BACKTESTING VAR ESTIMATION UNDER GARCH AND GJR-GARCH MODELS

Aleš Kresta

Abstract

The important and no less interesting part of financial risk management is the risk modelling. Commonly utilized measure of risk (not only by banks and insurance companies) is Value at Risk. Since the financial time series are typical by non-constant volatility over time, it is crucial for Value at Risk calculation to model the standard deviation of returns correctly. In the paper we assume (relatively simple) models based on GARCH and GJR-GARCH models with Student distributions of innovations. These models are back-tested assuming the investment into Prague stock market index. The period utilized for back-testing is from 1993 till 2012, i.e. 4,627 daily values. The evaluation is made by means of the detected number of exceptions, i.e. the cases in which the observed losses were bigger than estimated Value at Risk on a given probability level. Also well-known statistical tests due to Kupiec and Christoffersen are utilized. According to results the assumed models are not accurate – the risk is underestimated, but bunching of the exceptions is not present.

Key words: back-testing, Value at Risk, risk management

JEL Code: G17, G32

Introduction

Value at Risk (henceforth VaR) is generally accepted as a measure of risks which the financial institutions are exposed to. Simply speaking, the Value at Risk represents maximum loss observed with a given confidence level, i.e. there will be loss bigger than VaR only with the probability level of one minus confidence level (for banks and insurance companies the confidence levels are 99% and 99.5%).

For Value at Risk estimation we can generally recognize three groups of methods: variance-covariance method, (filtered) historical simulation and Monte Carlo simulation. While it is crucial for all the methods to estimate the future volatility correctly, we can distinguish also the methods which assume the volatility to be constant over time and methods modelling variance over time. Both type models were tested for example by Alexander and Sheedy (2008).

Majority of the recent papers published on risk backtesting topic is focused on the dependency modelling (i.e. the authors assume the portfolio composed of more than one asset). We can mention for instance papers published by Huang, Lee, Liang, and Lin (2009), Ignatieva and Platen (2010) or Kresta and Tichý (2012). In these papers the accuracy of models was assessed, but there were not given much attention to the length of period utilized for parameters estimation.

In the paper we examine the effect of the chosen period size utilized for parameters estimation on the backtesting results. For the computational simplicity we focus on one asset portfolio, i.e. only one risk factor is assumed and thus the dependence modelling problems are avoided. In the paper we assume (relatively simple) autoregressive models with heteroscedasticity modelled by GARCH and GJR models with Student innovations.

The goal of the paper is to backtest these volatility models for different estimation period sizes. The backtesting is performed for the investment into the Prague stock market index over the period 1993-2012.

The paper is organized as follows. Applied volatility models are defined in the next section. Then, Value at Risk and method of its backtesting are described. In the last section, utilized dataset is described and backtesting results are presented.

1 Volatility models

Volatility models have become important tool in time series analysis, particularly in financial applications. Engle (1982) observed that, although the future value of many financial time series is unpredictable, there is a clustering in volatility. He proposed autoregressive conditional heteroskedasticity (ARCH) process, which has been later expanded to generalized autoregressive conditional heteroskedasticity (GARCH) model by Bollerslev (1986). Further extension assumed in this paper is asymmetric GJR model proposed by Glosten, Jagannathan, and Runkle (1993). There were also proposed other volatility models such as IGARCH, FIGARCH, GARCH-M, EGARCH, etc. However these are not subject of study in this paper.

For both models the conditional mean will also be assumed. Thus the models of time series $\{x_t\}_{t=1}^{N}$ will be of general form as follows,

$$x_t = \mu_0 + \sum_{i=1}^{R} \mu_i \cdot x_{t-i} + \sigma_t \cdot \widetilde{\varepsilon}_t, \tag{1}$$

$$\tilde{\varepsilon}_t \sim t_\nu(0,1), \tag{2}$$

where μ_0 is unconditional mean of the series, μ_i are autocorrelation coefficients for lag 1 up to R, σ_t is modelled standard deviation (volatility) and $\tilde{\varepsilon}_t$ is a random number from Student probability distribution (henceforth t).

1.1 GARCH model

The GARCH model was proposed as the extension of ARCH model in order to avoid problematic parameters estimation, when there are many of them. The model takes the following form,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{P} \alpha_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \cdot \varepsilon_{t-j}^2 , \tag{3}$$

where α_0, α_i and β_j are parameters needed to be estimated. The positive variance is assured if $\alpha_0 > 0$, $\alpha_i \geq 0\ \forall i$ and $\beta_j \geq 0\ \forall j$. The model is stationary if $\sum_{i=1}^{P}\alpha_i + \sum_{j=1}^{Q}\beta_j < 1$.

1.2 GJR model

It was shown, firstly by Black (1976), that there is usually different impact of the positive and negative shocks on the volatility. GJR model, proposed by by Glosten et al. (1993), takes this into account. It is similar to the GARCH model (3), but if the previous innovation was negative (dummy variable i_{t-j}) the impact on volatility is bigger (by the parameter γ_j). The model takes the following form,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{P} \alpha_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \cdot \varepsilon_{t-j}^2 + \sum_{j=1}^{Q} \gamma_j \cdot i_{t-j} \cdot \varepsilon_{t-j}^2 , \tag{4}$$

where i_{t-j} is a dummy variable which equals to one when innovation ε_{t-j} is negative and null otherwise. Variables α_0, α_i, β_j and γ_j are parameters needed to be estimated. The positive variance is assured if $\alpha_0 > 0$, $\alpha_i \geq 0\ \forall i$, $\beta_j \geq 0\ \forall j$ and $\beta_j + \gamma_j \geq 0\ \forall j$ and model is stationary if $\sum_{i=1}^{P}\alpha_i + \sum_{j=1}^{Q}\beta_j + \frac{1}{2}\sum_{j=1}^{Q}\gamma_i < 1$.

2 VaR and backtesting procedure

Value at Risk (VaR) is nowadays commonly accepted measure of the risk. If we assume a random variable X – the profit from asset / portfolio with the (un)known distribution function F_X, VaR at a given probability level α is the maximum loss which will occur in $1-\alpha$ cases (confidence level),

$$VaR_\alpha(X) = \sup\{x \in R : F_X(-x) \geq \alpha\}. \tag{5}$$

VaR is usually estimated for one day ahead period and then (if needed) recalculated for longer periods. Mostly utilized values of α are 15%, 5%, 1% and 0.5%. For further explanation of VaR concept and methods utilized for its estimation see e.g. (Jorion, 2006).

There are three basic approaches to VaR estimation – (i) analytical formula utilizing parametrical probability distribution function, (ii) stochastic (Monte Carlo) simulation that estimates the quantile of a given distribution numerically, and (iii) (filtered) historical simulation that relates VaR estimation to the quantile obtained from historical observations. For all the models the accurate prediction of returns' volatility is fundamental. In this paper we assume that the financial returns can be modelled by the AR-GARCH and AR-GJR processes, which were described in the previous section. When parameters of these processes are estimated, VaR can be calculated as follows,

$$VaR_{\alpha,t+1} := VaR_\alpha(x_{t+1}) = \hat{\mu}_0 + \sum_{i=1}^{R} \hat{\mu}_i \cdot x_{t-i+1} + \hat{\sigma}_{t+1} \cdot q_\alpha, \tag{6}$$

where $\hat{\mu}_0$ and $\hat{\mu}_i$ are estimated coefficients of conditional mean, $\hat{\sigma}_{t+1}$ is estimated standard deviation by one of the models defined in previous section and q_α is an appropriate quantile of the innovations distribution (Student distribution).

By means of backtesting procedure the model is verified. This procedure is based on the comparison of the risk estimated at time t for time $t+1$ with the true loss observed at time $t+1$. Within the backtesting procedure on a given time series the following two situations can arise – the loss is higher or lower than its estimation,

$$I_t = \begin{cases} 1 & if\ r_t < -VaR_t \\ 0 & if\ r_t \geq -VaR_t \end{cases}. \tag{7}$$

While the former case is denoted by 1 as an exception, the latter one is denoted by zero. If the model is accurate, than roughly $(1-\alpha)\cdot n$ exceptions (where n is the length of the data set utilized for backtesting) should be experienced. Bigger quantity of exceptions means, that the model underestimates the risk and vice versa. For further details see e. g. (Hull, 2006).

Mostly applied statistical tests are due to Kupiec (1995) and Christoffersen (1998). Kupiec's test (henceforth K-test) is derived from a relative amount of exceptions, i.e. whether their quantity is from the statistical point of view different from the assumption. The null hypothesis is that the observer probability of exception occurring is equal to the assumed. A given likelihood ratio on the basis of χ^2 probability distribution with one degree of freedom is formulated as follows:

$$LR = -2\ln\left[\frac{\pi_{ex}^{n_1}\left(1-\pi_{ex}\right)^{n_0}}{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}\right], \tag{8}$$

where π_{ex} is expected probability of exception occurring, π_{obs} is observed probability of exception occurring, n_0 is the number of zeros and n_1 is the number of ones (exceptions).

By contrast, in order to assess whether the exceptions are distributed equally in time, i.e. without any dependence (autocorrelation), we should define the time lag first. Therefore, we replace the original sequence by a new one, where 01, 00, 11 or 10 are recorded. The null hypothesis is that the probability of exception occurring is independent on the information whether the exception has occurred also previous day. Then we have the likelihood ratio as follows (C-test):

$$LR = -2\ln\left[\frac{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}{\pi_{01}^{n_{01}}\left(1-\pi_{01}\right)^{n_{00}}\pi_{11}^{n_{11}}\left(1-\pi_{11}\right)^{n_{10}}}\right], \tag{9}$$

where $\pi_{ij} = \Pr\left(I_t = j \middle| I_{t-1} = i\right)$ and $\pi_{obs} = \dfrac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}$. This test statistic has χ^2 probability distribution with one degree of freedom.

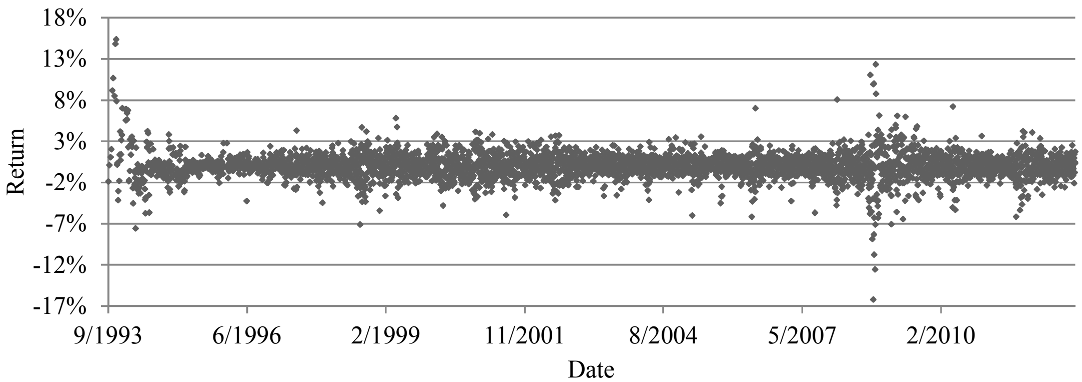
3 Results

Particular models described in the previous section will be applied on data series of log-returns calculated from historical series of Prague stock market index (index PX-50 and PX, henceforth only index) for period from September 5, 1993 till October 5, 2012.[1] The input

[1] The index PX is calculated from March 20, 2006. In that date it took over the values of the index PX-50, which is calculated from April 5, 1994. The previous values of index PX 50 has been calculated ex-post.

data was downloaded from the webserver www.czechwealth.cz. The size of the series is 4,627 daily returns, from which the first 250 observations are left for initial estimation of parameters and 4,377 observations are utilized for backtesting procedure. The evolution of the log-returns is depicted in Fig. 1.

Fig. 1 The evolution of returns of index from 5. 9. 1993 till 5. 10. 2012.



Source: data from www.czechwealth.cz, own elaboration

It is apparent that there are clusters with higher volatility of returns. The highest volatility is in the second half of the year 2008 – last financial crisis. Also other periods possess increased volatility of returns –summer 2006, second half of the year 1998, the end of 1993, etc. Basic descriptive characteristics of returns are depicted in Tab. 1.

Tab. 1: Statistic description of observed returns

Characteristics	Value
Mean	0.023%
Median	0.028%
Standard deviation	1.529%
Skewness	0.301
Kurtosis	13.857
Returns autocorrelation	0.191
Autocorrelation of the square of the returns	0.495

Source: own calculation

The following results from the characteristics are apparent: (i) although the median is close to the mean, due to nonzero skewness we can conclude that the probability distribution of the returns is skewed, (ii) also high kurtosis suggests the presence of heavy tails so that the normal distribution cannot be utilized for modelling purposes, (iii) the autocorrelation of the mean is small, but present and (iv) the autocorrelation of the square of the returns is relatively high – this suggest that the use of conditional volatility model is necessary.

For the parameters estimation of whole period process only the first autoregressive coefficient and only first order coefficients for volatility equation were found statistically significant for both assumed models. Thus we further assume only 1-1-1 models. In the Tab. 2 we provide the summary of parameters estimated from the whole dataset as well as the values of log-likelihood function (LLF), Akaike information criteria (AIC) and Bayesian information criteria (BIC).[2] As can be seen from the table, LLF (and thus also AIC and BIC) is slightly higher for the GJR model, but estimated parameters are similar for both models (parameter γ_1 is relatively small).

Tab. 2: Estimated parameters of particular models from the whole period

Model	AR(1)-GARCH(1,1)-t	AR(1)-GJR(1,1)-t
Estimated parameters	$\mu_0 = 0.00057$ $\mu_1 = 0.129$ $\alpha_0 = 2.482e-6$ $\alpha_1 = 0.848$ $\beta_1 = 0.151$ $\nu = 7.589$	$\mu_0 = 0.00048$ $\mu_1 = 0.131$ $\alpha_0 = 2.755e-7$ $\alpha_1 = 0.846$ $\beta_1 = 0.115$ $\gamma_1 = 0.0695$ $\nu = 7.767$
LLF	1,4015.4	1,4022.9
AIC	-2,8017	-2,8030
BIC	-2,7972	-2,7978

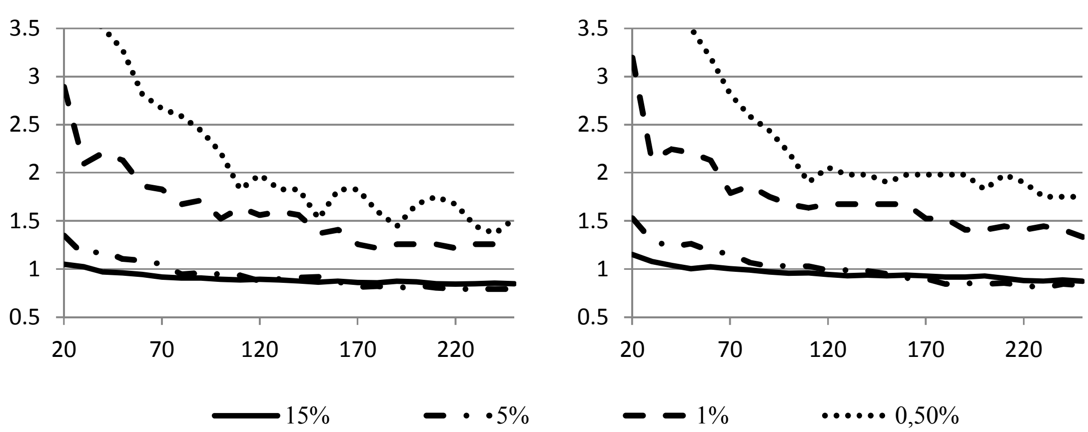
Source: own calculation

[2] Akaike information criteria is computed as $2 \cdot k - 2 \cdot LLF$ and Bayesian information criteria (BIC) is computed as $\log(n) \cdot k - 2 \cdot LLF$, where n is the number of observations, k is the quantity of parameters to be estimated and LLF is the value of log-likelihood function.

The models (each with $R=1$, $P=1$, $Q=1$) were backtested with different periods utilized for parameters estimation and the ratio of observed quantity of exceptions to the expected quantity were recorded. The results are depicted on Fig. 2. The ratios should be ideally equal to one or at least close to one (as we want the quantity of observed exceptions to be approximately equal to the expected quantity). As can be seen, the trend of the ratio is the same for both models (and also for all the probability levels). The ratio of observed quantity of exceptions to the expected quantity is decreasing with the increasing size of period utilized for parameter estimation. The explanation of this is probably following. With the increase of the period length the estimated parameter ν is increasing on average (as with the longer period the extreme values are more likely to be observed – tails are heavier), which increases the estimated Value at Risk.

From the figure we can clearly distinguish the difference between two groups of ratios. Ratios for probability levels 15% and 5% are around 1. For the short estimation periods (lower than 70 days) they are above 1 – the risk is underestimated; for longer estimation periods (above 70 days) they are below 1 – the risk is overestimated. Ratios for probability levels 1% and 0.5% are above 1 for all the assumed estimation periods. Thus, the risk is always underestimated. For the best risk estimation, the period of length 240 days should be utilized for parameters estimation.

Fig. 2 The ratios of observed and expected quantities of exceptions for different periods utilized for parameters estimation (from 20 days till 250 days with the step of 10 days) and different models (GARCH-t in left, GJR-t in right)



Source: own calculation

In the Tab. 3 and Tab. 4 the p-values of the defined statistical tests are summarized for estimation period size of 70 and 240 days. According to the p-values we can reject the bunching of exceptions (all the p-values of C-test are higher than 5%). However, the accuracy of the model is problematic. For probability levels 15% and 5% the models are accurate only when parameters are estimated from last 70 days. For the probability levels 1% and 0.5% (the levels important for financial institutions and their regulatory authorities) the GJR model can be rejected as inaccurate (all the p-values are lower than 5%) and the GARCH model is accurate for 240 days period utilized for parameters estimation.

Tab. 3: P-values of tests for utilized period of parameters estimation of 70 days

model	GJR-t				GARCH-t			
probability level	15%	5%	1%	0,50%	15%	5%	1%	0,50%
P-value of K test	95,2%	7,0%	0,0%	0,0%	6,9%	55,2%	0,0%	0,0%
P-value of C test	94,7%	44,7%	19,0%	82,4%	8,6%	6,8%	8,9%	61,8%

Source: own calculation

Tab. 4: P-values of tests for utilized period of parameters estimation of 240 days

model	GJR-t				GARCH-t			
probability level	15%	5%	1%	0,50%	15%	5%	1%	0,50%
P-value of K test	1,3%	6,2%	4,7%	1,4%	0,1%	1,1%	20,4%	20,2%
P-value of C test	66,2%	6,3%	55,0%	19,7%	49,0%	7,5%	35,9%	61,8%

Source: own calculation

Conclusion

The development of the accurate model for the risk estimation is essential for all the financial institutions. Challenging part of each method is volatility modelling and prediction. In this paper we have shown that the backtesting results are also influenced by the size of the period utilized for parameter estimation. For the investment into Prague stock market index we have shown that for the backtesting results of described models the bunching can be rejected, i.e. models react to the volatility increase instantly. But these models are not accurate and more sophisticated models should be assumed. We also found out that with the increasing size of the period for parameters estimation the quantity of exceptions is decreasing.

Acknowledgment

The research was supported by the European Social Fund under the Opportunity for young researchers project (CZ.1.07/2.3.00/30.0016) as well as by GAČR (Czech Science Foundation – Grantová Agentura České Republiky) under the project no. GP13-18300P. The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated.

References

Alexander, C., & Sheedy, E. (2008). Developing a stress testing framework based on market risk models. *Journal of Banking and Finance*, 32(10), 2220-2236. doi: 10.1016/j.jbankfin.2007.12.041

Black, F. (1976). The pricing of Commodity Contracts. *Journal of Financial Economics*, 3(1-2), 167-179. doi: 10.1016/0304-405X(76)90024-6

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics,* 31(3), 307-327. doi: 10.1016/0304-4076(86)90063-1

Engle, R. F. (1982). Auto-regressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007. doi: 10.2307/1912773

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility on the nominal excess returns on stocks. *Journal of Finance*, 48(5), 1779-1801. doi: 10.1111/j.1540-6261.1993.tb05128.x

Huang, J. J., Lee, K. J., Liang, H., & Lin, W.F. (2009). Estimating value at risk of portfolio by conditional copula-GARCH method. *Insurance: Mathematics and Economics*, 45(3), 315–324. doi: 10.1016/j.insmatheco.2009.09.009

Hull, J. (2006). *Risk Management and Financial Institutions*. Upper Saddle River: Prentice Hall.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841-862. doi: 10.2307/2527341

Ignatieva, K., & Platen, E. (2010). Modelling Co-movements and Tail Dependency in the International Stock Market via Copulae. *Asia-Pacific Financial Markets*, 17(3). 261–302. doi: 10.1007/s10690-010-9116-2

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. (3rd ed.). New York: McGraw-Hill.

Kresta, A., & Tichý, T. (2012). International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *Finance a úvěr – Czech Journal of Economics and Finance*, 61(2), 141-151. Retrieved from http://journal.fsv.cuni.cz/mag/article/show/id/1244

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73-84. doi: 10.3905/jod.1995.407942

Contact

Aleš Kresta
VŠB – Technical University of Ostrava, Faculty of Economics
Sokolská třída 33, Ostrava, Czech Republic
ales.kresta@vsb.cz