DYNAMIC ECONOMETRIC MODELS
Vol. 8 – Nicolaus Copernicus University – Toruń – 2008

Piotr Fiszeder
Nicolaus Copernicus University in Toruń

How to Increase Accuracy of Volatility Forecasts Based on GARCH Models

1. Introduction

There is a large literature on volatility forecasting (see Poon and Granger, 2003), but nevertheless it is difficult to extract a coherent set of prescriptions concerning the most appropriate empirical procedure for tackling this issue. The results of empirical analyses are unclear and often even contradictory. Various conditional variance specifications within the parametric GARCH class of models were proposed in the literature, but there is no consensus on the relative quality of out-of-sample forecasts of those formulations. Analyses with GARCH models for the Polish stock market were performed among others by Piontek (2003), Doman and Doman (2004), Fiszeder (2004a, 2004b, 2005), Osiewalski, Pajor and Pipień (2004) and Pipień (2006). Only in the investigations of Doman and Doman and Fiszeder were intraday data used for evaluation of forecasts quality. In this paper a significantly wider class of GARCH models, especially models extended with additional information, was used and the research period was wider, which could significantly influence results. The main purpose of this study is to compare a performance of the different specifications of GARCH models for predicting volatility. Additional information used in construction of the GARCH model or in estimation of its parameters does not always lead to an increase in accuracy of volatility forecasts.

The plan for the rest of the paper is as follows. Section 2 outlines the competing methods used in the analysis and the measures used to assess the performance of the candidate models. In section 3 accuracy of volatility forecasts for WIG20 index was analysed and section 4 concludes.

2. Methods of Volatility Forecasting and Evaluation Measures

The forecasting methods used in the paper can be divided into three groups. The first one includes GARCH models whose parameters were estimated on daily data with only closing prices: GARCH, a GARCH model with conditional Student-t distribution (GARCH-t), IGARCH, GARCH-M, a GARCH model with threshold GARCH-M effect (GARCH-MT), GJR, TGARCH and FIGARCH. An EGARCH model was also used; however, owing to problems with estimation of parameters for some periods (flat likelihood function) the results were omitted for this model[1]. The second one contains other methods for predicting volatility estimated on data with the closing prices[2]: a random walk model for variance (RW), a historical average model (HA), a moving average model for variance (MA), an exponential smoothing model for variance (ES) and a stochastic volatility model (SV). The third group includes methods which use additional information: a GARCH model estimated on scaled true range data (GARCH on TR), a GARCH models extended with additional explanatory variables from time $t-1$ - squared true range data (GARCH with TR), the squared difference between logarithms of daily high and low (GARCH with HL), a sum of squared 5-minute returns (GARCH with 5MR), the squared daily return from the S&P 500 stock index (GARCH with S&P). The following models can also be included in this group: GARCH models describing seasonal fluctuations (GARCH with seas.), a day of the week effect (GARCH with day) and an effect connected with holidays (GARCH with hol.) – models extended with dummy variables, a GARCH model estimated on residuals from the regression of daily returns of the WIG20 index on daily returns of the S&P 500 index (S&P in mean), a random walk model used for the sum of squared 5-minute returns (RW intra). The inclusion of the S&P 500 index in the analysis results from strong connections between the Polish stock index and indices of the U.S. stock market (see Fiszeder and Romański, 2002).

Because of the limited size of the publication, presentation of basic specifications of the models was omitted. They can be found, for example, in papers of Fiszeder (2004a, 2004b). Only infrequently used parameterisations of models were described below. The model GARCH with threshold GARCH-M effect (introduced by Fiszeder, 2005) can be presented as:

$$y_t = c + \delta^{+} I_t h_t + \delta^{-}(1 - I_t) h_t + \varepsilon_t, \qquad \varepsilon_t | \psi_{t-1} \sim D(0, h_t), \tag{1}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{2}$$

[1] For shorter periods, forecasts from this model were significantly less accurate than forecasts from the GARCH model (they were most often overestimated).

[2] Abbreviations given in the brackets are further used in tables.

where $I_t = 1$ when $x_{t-1} \geq 0$ and $I_t = 0$ when $x_{t-1} < 0$. Returns of the S&P 500 index are used as an exogenous variable x_t.

The moving average estimation period in the moving average model for variance and the value of the smoothing parameter in the exponential smoothing model for variance were chosen for each forecasting period separately to produce the best fit by minimizing the RMSE in the pre-sample. Additionally a moving average equal to 25 and a smoothing parameter equal to 0.94 (value used in RiskMetrics procedure) were also applied.

Fiszeder (2005) suggested estimating parameters of the GARCH model on scaled true range data. True range data for time t is calculated as:

$$TR_t = \max\{(H_t - L_t), |C_{t-1} - H_t|, |C_{t-1} - L_t|\}, \tag{3}$$

where H_t and L_t are the maximum and the minimum price of an asset at time t, C_{t-1} is the closing price at time $t-1$.

Volatility represented by the true range data is higher than volatility represented by squared daily returns, which is why such data should be scaled. Scaling can be applied in two ways, using squared returns or absolute values of returns as follows:

$$STR_t = \frac{a}{b} TR_t, \tag{4}$$

where $a = \sqrt{\frac{1}{t}\sum_{i=1}^{t} r_i^2}$, $b = \sqrt{\frac{1}{t}\sum_{i=1}^{t} TR_i^2}$ or $a = \frac{1}{t}\sum_{i=1}^{t} |r_i|$, $b = \frac{1}{t}\sum_{i=1}^{t} |TR_i|$ depending on the selected scaling method[3].

Scaling for absolute values of returns usually leads to underestimated forecasts, which is why scaling for squared returns is suggested.

Evaluation of forecast accuracy was performed on the basis of the following measures: the relative mean error (RME), the mean absolute error (MAE), the root mean squared error (RMSE), the logarithmic loss function (LL), the heteroskedasticity adjusted mean absolute error (HAMAE), the heteroskedasticity adjusted root mean squared error (HARMSE), the R^2 in a regression of the ex post realized values of variance on its forecast values and the LINEX loss function. The percentage of over-predictions, the estimates of parameters in a regression of the ex post realized values of variance on its forecast values and the estimates of correlations between forecasts of conditional volatility from the simple GARCH model and from the other analysed models were calculated. A test for the equality of the MSEs (see West and Cho, 1995) was also performed.

---

[3] In all formulae it is assumed that mean return is not significantly different from zero.

3. Forecasting Volatility of Returns for the WIG20 Index

Evaluation of accuracy for selected methods of volatility forecasting is performed for the WIG20 index. The sample consists of 1504 daily returns over the period from January 2, 2001 to December 29, 2006[4]. Out-of sample one-day ahead forecasts are constructed for the period from January 2, 2004 to December 29, 2006 (757 sessions). Parameters of the all analysed models are estimated 757 times, extending each time one observation. Hence, the evaluation of forecast accuracy is performed for a relatively long period of time (757 daily observations, approximately 57 thousand 5-minute observations). The continuously compounded rates of return are calculated as $r_t = 100\ln(P_t / P_{t-1})$, where P_t is the closing price of the index at date t. Weak and unstable in time, autocorrelation of order one is present for the whole sample, however insignificant in separate periods of estimation and forecasting. As a measure of ex-post realized volatility the sum of squared 5-minute returns are used.

Parameters of the GARCH models were estimated by the quasi maximum likelihood method. An exception is the GARCH model with conditional Student-t distribution for which the maximum likelihood method was used. The choice of the GARCH orders for p and q (the lag lengths) was based on the minimization of the Schwarz information criterion and results of the test for ARCH. The GARCH(1,1) model has been found to be adequate in this study. Parameters of the stochastic volatility model were estimated by the quasi maximum likelihood approach that relies on a transformation of the model to a state-space form to apply the Kalman filter. Despite its inefficiency, the QML method is consistent and very easy to implement numerically.

In Table 1 the percentage of over-predictions, the values of RME and the values and rankings of all competing models under the MAE, RMSE and logarithmic loss function are reported. The results for the HMAE and HRMSE are omitted because rankings were similar to the results for the MAE and RMSE respectively. For most of the methods forecasts of volatility are more often overpriced but in total underpriced. In Table 2 estimates of parameters for regression of realized volatility on forecasted volatility, estimates of R^2 for this regression and estimates of LINEX loss function are presented.

The ranks of any forecasting model vary depending upon the choice of error statistic. The results suggest that no single model is superior for all evaluation measures. Omitting the LINEX loss function for $a = -1$, which penalizes over-predictions of volatility more heavily than under-predictions, the GARCH model estimated on scaled true range data was the best model. It seems that the moving average model for variance, with the moving average estimation period

[4] Intraday data which are used in the analysis are available for Polish stocks only after the introduction of the WARSET system, which took place in November 17, 2000.

chosen for each forecasting period in the pre-sample, is the second best method. However this model performs poorly for single stocks.

Table 1. Evaluation of forecast accuracy: the percentage of over-predictions, RME, MAE, RMSE and logarithmic loss function

Model	% of over-predictions	RME in %	MAE		RMSE		LL	
			Value	Rank	Value	Rank	Value	Rank
GARCH	62.75	3.90	0.835	11	1.850	12	0.310	9
GARCH-t	62.75	3.31	0.837	13	1.844	10	0.310	9
IGARCH	62.22	2.59	0.849	16	1.844	10	0.314	15
GARCH-M	61.96	4.14	0.835	11	1.855	15	0.310	9
GARCH-MT	62.09	5.54	0.828	7	1.869	19	0.308	7
GJR	62.75	3.53	0.837	13	1.851	13	0.312	13
TGARCH	65.26	2.69	0.855	19	1.882	20	0.323	17
FIGARCH	57.07	8.93	0.828	7	1.867	18	0.306	5
RW	30.78	9.04	1.794	25	3.148	25	7.673	25
HA	78.34	-23.34	1.282	24	2.184	23	0.674	24
MA	52.84	8.58	0.806	1	1.745	2	0.302	3
MA k = 25	51.39	9.49	0.848	15	1.810	4	0.325	18
ES	58.26	6.00	0.831	10	1.816	7	0.312	13
RiskMetrics	53.37	8.93	0.811	2	1.811	5	0.295	2
SV	48.61	19.69	0.825	6	1.933	22	0.331	21
GARCH on TR	66.05	-2.35	0.813	3	1.721	1	0.284	1
GARCH with TR	66.05	-1.73	0.859	20	1.802	3	0.311	12
GARCH with HL	65.65	-1.28	0.854	18	1.812	6	0.309	8
GARCH with 5MR	65.26	-2.29	0.879	22	1.827	8	0.330	20
GARCH with S&P	60.11	6.79	0.821	4	1.851	13	0.303	4
GARCH with seas.	61.43	5.15	0.863	21	1.883	21	0.328	19
GARCH with day.	59.45	6.02	0.853	17	1.865	16	0.337	22
GARCH with hol.	61.56	4.66	0.830	9	1.838	9	0.317	16
S&P in mean	61.03	7.15	0.821	4	1.865	16	0.307	6
RW intra	49.27	0.07	1.054	23	2.320	24	0.433	23

Symbols used in Table are explained in section 2.

The random walk models and the historical average model provide the least accurate forecasts under the most evaluation criteria. The volatility forecasts for the historical average model are significantly overvalued. Therefore the simplest models provide the least accurate forecasts.

The use of additional information in the GARCH model gives mixed results. The results depend not only on the kind of information but also on the way of using information (compare in Tables 1-3 for example GARCH on TR and GARCH with TR or GARCH with S&P and S&P in mean).

The forecasts from most methods are highly correlated with forecasts from the GARCH model (see Table 3). Similarly, the differences between estimates for some error statistics between many models are negligible. For example the hypothesis that MSEs for the GARCH model and for each of the separately analysed models are equal was rejected at the 5% level only for three methods: the random walk model for variance, the historical average model and GARCH models describing seasonal fluctuations (see Table 3).

Table 2. Evaluation of forecast accuracy: determination coefficient, LINEX loss function

Model	γ_0	γ_1	R^2		LINEX a = -1		LINEX a = 1 ($\times 10^9$)	
			Value	Rank	Value	Rank	Value	Rank
GARCH	-0.371	1.253	0.248	12	0.598	6	2.453	14
GARCH-t	-0.330	1.223	0.250	10	0.615	9	2.243	12
IGARCH	-0.125	1.097	0.244	14	0.695	17	1.846	10
GARCH-M	-0.328	1.232	0.242	15	0.599	7	2.727	17
GARCH-MT	-0.368	1.274	0.233	17	0.566	4	3.509	22
GJR	-0.397	1.264	0.246	13	0.599	7	2.497	15
TGARCH	-0.387	1.247	0.218	20	0.620	11	3.118	19
FIGARCH	-0.036	1.120	0.230	19	0.638	15	3.148	20
RW	1.462*	0.212*	0.094	24	7.3e+6	24	15.379	24
HA	4.567*	-1.234*	0.011	25	1.257	23	17.516	25
MA	-0.107	1.159	0.332	2	0.615	9	0.459	1
MA k = 25	0.094	1.048	0.275	4	0.751	18	0.707	4
ES	-0.063	1.101	0.269	6	0.663	16	1.337	8
RiskMetrics	-0.042	1.124	0.277	3	0.625	12	1.525	9
SV	-0.163	1.357	0.208	22	0.506	1	4.576	23
GARCH on TR	-0.319	1.149	0.345	1	1.180	22	0.659	2
GARCH with TR	-0.087	1.030	0.275	4	1.082	21	0.867	5
GARCH with HL	-0.083	1.032	0.268	7	0.931	20	1.078	6
GARCH with 5MR	-0.109	1.036	0.256	9	0.915	19	1.203	7
GARCH with S&P	-0.331	1.269	0.249	11	0.565	3	2.631	16
GARCH with seas.	-0.218	1.181	0.215	21	0.626	13	3.027	18
GARCH with day.	-0.185	1.173	0.232	18	0.628	14	2.044	11
GARCH with hol.	-0.369	1.263	0.259	8	0.575	5	2.407	13
S&P in mean	-0.336	1.277	0.238	16	0.558	2	3.160	21
RW intra	1.088*	0.399*	0.160	23	1.5e+9	25	0.668	3

The Table reports the estimated intercept (γ_0) and slope (γ_1) coefficients from the regression of the sum of square 5-minute returns on forecasted volatility and R^2 from this regression. For γ_0 and γ_1 the asterisk indicates that parameters are significantly different from zero and unity respectively at the 5% level. Newey-West robust standard errors are used. Symbols used in Table are explained in section 2.

On the other side, some differences are significant from the economic point of view. If one treats the additional estimate error as an additional risk, then for example the difference between estimates of RMSE for the GARCH model estimated in a traditional way and for the GARCH model estimated on scaled true range data (0.129) gives approximately 5.8% additional risk (for standard deviation) for one year in the case of the traditional method of estimation.

In most research work related to volatility forecasting properties of considered time series are analysed only very generally. That is why it is impossible to formulate more general conclusions about specific methods of forecasting. In this analysis a number of tests concerning the correctness of considered specifications of GARCH models (most often by testing proper restrictions) were performed[5]. The conclusions from those tests were most often in agreement with conclusions from analysis of the Schwarz criterion (see Table 3).

Table 3. Correlations between forecasts of volatility, results of the test for equality of the MSEs and the Schwarz criterion

Model	ρ	χ^2	SC	SC ranking
GARCH	1	-	5306	10
GARCH-t	0.999	0.945	5291	3
IGARCH	0.998	0.121	5302	9
GARCH-M	0.999	0.984	5309	11
GARCH-MT	0.997	1.083	5299	5
GJR	0.998	1.389	5313	14
TGARCH	0.986	2.522	5319	16
FIGARCH	0.968	2.407	5300	6
RW	0.364	21.394*	-	-
HA	-0.055	6.626*	-	-
MA	0.932	1.640	-	-
MA k = 25	0.958	0.523	-	-
ES	0.981	0.719	-	-
RiskMetrics	0.984	1.520	5298	4
SV	0.933	2.458	5267	2
GARCH on TR	0.890	2.183	-	-
GARCH with TR	0.967	0.771	5302	8
GARCH with HL	0.972	0.777	5301	7
GARCH with 5MR	0.988	0.321	5310	12
GARCH with S&P	0.999	0.150	5312	13
GARCH with seas.	0.955	3.577	5374	18
GARCH with day.	0.967	7.858*	5327	17
GARCH with hol.	0.987	2.907	5316	15
S&P in mean	0.998	1.280	5260	1
RW intra	0.526	2.510	-	-

The column ρ presents the estimates of correlations between forecasts of volatility from the GARCH model and from the other analysed models. The χ^2 column presents the estimates of statistic for the hypothesis that MSEs for the GARCH model and each of the separately analysed models are equal. The asterisk indicates that the hypothesis is rejected at the 5% level. Symbols used in Table are explained in section 2.

The results from the performed tests and analysis of the Schwarz criterion are often not in agreement with the accuracy of the models in volatility forecasting. In construction of information criteria and often in restriction testing the whole model is considered (for example equations for mean and for variance, conditional density). Taking into account in modelling specific properties of stock returns does not always increase the accuracy of volatility forecasts based on those models. It appears that some characteristics are important in the evaluation of the general fit of the model in the sample, but their influence is not significant on the accuracy of forecast.

[5] The results are omitted owing to the limited size of this paper.

4. Conclusions

This study evaluates the performance of seventeen parameterisations of GARCH models and eight other methods for predicting volatility of the Polish stock index WIG20. The results of research have shown that the use of additional information in construction of a GARCH model or in the estimation of its parameters does not always increase the accuracy of volatility forecasts. However, it is possible to increase significantly from the economic point of view[6] the accuracy of forecasts constructed based on the GARCH model estimated on scaled true range data.

References

Doman, M., Doman, R. (2004), *Ekonometryczne modelowanie dynamiki polskiego rynku finansowego*, AE w Poznaniu, Poznań.

Fiszeder, P. (2004a), *Forecasting Volatility with GARCH Models*, materiały konferencji MACROMODELS'2003, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Fiszeder, P. (2004b), Prognozowanie zmienności na podstawie modeli GARCH, *Rynek Terminowy*, 25, 121–128.

Fiszeder, P. (2005), Forecasting the Volatility of the Polish Stock Index – WIG20, *Forecasting Financial Markets. Theory and Applications*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Fiszeder, P., Romański, J., (2002), Looking for the Pattern of GARCH Type Models in Polish Stock Returns. Comparison with Indices of the EU and the East European Stock Markets, *East European Transition and EU Enlargement, A Quantitative Approach*, Physica-Verlag, A Springer-Verlag Company, Heidelberg.

Osiewalski, J., Pajor, A., Pipień, M., (2004), *Bayesowskie modelowanie i prognozowanie indeksu WIG z wykorzystaniem procesów GARCH i SV*, XX Seminarium Ekonometryczne im. Profesora Zbigniewa Pawłowskiego, AE w Krakowie, Kraków.

Piontek, K. (2003), Weryfikacja wybranych technik prognozowania zmienności – Analiza szeregów czasowych*, Inwestycje finansowe i ubezpieczenia - tendencje światowe a polski rynek*, Prace naukowe AE we Wrocławiu, nr 991.

Pipień, M. (2006), *Wnioskowanie bayesowskie w ekonometrii finansowej*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Poon, S-H., Granger, C. (2003), Forecasting Volatility in Financial Markets: *A Review, Journal of Economic Literature*, 41, 478–539.

West, K. D., Cho, D. (1995), The Predictive Ability of Several Models of Exchange Rate Volatility, *Journal of Econometrics*, 69, 367–391.

[6] Although insignificantly from a statistical point of view according to RMSE.