ON CONDITIONAL VOLATILITY MODEL COMPARISON

Petr Seďa

Abstract

Volatility can be defined and measured as a risk of financial instrument over a specified time period. In this paper, we deal with volatility model selection and comparison in a specific framework. In particular, univariate volatility models like traditional ARCH model and its extensions will be object of our interest. Selection of the best suitable model may be usually based on in-sample or out-of-sample criteria. In empirical studies, we usually favour model that can capture real features of the data analysed and, in addition, can provide the most accurate out-of-sample forecast quality. In this paper, we focus just on out-of-sample comparison of linear and nonlinear ARCH family models which may follow two different approaches. In first approach, alternative models are contrasted by different loss functions based directly on variance forecast and Diebold-Mariano type tests. The second approach includes indirect evaluation methods which consider using of alternative variance forecasts. In our study, we deal with the evaluation of alternative ARCH family models within a VaR framework. Empirical comparison of the methods discussed above will be demonstrated on illustrative example using sample data from U.S. stock market. We consider daily data of S&P500 index in the period of 2007 - 2012 years which includes the stage of recent global financial crisis of 2008-2009 years.

Key words: conditional volatility, Diebold-Mariano test, MCS approach, out-of-sample forecast, VaR.

JEL Code: C52, C53, C58.

Introduction

Model comparison and model selection, of the second moment of a given random variable, have been widely discussed topic in the financial engineering for many years. In the financial econometrics theory there have been derivate and proposed several approaches to verify if a given model is able to replicate empirical features observed on sample data. In this paper, we will focus on model comparison and model selection taking in consideration univariate conditional volatility models.

The GARCH or generalized autoregressive conditional heteroskedasticity models have become a standard tool in modeling the conditional variances of returns from financial time series data; see (Bollerslev, 1986). The popularity of these models lies in their compatibility with major facts for asset returns, the existence of efficient statistical methods for estimating model parameters, and the availability of useful volatility forecasts.

Traditional methods for model comparison and selection can be relatively simply adjusted and applied within specific group of models, see (Patton & Sheppard, 2009). The aim of this paper is to provide out-of-sample comparison of linear and nonlinear ARCH family models based on out-of-sample forecast using two different approaches. In first approach, alternative models are contrasted by different loss functions based directly on variance forecast and Diebold-Mariano type tests (Diebold & Mariano, 1995). The second approach includes indirect evaluation methods which consider using of alternative variance forecasts, in particular a VaR framework (Caporin, 2008).

The studies conducted in this paper will be summarized in the following way. First, the conditional heteroskedasticity models will be introduced. Second, it will be discussed the model selection distinguishing direct and indirect model evaluation based on out-of sample forecast. In Section 3, the data sample will be described and statistically analyzed. Section 4 reports an empirical example on a set of stock market indices, and Section 5 concludes the analysis.

1 Conditional volatility model specification

In this section, we define some simple and univariate volatility models. We consider traditional symmetric GARCH (1, 1) model and its extension, the EGARCH (1, 1) model, which is able to capture asymmetry in variance. Normally, the GARCH (1, 1) model taking the form of discrete data and considering heteroskedasticity may be according to (Bollerslev, 1986) defined as follows:

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \tag{1}$$

$$\varepsilon_t = \xi_t \sigma_t, \tag{2}$$

$$\xi_t \sim N\left(0,1\right), \tag{3}$$

where σ_t^2 is conditional variance estimated here. The GARCH heteroskedasticity is defined by (1). Parameters $\beta_0, \beta_1, \beta_2$ are constants satisfying the conditions of $\beta_0 > 0,\ \beta_1 + \beta_2 < 1$. In the GARCH (1, 1) model, there is a stringent trade-off between the possibility of having sharp changes in the short term volatility represented by high value of the parameters $\beta_1 + \beta_2$ and the ability to capture the long memory behavior of volatility through $\beta_1 + \beta_2$.

A GARCH-class model named Exponential GARCH proposed by (Nelson, 1991) allows for asymmetric effects and therefore solves one of the important shortcomings of the symmetric models. While the GARCH (1, 1) model imposes the nonnegative constraints on the parameters, the EGARCH (1, 1) models the log of the conditional variance so that there are no restrictions on these parameters:

$$\log\left(\sigma_t^2\right) = \beta_0 + \beta_1 \left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \beta_2 \log\left(\sigma_{t-1}^2\right) + \gamma_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}}. \tag{4}$$

Note that the left side is the logarithm of the conditional variance. This implies that the leverage effect is exponential and that forecast of the conditional variance is guaranteed to be nonnegative. The presence of leverage effects can be tested by the hypothesis that $\gamma_1 < 0$. Bad news can have a larger impact on volatility, and the value of γ_1 is expected to be negative.

2 Model selection and evaluation

In this section, it will be illustrated some approaches to model comparison and model selection using well-known univariate volatility models. Selection of the best suitable model might be based on in-sample or/and out-of-sample criteria. In this paper, we will focus just on out-of-sample criteria. However, a practical identification of an optimal model requires an optimal balance between these two groups of criteria. In empirical studies, a trade-off between these two approaches actually exists; see (Patton & Sheppard, 2009). A forecasting will have a greater emphasize on out-of-sample fit, while a structural analysis will concentrate on in-sample outcomes.

The out-of-sample criteria of ARCH family models selection may be considered from two different perspectives. A direct evaluation and comparison of volatility forecast represents first alternative, while an indirect comparison of volatility models through the possible uses of the variance forecast constitutes the second approach.

2.1 Direct model evaluation

Within direct model evaluation and comparison, some alternative models are usually compared by tests which are directly based on variance forecast. Let define $\hat{\sigma}_t^2$ as variance forecast of a model at time t and σ_t^2 as unknown variance at time t. For each estimated models, we can evaluate a set of standard quantities like mean absolute error (MAE) or mean squared error (MSE).

However, model equivalence can be verified by more formally using the approach defined by (Diebold & Mariano, 1995). This approach compares alternative models using loss functions differentials. According (Patton, 2011), using the proxies for the underlying volatility cause distortions for some loss functions. Therefore (Patton, 2011) proved that some class of homogeneous loss functions is robust, and accordingly allows an unbiased model comparing. In this paper, the MSE and QLIKE loss functions will be used as given below:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(s_t^2 - \hat{\sigma}_t^2\right)^2, \tag{5}$$

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}\left(\ln \hat{\sigma}_t^2 + \frac{s_t^2}{\hat{\sigma}_t^2}\right)^2, \tag{6}$$

where s_t^2 is a proxy for unobserved volatility σ_t^2. Alternative models can be compared by testing the predictive ability that is associated with the null hypothesis of the unconditional expected null loss function difference as referred below:

$$H_0 : E\left[MSE(j)\right] - E\left[MSE(i)\right] = E\left[MSE(j) - MSE(i)\right] = 0. \tag{7}$$

A similar expression can be written for QLIKE loss function. According (Diebold & Mariano, 1995) under mild conditions of the loss function differences, the test statistic τ_{MSE} takes the form:

$$\tau_{MSE} = \frac{\overline{lf}_{MSE}(j,i)}{\sqrt{Var\left[lf_{MSE,t}(j,i)\right]}}, \tag{8}$$

where $\overline{lf}_{MSE}(j,i) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(s_t^2 - \hat{\sigma}_{j,t}^2\right)^2 - \left(s_t^2 - \hat{\sigma}_{i,t}^2\right)^2\right]$, and $Var\left[lf_{MSE,t}(j,i)\right]$ is a variance estimator which is heteroskedasticity and autocorrelation consistent.

The comparison of models based on Diebold-Mariano type test is limited since it represents pairwise comparison only. Therefore, it is not possible to exclude having different models rankings connected with different loss functions. In order to resolve this limitation, several approaches have been defined in literature; see (Hansen, Lunde & Nason, 2011). They developed the model confidence set (MCS) approach. The MCS approach constitutes a testing framework for the null hypothesis of equivalence across models. Let assume that the set *M* contains a number of different models which have been used for forecasting. The null hypothesis of the MCS is defined as:

$$H_0 : E\left[lf_{MSE,t}\left(j,i\right)\right] = 0, i > j, \forall i, j \in M. \tag{9}$$

The null hypothesis can be verified using the following two test statistics:

$$t_R = \max_{i,j\in M} \left| \frac{\overline{lf}_{MSE}\left(j,i\right)}{\sqrt{Var\left[lf_{MSE,t}\left(j,i\right)\right]}} \right|, \tag{10}$$

$$t_{SQ} = \sum_{i,j\in M, j>i} \left(\frac{\overline{lf}_{MSE}\left(j,i\right)}{\sqrt{Var\left[lf_{MSE,t}\left(j,i\right)\right]}} \right)^2. \tag{11}$$

The both test statistics are based on bootstrap estimation of variance $Var\left[lf_{MSE,t}\left(j,i\right)\right]$. Since the distribution is non-normal, the rejection area is identified using bootstrap *p*-values under the null hypothesis. If the null hypothesis of equal ability of prediction among all models is rejected, the worst model is then excluded from the set *M*. To identify such a model we can use following elimination rule:

$$j = \arg\max_{j\in M} \left(\sum_{i\in M, i\neq j} \overline{lf}_{MSE}\left(j,i\right) \right) \left(Var\left[\sum_{i\in M, i\neq j} \overline{lf}_{MSE}\left(j,i\right) \right] \right)^{-\frac{1}{2}}. \tag{12}$$

More details on the MCS method with elimination rule are included in (Hansen, Lunde & Nason, 2011).

2.2 Indirect model evaluation

Since traditional econometric methods were innovated in last years; see (Hančlová & Nevima, 2008), indirect methods of model evaluation take into consideration alternative variance forecasts. The literature has recently focused mainly on evaluation of alternative GARCH specification within a Value-at-Risk (VaR) framework; see (Caporin, 2008). Empirical research dealt with test for the evaluation of VaR forecast. Within this framework, we consider a variable which displays heteroskedasticity, characterized by a time-varying mean and conditional density according to:

$$x_t \mid I^{t-1} \sim f\left[x_t, \mu_t, \sigma_t^2, \theta\right]. \tag{13}$$

where I^{t-1} denotes the information set up to time t-1 and θ is a vector containing additional parameters. The one day VaR for x_t is defined as given below:

$$\alpha = \int_{-\infty}^{VaR(x_{t+1},\alpha)} f\left(x_{t+1}, E\left[\mu_{t+1} \mid I^t\right], E\left[\sigma_t^2 \mid I^t\right], \hat{\theta}\right) dx_{t+1}, \tag{14}$$

where time varying mean and variance are substituted by their conditional expectations, and α represents the VaR confidence level. In this paper, we provide an interpretation of VaR model comparison using the MCS approach which was recently introduced. Loss functions based on VaR forecast have been defined by (Lopez, 1999) and (Caporin, 2008) as follows:

$$\mathrm{IF} = I\left(x_t < \mathrm{VaR}\left(x_t, \alpha\right)\right), \tag{15}$$

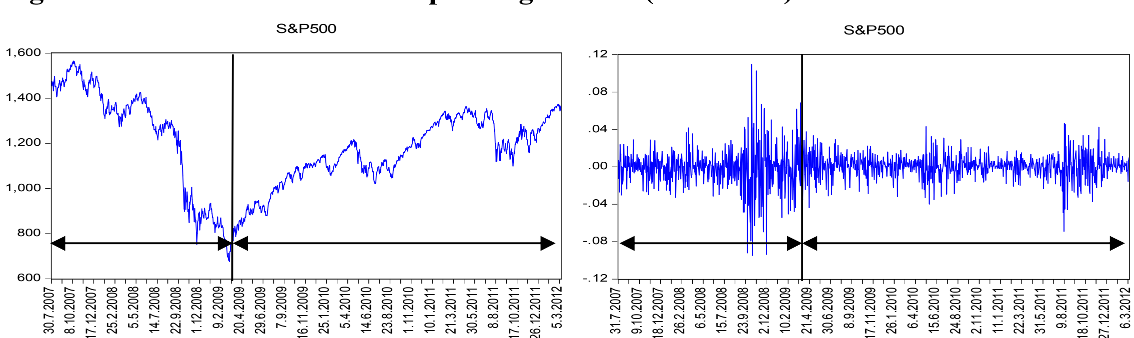
$$\mathrm{PIF}_t = I\left(x_t < \mathrm{VaR}\left(x_t, \alpha\right)\right)\left(1 + \left(x_t - \mathrm{VaR}\left(x_t, \alpha\right)\right)^2\right). \tag{16}$$

The indicator loss function, IF identifies exceptions, while the penalized indicator function, PIF penalizes exceptions by means of squared deviation between respective returns and VaR. The main benefit of these loss functions lies in a fact that they don´t rest on the true volatility but depend on the volatility forecast. These methods can be therefore used within the MCS approach to compare alternative models.

3 Data sample

Empirical analysis is performed on a data set of daily closing rates of S&P500 index in the period of 2007-2012 years. It includes total of 1213 daily observations. This period was chosen purposely, to investigate changes of the U.S. equity market volatility during time with a special emphasis on the behavior in the time during and after the global financial crisis of 2008-2009 years. The returns r_t at time t were defined as the logarithm of S&P500 index p, that is, $r_t = \log\left(p_t - p_{t-1}\right)$. Visual inspection of the plot of daily values and returns series of respective index proved very useful, see Fig. 1. Following the spread of bad news about U.S financial crisis, we have seen a more than 60 percent decline of S&P500 index in 2008. It can be seen that from Fig. 1 that return fluctuates around mean value that is close to zero.

Fig. 1: S&P500 values and corresponding returns (2007-2012)



Source: own calculations, data: *www.standardandpoors.com*

Volatility is low for certain time periods and high for other periods. The movements are in the positive and negative territory and larger fluctuations tend to cluster together separated by periods of relative calm. However, the global financial crisis of 2008-2009 years changed it. Most asset classes experienced significant pullbacks, the correlation between asset classes increased significantly and the markets have become extremely volatile, see (Seďa, 2012). Since the volatility was highest in 2008 when the values of S&P500 index reached the minimum values in investigated period, we divided the basic period of 2007-2012 years into two testing periods. First, the crisis period, started in July 2007 and finished by March 2009, while second, the post-crisis period, was defined from April 2009 to March 2012. Tab. 1 shows several descriptive statistics and the results of the unit root test for S&P500 stock returns. Symbol (*) means a rejection of relevant null hypothesis at the 5% significance level.

Tab. 1: Descriptive statistics of S&P500 index

Period	Mean	Stand. dev.	Skewness	Kurtosis	*J-B* test	*L-B Q*(12)
crisis	-0,0019	0,0223	-0,1069	7,5169	365,5049*	36,431*
post-crisis	0,0009	0,0131	-0,0783	6,4942	399,6424*	30,587*

Source: own calculations

The means of both sample returns are quite small, while the standard deviations are significantly higher. Based on the values of skewness, increased values of kurtosis and Jarque-Bera (J-B) test of normality, the daily S&P500 return series show mostly leptokurtic distribution which has a higher peak and heavy tail, instead of normal distribution. The Ljung-Box (LB) statistics Q_n for the squared return series are extremely high which indicate rejection of the null hypothesis of no correlation. The sample statistics of analyzed time series data indicate that it is desirable to consider heteroscedasticity and jump risk when estimating the volatility of the U.S. stock market.

4 Empirical analysis

The aim of this section is to demonstrate and present an empirical comparison of the methods discussed and described in the previous sections. For both return series we fit two specific conditional heteroskedasticity models, namely symmetric linear GARCH (1, 1) and nonlinear EGARCH (1, 1) models. Both models are estimated under assumption of GED distribution of errors. These two models are then used to produce one-step-ahead volatility forecast. The models are compared using the methods described in the previous sections. In particular, we consider the Diebold-Mariano test using the MSE a QLIKE loss functions across the models, the MCS approach using again the MSE a QLIKE loss functions, and the loss functions defined by (15) and (16) at 5% VaR level.

The out-of-sample comparisons start from the outcomes of the Diebold-Mariano test by means of MSE and QLIKE loss functions. Our aim is to evaluate and compare GARCH and EGARCH models performance across different market phases (crisis and post-crisis periods). Therefore, we consider separately 2008 year which belongs to the global financial crisis period, and 2011 year that is a part of post-crisis period. We compare two years with very different volatility and returns.

Tab. 2 reports some characteristic empirical findings. Diebold-Mariano test evaluates the null hypothesis of zero expected difference between loss functions of GARCH and EGARCH models. Significant values at 5% confidence level indicate a preference of GARCH model if the value is positive (in italics) or asymmetric EGARCH model when the value of test statistic is on the other hand negative (in bold).

Tab. 2: Diebold-Mariano test statistics for 2008 and 2011 years

Test	MSE loss fun. (2008)	QLIKE loss fun. (2008)	MSE loss fun. (2011)	QLIKE loss fun. (2011)
Die.-Mar.	-1,045	0,614	**-2,152**	-1,058

Source: own calculations

Focusing on MSE loss function, the null hypothesis of zero loss function differentials rejected in 2008 year, while in post crisis year 2011 the value of test statistic is negative and statistically significant at 5% significance level which prefer nonlinear EGARCH model. When we consider QLIKE function, the null hypothesis is not rejected for both sample years.

It is clear that some preferences between models may occur in favour of asymmetric EGARCH model. Though it is not a case of this study, the limitation of the Diebold-Mariano test lies in possibility to consider just pairwise comparison. As suggested in section 2.1, the MCS method overcomes this restrictive comparison.

The MCS results based on the t_R statistics for selected out-of-sample periods are reported in Tab. 3. For S&P500 return series we evaluated two alternative volatility models by means of the MSE and QLIKE functions as well as the loss functions defined in (15) and (16). Tab. 3 shows the model confidence set over different loss functions and periods. Bold values denote that models are included at 10% confidence interval in the confidence set. These models are statistically equivalent if compared using the loss functions.

Tab. 3: Model confidence set for 2008 and 2011 years

	MSE loss function	QLIKE loss function	IF (5%)	PIF (5%)
GARCH (2008)	**0,27**	0,04	**1,00**	**1,00**
EGARCH (2008)	**0,52**	0,04	**0,25**	**0,14**
GARCH (2011)	0,05	0,06	0,04	**0,19**
EGARCH (2011)	**1,00**	**1,00**	**0,47**	**0,37**

Source: own calculations

If we consider S&P500 return series, the results differs between models and out-of-sample testing periods. In 2008, MSE considers both models equivalent, QLIKE doesn´t include models into confidence set at all, while IF and PIF indicate again rather preference for EGARCH. For 2011, some differences appear across the loss functions and models. For MSE and QLIKE, the optimal model is clearly EGARCH. IF excludes GARCH model from the confidence set, while PIF considers both models equivalent. To sum up, our findings may be interpreted as confirmation of the direct model comparison outcomes, which indicates a preference rather for EGARCH model. In other words, more flexible model is to be preferred.

Conclusion

In this paper, we focused on some existing methods for univariate conditional volatility model selection and model comparison. We consider selected relatively new out-of-sample approaches which are based on model forecast evaluation. First, we focused on direct evaluation methods, namely MSE and QLIKE criteria and relevant Diebold-Mariano test. In order to solve limitation associated with these methods we also considered the MCS approach. In the next step, we moved on indirect methods of model evaluation which consider more general loss functions based on VaR forecasts, compared using the MCS approach.

Finally, we presented an empirical study by means of the less usual approaches for model selection and comparison. Specially, non-nested hypothesis testing and VaR-based loss functions. In this paper, we focused just on out-of-sample comparison of linear and nonlinear ARCH family models using sample data from U.S. stock market, namely daily returns of S&P500 index. Sample results suggest that there is a preference for nonlinear model. However, model preference depends in general on the sample period used for evaluation and on the loss function we consider.

The loss functions approaches can be simply used on the forecasts produced by other univariate models as well as multivariate specifications. The use of these techniques can be extended to cover also other ARCH family models as well stochastic volatility models.

Acknowledgment

This paper was supported by the Czech Science Foundation through project No. 13-13142S and the European Social Fund within the project CZ.1.07/2.3.00/20.0296.

References

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31,* 307-327.

Caporin, M. (2008). Evaluating Value-at-Risk measures in presence of long memory conditional volatility. *Journal of risk, 10,* 79-110.

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

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79,* 453-497.

Hančlová, J. & Nevima, J. (2008). *Innovation in education in the field of econometrics*. In *5th International Conference on Efficiency and Responsibility in Education* (pp. 61-66). Prague: Czech University of Life Sciences Prague.

Lopez, J. (1999). Regulatory evaluation of value-at-risk models. *Journal of Risk, 1,* 37-64.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59,* 347–370.

Patton, A. J. & Sheppard, K. (2009). *Handbook of Financial Time series*: *Evaluating volatility and correlation forecasts* (1st ed., pp. 801-838). Berlin: Springer Verlag.

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

Seďa, P. (2012). *Impact of the global financial crisis on stock market volatility: Evidence from central European stock market*. In *Mathematical Methods in Economics 2012* (pp. 787-792). Karviná: Silesian University.

Contact

Petr Seďa

Department of Mathematical Methods in Economics,

Faculty of Economics, VSB-TU Ostrava,

Sokolská třída 33, 701 21 Ostrava

petr.seda@vsb.cz