International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:5, No:11, 2011

VaR Forecasting in Times of Increased Volatility

Ivo Jánský, Milan Rippel

***Abstract*—**The paper evaluates several hundred one-day-ahead VaR forecasting models in the time period between the years 2004 and 2009 on data from six world stock indices - DJI, GSPC, IXIC, FTSE, GDAXI and N225. The models model mean using the ARMA processes with up to two lags and variance with one of GARCH, EGARCH or TARCH processes with up to two lags. The models are estimated on the data from the in-sample period and their forecasting accuracy is evaluated on the out-of-sample data, which are more volatile. The main aim of the paper is to test whether a model estimated on data with lower volatility can be used in periods with higher volatility. The evaluation is based on the conditional coverage test and is performed on each stock index separately. The primary result of the paper is that the volatility is best modelled using a GARCH process and that an ARMA process pattern cannot be found in analyzed time series.

***Keywords*—**VaR, risk analysis, conditional volatility, garch, egarch, tarch, moving average process, autoregressive process

I. Introduction

THE objective of the paper is to analyze VaR forecasting methods based on several conditional mean and conditional variance modeling processes subject to a sharp increase in volatility, such as during the recent financial crisis. The data for the analysis come from six world stock indices. Moreover, the methods are to be evaluated without any prior assumptions on the particular parameters of the conditional mean and conditional variance processes as well as the shape of the distribution of the log returns. Accuracy in the sense used in this paper is represented by the ability to provide results that closely follow the actual market development.

The paper is structured into five parts. First, literature overview briefly discusses several important papers that are concerned with the problematic of forecasting VaR. Then a model specification and methodology are discussed, which leads to the actual application of the selected VaR methods on the data from several market indices and their evaluation. The fifth part summarizes the obtained results. Finally, a short conclusion is provided.

II. Literature Overview

The original inspiration for this paper stemmed from the work of Angelidis, Benos & Deggianakis [2] who performed a similar analysis of several stock based indices. In their paper, the authors analyzed five European stock indices and estimated several GARCH based models on the data for the time span 1987 – 2002. This paper focused on analyzing the data that contain a relatively stable period in terms of volatility in the in-sample part and quite a turbulent volatility in the out-of-sample subset. The reason for such deviation was to perform analysis that attempts to provide answers on how the models behave in times of unexpectedly increased volatility.

Similarly to [2] our paper uses the so called conditional coverage framework developed by Christoffersen [6], in order to test the performance of the estimated models. The advantage of using the conditional coverage framework lies in the fact that the conditional coverage test is not only capable of evaluation the failure ration of a given model, moreover it is able to rule out such models that cluster violations. In other words, the conditional coverage framework ensures that the violations are randomly distributed over the critical period.

Another quite influential study for this paper was the paper written by Costello, Asem & Gardner [7], who evaluated several ARMA-GARCH based models on the data for Brent Crude Oil between the years 1987 and 2005. The authors took a similar approach in the evaluation of GARCH based models with the addition that their GARCH models were parameterized.

III. Model Specification and Methodology

For modeling data series we used two common concepts of conditional mean- the AR process and the MA process. The AR process is described by (1):

$$y_t = \mu + \sum_{i=1}^{p} \rho_i * y_{t-i} + \varepsilon_t, \quad (1)$$

where p is the lag parameter of the observed variable, y_t is the random observed variable at time t depending on the previously realized values of y_{t-i}, μ is the mean constant and ε_t is the white noise.

The MA process is described by (2):

$$y_t = \mu + \sum_{i=1}^{q} \theta_i * \varepsilon_{t-i} + \varepsilon_t, \quad (2)$$

where q is the number of lags of the error term, y_t is the random observed variable depending on the previously realized values of error term ε_{t-i}, , θ_i is the parameter,μ is the mean constant, ε_t is the white noise.

The combination of both gives us the ARMA process described by (3):

$$y_t = \mu + \sum_{i=1}^{p} \rho_i * y_{t-i} + \sum_{i=1}^{q} \theta_i * \epsilon_{t-1} + \epsilon \quad (3)$$

As the financial data time series shows heteroskedasticity (see [1]) a model dealing with conditional heteroskedasticity must be used. We use the GARCH model introduced in [4], which is a generalization of the ARCH model that was

Ivo Jánský is a PhD student at Charles University in Prague, Czech Republic (e-mail: ivo.jansky@gmail.com).
Milan Rippel is a PhD student at Charles University in Prague, Czech Republic (e-mail: milanrippel@seznam.cz).

originally developed in [9]. The ARCH model allows for long lags in conditional variance and the GARCH model extends it in the way that it allows for both long lags in conditional variance and a more flexible lag structure. The definition of the GARCH(p,q) model is described by (4):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i * \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i * \sigma_{t-i}^2, \qquad (4)$$

where p is the order of the GARCH terms σ_t^2 and q is the order of the ARCH terms ε_t^2.

GARCH model is able to deal with common financial data time series characteristics such as thick tails and volatility clustering, as pointed out in [11] and [12]. There are, however, some characteristics of financial time series that the GARCH model is not able to deal with. The main disadvantage of the GARCH model is that conditional variance depends on the squared value of ε_t, which in turn means that the model is sensitive only to the absolute magnitude of the variable but not to its sign leading to a presence of a leverage effect (see [3]), which represents a negative correlation between asset returns and volatility of returns.

Some authors such as Brooks & Persand in [5] and Rabemananjara & Zakoïan in [15] therefore suggest using models that take into consideration asymmetries in volatility of returns. Two most popular asymmetric models are the EGARCH model of Nelson in [13] and the TARCH model created by Zakoïan in [16].

In order to test for serial autocorrelation of residuals and squared residuals, the Portmanteau Q test is employed in the form represented by the Ljung-Box implementation described by (5):

$$Q = n * (n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}, \qquad (5)$$

where n is the number of observations, h is the number of lags under test, k is the current lag and $\hat{\rho}_k^2$ is the autocorrelation at current lag. The null hypothesis of the Ljung-Box test states that the data in the time series is independently distributed. Moreover, the test is performed at various lags, which means that it tests the overall randomness of the time series under observation. With the help of the Portmanteau Q test, we are able to find out whether or not the ARMA-GARCH model is appropriate for a particular time series and is able to capture the serial autocorrelation.

The tested models are exclusively combinations of the logarithm form of ARMA process with up to two lags for the conditional mean, as described by (3), and one of the GARCH, TARCH and EGARCH processes with up to two lags for the conditional volatility, as denoted in (4). This gives a total of 648 models that have been estimated and evaluated on the datasets. The stationarity condition of log returns is achieved by differencing logarithmic prices in the time series. The distribution function that models residuals (error term) represents another aspect in the definition of the models. To perform a more thorough evaluation, the normal distribution function, the Student-t distribution function, and the GED were chosen as possible distributions for the error term.

The analysis is performed on six stock indices DJI, GSPC, IXIC, FTSE, GDAXI and N225. Datasets have been selected with the requirement to grasp indices from different parts of the world. In order to test the forecasting accuracy of selected VaR models, the data are divided into two groups. Each dataset includes exactly 1500 observations from which the first 1000 represent the in-sample subset and the remaining 500 represent the out-of-sample subset. Observations in all datasets are centered on the 31st December 2007. The reason is to have an artificial point for the division of each dataset into in-sample and out-of-sample subsets due to the fact that the stock indices come from several countries and trading dates are not internationally standardized.

The models are evaluated on the in-sample subset and the top performing in-sample models based on the AIC value are selected and discussed in the analysis for each index. One of the models with the best AIC value is then described in more details at the end of the paper. The models are then applied to the out-of-sample data, where their forecasting capability is tested. Some papers such as [14] discourage from evaluating models using information criteria[1], since a superior in-sample performance of a particular model does not necessarily mean a superior out-of-sample performance. The same argument applies to rankings based on the log likelihood function. For this reasons, the evaluation procedure is based on the out-of-sample performance measured by the p-values of both the unconditional and conditional coverage tests, since they are able to capture both the violations rate and the independence of violations. In situations when the p-values cannot be used, the values of loss functions according to the framework presented in [8] are employed.

IV. Data Analysis

The hypothesis that in-sample log returns are normally distributed is tested using the Jarque-Bera test statistic and the results confirm that the null hypothesis of normality is rejected at the significance level of 5% for each index, except for the IXIC index, where the null hypothesis cannot be rejected at 1% significance level. From the volatility plots it is obvious that the time series experienced a sharp increase in their volatility around the year 2008-2009.

The out-of-sample subset has quite different characteristics than the in-sample subset. P-values of the Jarque-Bera test statistic reject the null hypothesis of normality at all significance levels for each index and the variance of the out-of-sample is greater. The increase amounts to approximately 5 to 8 times the variance of the in-sample subset, depending on the particular index. The out-of-sample subset is therefore quite turbulent, which is very suitable for the purpose of the analysis.

The autocorrelation and partial autocorrelation functions of residuals suggest that there might be AR and MA processes in the subset for each index. As well as in the in-sample subset, the actual order of AR and MA processes varies among

[1]Bayesian Information Criterion or Akaike Information Criterion

indices and it seems that AR and MA processes are of a high order of up to ten significant lags. As well as in the in-sample subset, the GARCH family of processes seems to be present in the out-of-sample subset for all indices except for the GDAXI index. The highest order GARCH process seems to be present in the FTSE and the N225 indices.

Each in-sample subset has been tested for stationarity by the Augmented Dickey-Fuller test for the presence of a unit root and the null hypothesis of a unit root up to the fifth lag has been rejected at all significance levels for each index. Stationarity of the out-of-sample subset is tested using the Augmented Dickey-Fuller test and the null hypothesis of a unit root up to the fifth lag has been rejected at all significance levels for each out-of-sample subset.

From the analysis of the in-sample and the out-of-sample subsets it is clear that the data have quite different properties. Since the models are estimated on the in-sample data and then applied on the out-of-sample data, it is interesting to observe, whether or not the models are able to use the past realized log returns at time t in order to provide an appropriate VaR forecast for time $t + 1$.

V. Models Estimation and Evaluation

The following section contains a comparison of performance of models for in-sample and out-of-sample datasets for each of the selected indices. The overall results are then commented in the following chapter, the overview of the best-performing models is given in Table I and Table II.

According to the Jarque–Bera statistic, the log returns of all indices are not normally distributed for the in-sample subsets. Therefore when the estimated models are ranked by their AIC and LL values, the top ranking models are the ones that assume either the Student-t or the GED for the distribution of the residuals. The estimated degrees of freedom parameter of the Student-t distribution for residuals lies between 7 and 8 (16 to 21 for the IXIC and FTSE indices) and the shape parameter for the GED lies between 1.4 and 1.5 (1.8 for the FTSE index).

For each estimated model there were 500 forecasts calculated in the first run. The models were not periodically re-estimated, since the computational requirements accounted to 50 hours for the base estimation. Therefore periodical re-estimation after 125 days would require almost one month of continuous estimation using the statistical software. Due to such high demand on computation resources, the re-estimation will be subject to future research.

A. DJI (Dow Jones Industrial Average)

The top performing in–sample model is AR(2)-MA(1)-TARCH(2,2)-GED. To confirm that the best in–sample model based on the conditional volatility process TARCH(2,2) has a positive effect on the serial autocorrelation of squared residuals, the Portmanteau Q test has been applied on the residuals. The null hypothesis of the test cannot be rejected at 6.54% significance level in all cases. The top ranking in–sample model is dominated by the order of the particular TARCH process and the orders of the AR and MA processes seem to play a minor role. In other words, the most dominant factor is the conditional volatility.

To test the accuracy of the one-day-ahead VaR forecast, the estimated models are applied to the out-of-sample data. Since the out-of-sample data exhibit higher volatility, it is not surprising that even though the TARCH model is selected during the in-sample estimation, it exhibits very poor results in the out-of-sample evaluation. The most accurate VaR forecasts are achieved with the Student-t distribution with 8 degrees of freedom. The top eight out-of-sample models indicate that for the modeling of the conditional volatility the GARCH(2,2) process is optimal at the confidence level $\alpha = 0.95$. On the other hand the conditional mean process does not seem to follow any particular pattern. All the models underestimate the VaR. The failure rate, ranging from 7.0% to 7.4%, is more than the expected rate of 5%. The result of such underestimation is projected to the low p-value of the unconditional coverage. To select a single model with the best performance, the realized values of the loss functions are compared. The best performing VaR model for the DJI index at confidence interval $\alpha = 0.95$ is the AR(1)-MA(0)-GARCH(2,2)-T model.

The results prove that in the case of the DJI index it is not optimal to calculate one-day-ahead VaR using models that attain the highest values of either AIC or LL. Even though these models might perform relatively well on the in-sample subset, they do not provide adequate results when applied to the out-of-sample subset. Conditional volatility processes such as EGARCH and TARCH seem to be adequate in quite stable times in terms of volatility. On the other hand, when a period with higher volatility is expected, the results suggest using a simple GARCH model, since it outperforms the other models in terms of the conditional coverage test, which captures both accuracy of the VaR forecast and the independence of the realized violations of the predicted VaR.

B. GSPC index (Standard and Poor's 500)

Due to the serial autocorrelation of the data, the best in-sample models should include a conditional volatility modeling process. This is indeed the case, as all the best in-sample models use the EGARCH(2,2) process in order to model conditional volatility. The top in-sample model is AR(1)-MA(1)-EGARCH(2,2)-T. The best in-sample VaR models are dominated by the conditional volatility part. The conditional mean is modeled with multiple combinations of AR and MA processes with orders ranging from 0 to 2. Therefore, it is obvious that volatility plays the most important role in the prediction of the one-day-ahead VaR.

The application of the estimated models on the out-of-sample data is a good stress test, since the out-of-sample subset is more volatile than the in-sample subset. None of the top eight in-sample models belongs within the best performing models for the out-of-sample period. The best results are achieved with the Student-t distribution and the estimated number of degrees of freedom is approximately 7. The conditional variance in these models is modeled using the GARCH process, as opposed to the in-sample EGARCH

process. The top eight out-of-sample models underestimate the VaR, as the failure rates lay between 7.2% and 8.2%. The violations of the log returns are followed one by another, which worsens the value of the conditional coverage. According to the p-values and the values of the loss functions, the best performing model for the GSPC index at $\alpha = 0.95$ confidence interval is the AR(1)-MA(0)-GARCH(2,1)-T model.

As well as for the DJI index, the most accurate one-day-ahead VaR forecasts are not achieved with in-sample models that prove to be the best performing ones based on AIC and LL values. The top in-sample models model conditional volatility exclusively with the EGARCH process; however, the out-of-sample data are quite different from the in-sample data and therefore the conditional volatility process is not the same, GARCH is performing better than EGARCH.

C. IXIC index (NASDAQ Composite)

The serial autocorrelation hypothesis is rejected only for the 2nd lag. Therefore it can be expected that the best performing in-sample models should be the ones that model conditional variance with a process of at least two orders. The expectation is confirmed as all of the top eight in-sample models employ an EGARCH process with two lags for the autocorrelation term as the conditional volatility process. The best in-sample model is AR(0)-MA(2)-EGARCH(2,2)-T.

Since the out-of-sample subset has greater volatility and kurtosis, the application of the models on the out-of-sample subset might change the order of the best performing models. Among the successful models at $\alpha = 0.95$ confidence interval there are models with two types of the conditional volatility process. There are six models that employ the GARCH(1,1) process and there are two models that take advantage of the GARCH(2,2) process. All of the models use the Student-t distribution for the error term. Unfortunately, the results of models applications on the out-of-sample subset do not provide satisfactory results. The failure rates range from 7.8% to 8.0%, which is significantly higher than the expected 5% failure rate. The four models with the highest values of the conditional coverage are then compared based on loss functions values. Based on this criteria, the AR(1)-MA(0)-GARCH(1,1)-T model is selected as a model with the best performance at the confidence interval of $\alpha = 0.95$. It is important to note, however, that based on the values of the conditional coverage, none of the models proves to be an adequate one.

The outcome of the analysis is quite similar to the outcomes of the previous two indices. Even though the EGARCH process performs very well for the in-sample subset, which exhibits lower volatility, the one-day-ahead VaR forecasts based on the out-of-sample subset provide different results. The best accuracy is achieved when the conditional volatility process is modeled using the GARCH process. Thus also the results for the IXIC index confirm that the best in-sample models based on the AIC and LL values are not the suggested models for periods with higher volatility.

D. FTSE index (Financial Times Stock Exchange 100)

The top eight in-sample models are mostly the ones using the TARCH process for the conditional volatility process, with one exception that is represented by the EGARCH process. The EGARCH(1,1) process is successfully able to remove the serial autocorrelation, since the Portmanteau Q test indicates that the serial autocorrelation is effectively captured. Seven out of the eight models work under the assumption of the Student-t distribution for the error term and one is taking advantage of the GED distribution. The best in-sample model is AR(2)-MA(2)-EGARCH(1,1)-T.

Although the out-of-sample subset of the FTSE index exhibits a value of skewness closer to zero, the kurtosis is higher. The value of the Jarque-Bera test statistic is even higher than in the in-sample subset, which means that the out-of-sample subset is not normally distributed. It is therefore expected that the order of the best performing models might be quite different from order of the in-sample models. All top performing models assume the Student-t distribution for the error term and the degrees of freedom parameter is estimated to be equal to 12. Looking at the failure rates, it is obvious that the models are not able to predict the one-day-ahead VaR very accurately. The failure rates range from 7.8% to 8.0%, which is quite large. When the models are further on compared based on the values of their loss functions, there is one model with the lowest achieved value in two of the three selected loss functions. Therefore, the AR(0)-MA(2)-GARCH(2,1)-T model seems to be the most adequate one for the FTSE index at $\alpha = 0.95$ confidence interval.

The application of the top AIC and LL based models on the FTSE index out-of-sample data is not an optimal choice. None of the in-sample models is able to accurately forecast the one-day-ahead VaR. Interesting observation is that even for the FTSE index it is optimal to employ models that take advantage of a GARCH conditional volatility process; as such models provide the best forecasts. It is important to note, however, that even those models do not provide acceptable forecasts.

E. GDAXI Index (Deutscher Aktien IndeX)

After the application of the conditional volatility process TARCH(2,2) the serial autocorrelation is effectively captured. The top performing model for in-sample data is AR(2)-MA(2)-TARCH(2,2)-GED. Interesting observation is that the model works exclusively with the GED distribution for the error term. The shape parameter of the distribution is estimated to be approximately 1.5. As suggested by the Portmanteau Q test, the order of the autoregressive part of the TARCH process is equal to 2.

Both the skewness and the kurtosis of the out-of-sample log returns differ significantly from the in-sample subset and so does the volatility. Based on the results of the evaluation of the previous indices, it is expected that the top in-sample model might not score among the top out-of-sample models, simply due to the changed properties of the subset. The results of the models application at $\alpha = 0.95$ confidence interval are again not satisfactory. The selected models employ the

Student-t distribution for the error term with 8 degrees of freedom and the conditional volatility process is modeled by a GARCH processes with multiple orders. None of the EGARCH or TARCH processes appears among the top eight results for this confidence interval. The one-day-ahead VaR forecasts are violated approximately in 8.0%. The values of the loss functions select the AR(1)-MA(2)-GARCH(1,1)-T model as the one with the lowest realized loss and highest p-value of the conditional coverage test for the GDAXI index at the $\alpha = 0.95$ confidence interval.

Although the in-sample estimation suggests using models that take advantage of the TARCH process with GED as the distribution for the error term, the application of these models on the out-of-sample subset provides quite different results. The best relative accuracy is achieved with models that model conditional volatility with the GARCH process and the Student-t distribution. Unfortunately, the evaluation of the models at $\alpha = 0.95$ confidence interval does not provide satisfactory results, since all of the models are rejected by both the unconditional and conditional coverage tests. Therefore it cannot be confirmed that there is a one particular model that is able to capture the behavior of the GDAXI index at the best.

F. N225 index (Nikkei 225)

Due to the serial autocorrelation the conditional volatility should be modeled using one of the available conditional volatility processes. After the application of the conditional volatility process EGARCH, the serial autocorrelation is effectively captured. Even though the conditional volatility process, EGARCH, is the same for all top in-sample models, its parameters differ and so does the assumed distribution of the error term. The best performing model for the in-sample subset is AR(0)-MA(1)-EGARCH(2,1)-T.

The skewness of out-of-sample subset is very similar to the skewness of the in-sample subset. On the other hand, the kurtosis is significantly higher. Moreover, volatility is more turbulent. Therefore the top performing in-sample models might most likely fail to capture the change in the parameters of the subset. The top eight out-of-sample models at $\alpha = 0.95$ confidence interval employ the Student-t distribution with estimated 10 degrees of freedom for the error term. The failure rates for the one-day-ahead VaR at $\alpha = 0.95$ confidence interval are the lowest failure rates from the selected indices, as they range from 6.2% to 6.6%. The values of the conditional coverage are also quite satisfactory. Based on these values, the best performing model is the AR(2)-MA(1)-GARCH(1,1)-T model. This model has also the highest realized p-value of the conditional coverage of all tested models and all indices.

Interesting observation occurs in $\alpha = 0.95$ confidence interval, where the closest match to the expected failure rate of 5% is achieved among all indices. The EGARCH process for conditional volatility is suggested by the in-sample estimation procedure. However, the application of the estimated models on the out-of-sample subset indicates that there are other models that are able to achieve better VaR forecasting accuracy. As well as for the other indices, the best accuracy is achieved when the conditional volatility is modeled using the GARCH process.

VI. Results

The top in-sample models (see Table I), ranked by AIC, were mostly EGARCH or TARCH based with either Student-t or the GED as the most appropriate distribution for the error term. These estimates proved that models that work with conditional volatility modeled as an asymmetric process provided more accurate and significant estimates than the typically used models with a symmetric GARCH process. The interpretation of this outcome is that the markets treat positive shocks in a different manner than negative shocks. Considering the fact that the in-sample subsets were less volatile than the out-of-sample subsets, it seems that in periods with relatively stable volatility, asymmetric conditional volatility processes provide better estimates than the symmetric conditional volatility processes. On the other hand, the orders of the AR and MA processes did not seem to be of a particular importance, as their orders quite varied and seemed to play a minor role in the actual specification of the models. The implication of this result is that the markets treat the volatility of a particular stock index as a dominant factor, rather than the previously realized value of the stock index.

TABLE I
Top performing in-sample models

Index	Model	LL	AIC
DJI	AR(2)-MA(1)-TARCH(2,2)-GED	3598.18	-7172.35
GSPC	AR(1)-MA(1)-EGARCH(2,2)-T	3562.49	-7102.98
IXIC	AR(0)-MA(2)-EGARCH(2,2)-T	3273.93	-6525.86
FTSE	AR(2)-MA(2)-EGARCH(1,1)-T	3634.53	-7249.06
GDAXI	AR(2)-MA(2)-TARCH(2,2)-GED	3329.38	-6632.77
N255	AR(0)-MA(1)-EGARCH(2,1)-T	17898.67	-35779.33

The models were then compared based on the values of the conditional coverage test. Unlike most works with a similar topic, the paper applied a less known framework for the evaluation of a one-day-ahead VaR - the conditional coverage framework. The advantage of applying the conditional coverage lies in the fact that it does not only test whether the number of extreme cases corresponds to the selected confidence interval, it also tests whether the violations are clustered or not.

TABLE II
Top performing out-of-sample models

Index	Model	LL	AIC	Fails	p-val[2]
DJI	AR(1)-MA(0)-GARCH(2,2)-T	3564.40	-7112.80	7.0 %	8.%
GSPC	AR(1)-MA(0)-GARCH(2,1)-T	3531.67	-7049.35	7.2 %	0.7%
IXIC	AR(1)-MA(0)-GARCH(1,1)-T	3250.67	-6489.35	7.8 %	0.1%
FTSE	AR(0)-MA(2)-GARCH(2,1)-T	3526.51	-7037.03	7.8 %	2.8%
GDAXI	AR(1)-MA(2)-GARCH(1,1)-T	3300.20	-6584.40	8.0 %	0.5%
N255	AR(2)-MA(1)-GARCH(1,1)-T	3129.20	-6242.41	6.2 %	37.7%

[2]P-value of Conditional Coverage test.

The evaluation of the out-of sample models was performed for the 95% confidence interval and the results were quite different from the in-sample results (see Table II). All of the top scoring models were based on modeling the conditional volatility using the GARCH process. Even though most of the models underestimated the VaR by some amount the majority of the models were not rejected based on the p-value of the conditional coverage test.

It is quite interesting that the GARCH process offers the best performance in all indices. The likely reason lies in the fact that the GARCH process treats both positive and negative shocks in the same way and that the situation on the markets during the out-of-sample period probably exhibited this exact behavior.

VII. Conclusion

The VaR forecasting theory has been applied on six selected stock indices (DJI, GSPC, IXIC, FTSE, GDAXI, N225) and the data for the one-day-ahead VaR forecast came from the years 2004 - 2007 for the in-sample subset and the years 2008 - 2009 for the out-of-sample subset, which served both as a source of data with higher volatility and as a benchmark. Even though it is quite common in the literature to work with the assumption of normally distributed log returns, the paper did not adhered to such simplification and tested the log returns for a variety of distributions including the popular Student-t and the GED distributions.

One of the most significant aspects of the paper was the variability of employed models, as the paper attempted to estimate 648 dynamic models. The top in-sample models, ranked by AIC, were mostly EGARCH or TARCH based with either Student-t or the GED as the most appropriate distribution for the error term. On the other hand, the top out-of-sample models were in all cases the GARCH models with Student-t distribution for the error term.

Many authors suggest using the GARCH(1,1) process for modeling the conditional volatility of stock indices. Although, the paper did not work with such prior assumption and tested a vast number of models, the suggestion of using GARCH(1,1) process for the conditional volatility process while forecasting VaR could not be rejected. Each one of the suggested out-of-sample models took advantage of the GARCH process. Even though the asymmetric models achieved the best results in terms of the AIC and LL values, the actual forecast capabilities were dominated by a symmetric conditional volatility process. It also confirms the hypothesis that an asymmetric EGARCH is outperformed by a GARCH based model when the magnitude of shocks, volatility, is high.

Therefore the suggestion to forecast VaR only with models that have the highest value of the AIC was not backed by the results of the paper. None of the models with the highest in-sample AIC values was among the top out-of-sample models. The results obtained in the paper suggest taking advantage of the commonly used GARCH process, which saves a lot of computation time and provides satisfactory results. In order to improve the VaR forecasts, the most obvious step would be to re-estimate the models on a daily basis and to increase the possible lags of the conditional volatility process. The downside of such approach is, however, a significantly higher demand for computational resources, as it takes quite some time to estimate the models. This extension of the paper is left for the future research.

When compared to the results presented in [2], this paper found is in line with the findings of the mentioned authors, as the mean process represented by an ARMA process has not proved to improve the predictive accuracy of the models. The suggestion is, therefore, to exclude the ARMA process from the analysis, as it adds more complexity than improvement. On the other hand, the results of the volatility process analysis partially differ from the results of [2]. This paper identified the GARCH process as the best conditional volatility process for the analyzed time series; however, the mentioned authors identified both GARCH and EGARCH as equally valid for the conditional volatility analysis. Both the authors and this paper identified the Student-t distribution as the most appropriate distribution for the analyzed time series.

To conclude, the paper loosely followed a typical one-day-ahead VaR evaluation procedure with a number of improvements introduced into the procedure. Such improvements included a no prior assumption on the distribution of the log returns, which proved to be a step in the right direction. Further on, the paper estimated a quite large number of models that allowed comparing the models with various conditional mean and conditional volatility processes, as well as with three distribution functions for the error term. The final part of the evaluation took advantage of a less known framework that is used to measure the accuracy of the forecasted models. Thanks to this, the paper was able to provide a new insight on the topic, which certainly belongs to the most discussed topics in the financial sector.

Acknowledgment

Financial support for this research from:

i) The Czech Science Foundation, project The Institutional Responses to Financial Market Failures, under No. GA P403/10/1235;
ii) The IES Institutional Research Framework 2005-2010 under No. MSM0021620841; and
iii) The Charles University Grant Agency, project The Importance of Risk Management and Financial Stability During Financial Crisis under No. GAUK 31610/2010

is gratefully acknowledged.

References

[1] Akgiray, V. (1989): "Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. The Journal of Business, 62, 55-80.
[2] Angelidis, T., Benos, A., & Deggianakis, S. (2003). The Use of GARCH Models in VaR Estimation.
[3] Black, F. (1976). Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section , 177-181.
[4] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327.

[5] Brooks, C., & Persand, G. (2003). The effect of asymmetries on stock index return Value-at-Risk estimates. The Journal of Risk Finance, 4, 29-42.
[6] Christoffersen, P. (1998). Evaluation interval forecasts. International Economic Review, 39, 841-862.
[7] Costello, A., Asem, E., & Gardner, E. (2008). Comparison of historically simulated VaR: Evidence from oil prices. Energy economics, 30, 2154-1266.
[8] Diebold, F., & Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business and Economic Statistics, 20 (1), 134-144.
[9] Engle, R. F. (1982). Autogregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50, 987-1008.
[10] Jánský, I. (2011). Value-at-risk forecasting with the ARMA-GARCH family of models during the recent financial crisis. Prague: Institute of Economic Studies.
[11] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. The Journal of Business, 36, 394-419.
[12] Mandelbrot, B. (1967). The Variation of Some Other Speculative Prices. The Journal of Business, 40, 393-413.
[13] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347-370.
[14] Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. Journal of Econometrics, 45 (1-2), 267-290.
[15] Rabemananjara, R., & Zakoïan, J. M. (1993). Threshold ARCH models and asymmetries in volatility. Journal of Applied Econometrics, 8, 31-49.
[16] Zakoïan, J. M. (1994). Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18, 931 - 955.