. *, ** and *** indicate significance of the Amisano-Giacomini test at the 0.1, 0.05 and 0.01 significance level, respectively.

Table 3. The differences between the average values of the LPSs and results of the Amisano and Giacomini test for TPSA, PKOBP and KGHM.

as compared with the Exp-ACV specification with the exception of horizon h=5 for PKOBP. Different situation can be observed in the case of TPSA. Density forecasts from the ACV model with the Weibull error distribution are qualitatively similar to density forecasts from the Exp-ACV model, that is differences between the values of the LPSs are statistically insignificant, except for h=1.

Conclusions

The objective of this study is to verify predictive performance (in terms of point and density forecasts) of the ACV model with the Weibull distribution for the error term for 5-minute intra-daily volume data. Our main findings can be summarised as follows.

It is concluded that in terms of point forecasts, MAFEs and RMSFEs the considered Bayesian linear ACV models significantly outperform such benchmarks as the naïve or the Rolling Means methods. Moreover, the exponential error ACV models generate more accurate point forecasts than the ACV specifications with the Weibull distribution, but the differences between forecast errors are not in all cases statistically significant. The comparison of the ACV models also allows us to conclude that in the case of more actively traded stocks (PKOBP and KGHM) the linear ACV model with the Weibull distribution provides significantly more accurate density forecasts than the linear ACV model with exponential innovations in terms of the log-predictive score.

The presented results are based on a pilot study taking into account only the ACV model with the Weibull distribution for innovations. In future research we will try to focus on alternative ACV model specifications with different error distributions such as the Burr or the generalized gamma distribution, longer forecast horizons and different sampling frequencies.

Acknowledgements

This work has been financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Andersen, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance, 51,* 169-204.

Amisano, G., & Giacomini, R. (2007). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics, 25(2),* 177-190.

Bauwens, L., & Veredas, D. (2004). The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics, 119,* 381-412.

Białkowski, J., Darolles, S., & Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking and Finance, 32,* 1709-1722.

Brownlees, Ch., Cipollini, F., & Gallo, G.M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics, 9,* 489-518.

Brownlees, Ch., & Vannucci, M. (2013). A Bayesian approach for capturing daily heterogeneity in intra-daily durations time series. *Studies in Nonlinear Dynamics & Econometrics, 17(1),* 21-46.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13(3),* 253-263.

Easley, D., & O'Hara, M. (1987). Price, trade size and information in securities markets. *Journal of Financial Economics, 19,* 69-90.

Engle, R.F., & Russell, J.R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica, 66,* 1127-1162.

Foster, D., & Viswanathan, S. (1990). A theory of intraday variations in volume, variance and trading costs in securities market. *Review of Financial Studies, 3,* 593-624.

Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press.

Gerlach, R., Peiris, S., & Lin, E.M.H. (2016). Bayesian estimation and inference for log-ACD models. *Computational Statistics, 31,* 25-48.

Huptas, R. (2014). Bayesian estimation and prediction for ACD models in the analysis of trade durations from the Polish stock market. *Central European Journal of Economic Modelling and Econometrics, 6,* 237-273.

Huptas, R. (2016). The UHF-GARCH-Type Model in the Analysis of Intraday Volatility and Price Durations – the Bayesian Approach. *Central European Journal of Economic Modelling and Econometrics, 8,* 1-20.

Ito, R. (2016). Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data. https://doi.org/10.17863/CAM.5751.

Kaastra, I., & Boyd, M. (1995). Forecasting futures trading volume using neural networks. *Journal of Futures Markets, 15,* 953-970.

Karpoff, J., & Boyd, M. (1987). The relationship between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis, 22,* 109-126.

Lo, A., & Wang, J. (2000). Trading volume: Definition, data analysis, and implication of portfolio theory. *Review of Financial Studies, 13,* 257-300.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets, 8,* 377-399.

Satish, V., Saxena, A., & Palmer, M. (2014). Predicting intraday trading volume and volume percentages. *The Journal of Trading, 9,* 15-25.

Tauchen, G., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica, 51,* 485-505.